Volume 08 Issue 04 Published, November 17, 2004 ISSN 1535-864X

Intel® Technology Journal

Toward The Proactive Enterprise

The Proactive Enterprise is the infrastructure for future generations of information technology. This issue of Intel Technology Journal (Volume 8, Issue 4) examines the technologies for the Proactive Enterprise and Intel's research and development efforts in these areas.

Inside you'll find the following papers:

More information, including current and past issues of Intel Technology Journal, can be found at:
http://developer.intel.com/technology/itj/index.htm

Volume 08 Issue 04 Published, November 17, 2004 ISSN 1535-864X

Intel® Technology Journal

Toward The Proactive Enterprise

Articles

THIS PAGE INTENTIONALLY LEFT BLANK

Preface

By Lin Chao

Publisher, Intel Technology Journal

Information Technology (IT) is a core foundational business capability for companies. At Intel, IT delivers critical information solutions needed to keep us running and growing. Over the past several years, Intel has focused on IT innovation. This issue of Intel Technology Journal (Volume 8, Issue 4) examines the technologies for the "Proactive Enterprise" and Intel's research and development efforts in these areas.

The Proactive Enterprise is the infrastructure for future generations of information technology. Its aim is to move current IT system architectures and usage models from their current reactive approach toward more proactive and systematic approaches. In this mega issue of Intel Technology Journal (Volume 8, Issue 4), we delve into the details of the Proactive Enterprise. The Enterprise's four-layer framework, or capability stack, comprises 1) infrastructure, 2) information management, 3) adaptive and integrated business processes, and 4) collaborative and intuitive applications. The first paper in this issue provides an overview of Proactive Enterprise.

Infrastructure. The second paper looks at PlanetLab, a world-wide research platform which can be a test bed for examining IT challenges in the areas of viruses, security, content distribution, network and systems management, and provisioning. The third paper examines an autonomic framework for self-configuring network services. This paper presents the requirements for dynamically self-configuring network services and looks at the IBM Autonomic Computing Toolkit. The fourth paper looks at wireless network requirements for interactive multimedia, which typically requires high bandwidth due to the data rates and workload size. This paper presents a scalable and adaptive system-level approach for wireless multimedia in the Proactive Enterprise environment. The fifth paper examines network security. The most common solution available today for cyber security is the hardening of systems via "patching." This paper discusses how policy-enabled network security, complemented by system hardening, can provide a proactive strategy by reducing the likelihood of cyber threats and by controlling their spread. The sixth paper examines the idea of "Internet Suspend and Resume" where a user's entire personal computing environment, including the operating system, and its applications, data files, customizations, and current computing state is maintained in centralized storage. By using virtual machine technology, this computing environment may be easily transported and quickly instantiated on any Internet suspend and resume (ISR) client machine.

Data Fusion/Information Management. Papers seven and eight look specifically at data fusion. The seventh paper discusses how statistical learning can be a tool for semiconductor manufacturing. Semiconductor manufacturing data include high dimensionality (many thousands of variables), mixtures of categorical and numeric data, non-linear relationships, noise and outliers in both x and y axis, and temporal dependencies. To address these challenges, statistical-learning techniques are applied; this paper describes applications currently under development within Intel. The eighth paper looks at metadata management. Metadata answers the "who, what, when, where, why, and how" of

every piece of data being documented throughout the enterprise. This data can be integrated and managed through the enterprise. To achieve this enterprise-wide information integration, companies need to describe and share, in a common way, the data in their different data sources.

Business Processes. The ninth paper looks at Service-Oriented Architecture (SOA) and how it can serve as a building block for Service-Oriented Enterprise (SOE) with special attention paid to modeling and implementing business processes. This paper introduces a set of principles and guidelines for defining a reference architecture that consistently and accurately represents the SOA.

Intuitive and Collaborative Applications. Papers 10 and 11 highlight research done in the areas of Bayes networks and global team collaboration across geographical and cultural boundaries. Bayes network is a formalism based on probability and graph theory. A Bayes network models the probability distribution of a set of random variables as nodes in a graph, and the probabilistic dependencies among variables as arcs in the graph. Cause-effect are both elicited from experts and learned statistically with support of operational data. An illustration is given of how Bayes network can model the diagnostics for a vacuum subsystem in silicon wafer manufacturing. The 11th paper describes a multi-level approach for a framework for global team collaboration, including unique findings about Intel's remote teams' application framework.

One of the principal objectives of any IT organization is to support and enhance the individual productivity of each employee. These papers illustrate the heart of the Proactive Enterprise, which strives to harness the recent advances in machine intelligence to do more on behalf of the user (us) and allow the user to interact with the enterprise in more efficient and intuitive ways. We look forward to using the proactive–and smarter–enterprise.

Foreword

Building Intel Leadership in Enterprise Computing

Doug Busch
Vice President and Chief Information Officer, Intel Corporation

Many people think of "enterprise computing" as another name for data center operations. In fact, enterprise computing encompasses all of the elements of designing, developing, and operating the information solutions that keep a large enterprise running. Enterprise computing includes clients, servers, storage and networks, but the hardware components of enterprise IT spending are a small part of the overall investment companies make in their IT systems. For example, at Intel, only about 15% of our total information systems budget is spent acquiring hardware. The remainder is spent acquiring software and services, developing and deploying solutions, operating the huge installed base of systems, and supporting the Intel employees, customers and suppliers who use the systems. We are not alone; the industry is increasingly focused on the whole picture of enterprise computing, rather than the merits of individual hardware components.

As a result of competitive factors and technical barriers at the silicon level, Intel's historical focus on the price/performance of our semiconductor products is shifting to a focus on adding other kinds of value. If we are to deliver high-value features to our enterprise customers, it's critical that we understand the whole business of enterprise computing as well as we have historically understood manufacturing processes or microprocessor and memory architectures.

Enterprise computing spans an enormous range of issues. At the smallest scale, it involves the features needed for a secure handheld device or notebook computer, accurate tracking of individual hardware and software assets, or providing the ability to move seamlessly between WLAN access points. At the largest scale, it involves the operations and security of planet-wide communications networks, control of the complex data structures involved in a global company, the efficient monitoring and control of tens-of-thousands of servers and hundreds-of-thousands of clients, or designing systems that can easily be used by millions of users.

We are researching and designing building blocks for the computing industry which will change the game of developing and running enterprise-scale systems. Enterprise customers understand the value of these capabilities, and need the operational efficiency, reliability and ease of use they can deliver.

In this issue of Intel Technology Journal, we take a broad look at enterprise computing. We discuss research and new directions that enable a *Proactive Enterprise,* including new approaches to global team collaboration, service-oriented architectures and metadata management to enhance enterprise productivity, and business process transformation and application development. We explore the use of statistical or machine learning and data mining to enable efficient and agile manufacturing. And we discuss how virtualization technologies and new system-wide methods are enabling capabilities

for delivery of enhanced IT services and management of corporate distributed resources; this includes policy-based methods to manage security attacks or controlling their spread. With the background provided in this issue of ITJ, we hope the reader will be better informed of the new directions in enterprise computing.

Technical Reviewers

Robert Adams, Corporate Technology Group
Tony Benedict, Information Services and Technology Group
Satpal Birak, Information Services and Technology Group
Jim Brennan, Information Services and Technology Group
Jolyon Clarke, Technology and Manufacturing Group
Jonathan Clemens, Information Services and Technology Group
Martin Curley, Information Services and Technology Group
Denver Dash, Corporate Technology Group
Carl Dellar, Corporate Technology Group
Tom Gardos, Information Services and Technology Group
Bill Guthridge, Information Services and Technology Group
Sally Hambridge, Information Services and Technology Group
Susan R. Harris, Intel Communications Group
Patrick Holmes, Information Services and Technology Group
Adrienne Hudson, Technology and Manufacturing Group
Rob Knauerhase, Corporate Technology Group
Sridhar Mahanakali, Information Services and Technology Group
Gene Matter, Intel Communications Group
Gene Meieran, Technology and Manufacturing Group
Ray Mendonsa, Information Services and Technology Group
Don Michie, Information Services and Technology Group
William Pawlowski, Information Services and Technology Group
Padmanabhan Pillai, Corporate Technology Group
Glen Sweeney, Information Services and Technology Group
Timothy Verrall, Information Services and Technology Group
John Vicente, Information Services and Technology Group
Michael Waithe, Technology and Manufacturing Group
Stephen Zambroski, Technology and Manufacturing Group
Nathan Zeldes, Information Services and Technology Group

THIS PAGE INTENTIONALLY LEFT BLANK

The Proactive Enterprise

George Brown, Information Services and Technology Group, Intel Corporation
Thomas Gardos, Information Services and Technology Group, Intel Corporation
Jay Hopman, Information Services and Technology Group, Intel Corporation
Hong Li, Information Services and Technology Group, Intel Corporation
Sigal Louchheim, Information Services and Technology Group, Intel Corporation
Cynthia Pickering, Information Services and Technology Group, Intel Corporation
Jeff Sedayao, Information Services and Technology Group, Intel Corporation
John Vicente, Information Services and Technology Group, Intel Corporation

Index words: information technology, enterprise computing, proactive computing

ABSTRACT

Today's Information Technology (IT) organizations face significant challenges in delivering business value in these times of rapid architectural evolution as well as challenges in their ability to manage, provision, trust, and integrate the various elements of IT systems. An important aspect of these challenges is the impact IT has on the user and the user's interaction with IT systems. Through a systems view of IT, Intel's Information Services and Technology Group (ISTG) proposes a new direction for future capabilities within the IT enterprise. We present a layered taxonomy of capabilities common to IT systems and promote areas for breakthrough research that will enable the proactive enterprise.

INTRODUCTION

Just as Information Technology (IT) provides a significant advantage in a corporation's ability to compete in the growing Internet marketplace, a flexible and robust enterprise infrastructure also plays an important, competitive role by enabling employee productivity and globalization, allowing pervasive and secure business communications anywhere and anytime, and finally, by managing operational complexity and providing greater utilization of resource assets. IT organizations supporting large enterprises are facing many new challenges, as enterprises rely more heavily on IT services to manage business transactions and interactions occurring electronically and through the Internet.

Today, Intel's enterprise supports nearly 90,000 employees distributed in manufacturing facilities, campuses, and sales offices worldwide. As of the end of 2003, nearly 70,000 notebook computers were deployed with over 30,000 employees accessing networks wirelessly. Intel's IT operations maintain over 50,000 servers, close to 160,000 Local Area Network (LAN) nodes and manage 3.2 petabytes of server-based storage capacity. Moreover, in 2003, Intel employees made about 19,000 conference calls a week and received about 4.2 million e-mail messages per day. More than 60 percent of Intel's materials transactions and 85 percent of customer orders are processed electronically. Intel is aggressively working towards conducting all interactions with customers, suppliers, employees, and affiliates through the Internet [1].

The challenges of provisioning and managing IT services, distributed applications, and infrastructure will noticeably increase over the next five to ten years as the workforce continues to be more mobile and dispersed geographically and as computer-to-computer transactions continue to replace human-to-computer transactions. Storage capacity will continue to grow exponentially with the need to converge and convert these data repositories into valuable assets to compensate for the growing maintenance costs and to provide a competitive advantage. Business processes are becoming increasingly complex and integrated both within internal corporate business functions (e.g., manufacturing, design engineering, sales and marketing, and enterprise services) and across the external supply chain. Finally, a corporation's business depends critically on a robust, flexible information and communication infrastructure to enable corporate agility and productivity for business processes and user collaborative services.

The challenges for today's CIOs are to manage Total Cost of Ownership (TCO), reduce operating costs; promote innovation through stable introduction of new services, and grow infrastructure in a timely fashion to remain competitive with the industry. Ultimately the CIO's success will be measured by the delivery of significant new business value while maintaining a sustainable IT budget in the face of ever-increasing demand for IT services.

In this paper, we describe an IT research framework with the aim of identifying and developing promising new technologies that have the potential to deliver breakthrough improvements and at the same time reduce costs.

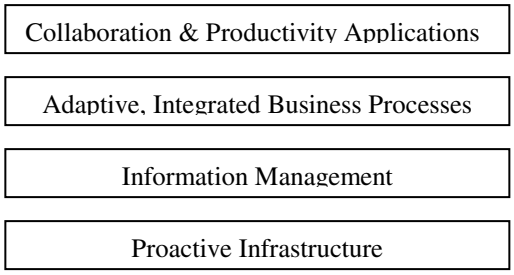

Figure 1: Proactive enterprise capability stack

The goal of our research is what we are calling the Proactive Enterprise and is based in part on Intel's Proactive Computing vision [2][3] that strives to move from human-centered computing to human-supervised computing. Similarly, our aim is to shift the current IT system architectures and usage models from their current, mostly reactive and human-centered state towards one that is more proactive, integrated, and human-supervised.

We illustrate our research on the Proactive Enterprise as a four-layer capability stack in Figure 1, and with more detailed research focus areas in Figure 2. The foundation of the Proactive Enterprise capability stack is the Proactive Infrastructure which addresses challenges in the distributed infrastructure of large enterprises such as security, mobility, virtualized platforms, communications and computing convergence, and autonomics. The layer above this deals with data management issues and the fusion of heterogeneous data and their repositories to form actionable information. The third layer, adaptive, integrated business processes, addresses issues critical to supply chain management and enabling of rapid business process systems. The top layer addresses user and application challenges of future enterprises with an emphasis on global team collaboration and individual as well as team productivity.

In the rest of this paper, we elaborate on the Proactive Enterprise and research focus areas, describe the research challenges and results, and present discovery opportunities for future work.

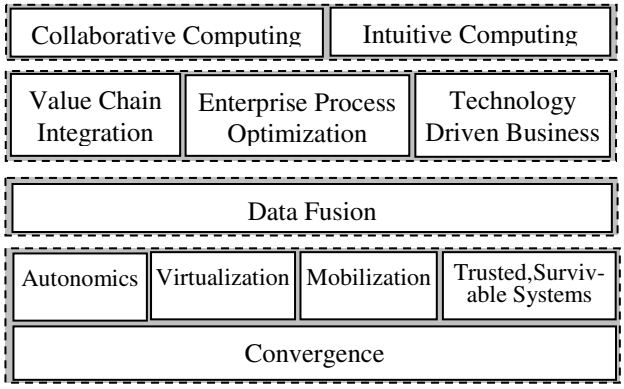

Figure 2: Proactive Enterprise research challenges

PROACTIVE INFRASTRUCTURE

IT managers should brace themselves for a significant challenge over the coming years in securing the enterprise and its users from un-trusted entities and security attacks, the expansion of alternative forms of physical devices or interconnection types, managing exponential TCO, and the ever-increasing end-user demands for rich media, and anywhere, anytime computing and communications.

The wider IT industry responses to these challenges have been incremental, and they have focused on reducing today's "pains" with point solutions, rather than integrated, system solutions. Fundamentally, the inherent issue with today's IT is its architecture, which is based on a discrete, static, and reactive framework. We envision the elements of the future IT enterprise infrastructure as structured and utilized in a highly virtualized, converged, dynamic, mobile, autonomic, trusted, and resilient manner. We refer to this layer as the "Proactive Infrastructure," since it fundamentally serves as the enabling layer of the Proactive Enterprise. To achieve the Proactive Infrastructure, we are promoting research in areas that either enable these characteristics or integrate aspects of these characteristics as infrastructure-based or service-oriented system solutions.

Autonomics

As the demands for distributed, virtualized, and global computing increase and Internet-based electronic businesses continue to grow, infrastructure growth and complexity increase along with them. While this growth is essential for business and the evolution of IT, it places rigorous demands on the operational environment. Network and system managers are faced with the challenge of managing the infrastructure for optimal service delivery while at the same time aiming to reduce

operational costs (e.g., manpower, operational tools). A current challenge for system managers is managing both the end-to-end service model and heterogeneous, discrete infrastructure elements, both in real-time and by using traditional offline methods. While traditional methods have been somewhat effective, they place rigorous demands on administrators in terms of skills, number of tools, and the need to reconcile end-end Fault, Configuration, Accounting, Performance, and Security management. We are presently doing systems research and development and use-case development in autonomic, adaptive and self-healing systems, which includes investigation into large-scale monitoring and provisioning systems and systems integration of component management systems. Current supported research and development includes applying methods of machine learning for statistical inference [4] active network management [5], real-time monitoring and fault detection; frameworks for automated network management, and distributed systems autonomics [6].

Virtualization

To support a global e-Business model, IT systems must have pervasive reach across physical and logical boundaries of the physical and service infrastructure. Geographic distances impose limitations on local computing services, secure access, and telecommunications infrastructure. Enterprise systems should offer services that are customized and configurable to enable users in different locations and job profiles to have similar access and availability to business information. Moreover, when employees travel and roam between Wi-Fi access points, whether it be at airport or coffee-shop wireless "hotspots," at home or in a hotel operating over a DSL or modem line, in a train or car using a cellular connection, or simply from a campus over a wireless or wire-line connection, the virtual enterprise must provide seamless service connectivity. Research in platform-based (i.e., hardware, OS) virtualization [7], distributed virtualization, and supporting technologies as well as in new virtualization usage models [8] that allow for segmentation of services, enable enhanced platform manageability and allow the exposure of new platform features out of band from operating system release schedules. Server virtualization will enable new manageability capabilities as well as greatly enhance upsizing and platform migration, and clustering and viability of secure GRID computing.

Mobilization

Today, demand for dynamic business environments has blurred the line between work and leisure and has further increased our mobility requirements. Sales forces, field service forces, fab workers, knowledge workers, and executives frequently find themselves in mobile situations. Even on the road, they have a critical need to maintain contact with events affecting their businesses and require accurate, up-to-date information concerning inventory, prices, and financial status, etc. Mobility to support reliable voice and data business communications and pervasive access to enterprise information and resources beyond the office building environment are key requirements, as telecommuters, road, and campus warriors need improved access to information, so that they can function as well outside the office as they can in the office. Smaller, powerful, less expensive equipment is now available; allowing access to e-mail and Web pages, while accommodations for screen size and processing power are scaling with these new devices. Further, notebook computers can rival most desktops in processing power, allowing applications to become mobile. On an enterprise level, the need for end-end security, QoS, seamless mobility, and operational resiliency is still a major challenge for the mobile enterprise, for delivery of traditional voice and data services as well as real-time, media-rich services. Systems research and development and use-case expansion of user mobility through various technologies including sensor or RFID*, BLUETOOTH*, ultra wideband (UWB), IEEE 802.11*, IEEE 802.16*, or third-generation wireless systems and mobile services including service discovery, mobile application software initiatives, personalization, and location-based services are relevant R&D that are being supported by IT research. Current research and development projects include cross-layer optimization, scalable, adaptive wireless multimedia streaming [9], overlay-based mobility, and mesh networks [10].

Convergence

The IT industry is undergoing significant change and reconfiguration worldwide due to a variety of industry forces [11]. This includes the growth of the Internet and e-Business, the deregulation and "horizontal" restructuring of telecommunication systems, the rapid rate of technology maturity in silicon and optics, the unification of voice and data communication services, and the wireless evolution and its momentum. These important changes have been prompted by the growing demand of the consumer base and the IT industry for increased infrastructure speed and media diversity, application flexibility, security, mobility, collaboration, virtualization, cost efficiency, and information advantage. Moreover, a

* Other brands and names are the property of their respective owners.

* Bluetooth is a trademark owned by its proprietor and used by Intel Corporation under license.

competitive battle is being waged by Original Equipment Manufacturers (OEMs), Operating System Vendors (OSVs), Independent Software Vendors (ISVs), and Internet Service Providers (ISPs) in data, voice, and media products and services towards a converged communication and computing industry.

Moreover, much of today's IT computing infrastructure and services are based on point-to-point linkages between clients and servers. Clients of a service typically must be explicitly configured with the servers they use. If those servers go down or are overwhelmed by a high demand, the service is unusable by such a client, even if other servers were still available. If an end user of a service moves to another location (particularly to another part of the world), the server that is configured may provide a poor response, with no automatic way to reconfigure the end-user's client software to use servers that are closer. We need ways to make applications pervasive, robust, adaptive, and scalable.

The area of convergence seeks to unify distributed IT services and infrastructure resources (bandwidth, compute, storage) within and outside the physical enterprise towards a utility-based system capable of pooling or segregating IT resources and services when needed, where needed, and in whatever forms necessary to improve performance, resiliency, and security.

A key effort in this space is our evaluation of PlanetLab and PlanetLab-developed technologies [12] for use by enterprises and within enterprises. We are also researching new ways of looking at IT systems using a "Cognitive Peers" approach [13]. This is a model for communication and data transfer and manipulation that has cognitive components.

Trustworthy and Survivable Systems

Computer networks and computing infrastructure have evolved from a static, one-dimensional, and physically connected model to a dynamic, multidimensional, and virtual model. This evolution, along with the growing threat of more sophisticated and dynamically adjusting exploits and attacks, has resulted in significant challenges to the ability of the IT infrastructure to provide essential services in the presence of attacks and failures as well as to recover to full services in a timely manner (survivability), and to the ability for network computing systems to provide users with services as intended without compromising service integrity, user privacy, and data confidentiality (trustworthiness). Our research focuses on system research, development, and use-case expansion to adaptive, real-time, and resilient security systems enabled by new technologies, services, and methods targeted at improving the survivability and trustworthiness of the IT infrastructure at high operational capacity and minimal TCO. We are collaborating and continuously exploring new joint research opportunities with both Intel and academic researchers in several areas including policy-enabled security management, correlated intrusion detection and response in large-scale networks, content (business content, privacy information, and intellectual property) protection, identity management, and trusted computer platforms. Among these topics, policy-enabled network security management [14][15] is an adaptive approach that enables a common security policy specification across a heterogeneous enterprise network, and allows correlation of abnormal events so that policies are dynamically linked to the threat environment. Correlated intrusion detection and response in large-scale networks allows faster and more accurate (lower false positive) detections of intrusions and attacks through distributed monitoring and intelligent correlation of intrusion data from distributed sources. For content protection, we are exploring emerging technologies such as Digital Rights Management (DRM), which refers to the access control and protection of digital contents distributed on-line. We focus our research on Enterprise Rights Management (ERM), that is, protection of digital content in an enterprise network environment through persistent usage policies that remain with the content throughout the lifecycle of the content. Our plan for identity management and trusted computing platforms is targeted at realizing trust across autonomous domains through standards and technologies such as federated identity management, manageable identities, and Trusted Platform Modules (TPM) [16]

DATA FUSION

Current data use is often traditional: it is assumed that users are familiar with the available data and know what they need to access. Both these assumptions break down when dealing with large data repositories, and these large repositories are becoming more common. Enterprise data volumes in corporations are doubling every 12-18 months, with data in some segments growing even faster [17]. With the increasingly large storage comes a dramatic increase in storage of rich media formats and data quality issues. To deal with the challenges emerging from the exponential growth in data, research is focusing on data and information systems: in other words, how to facilitate user access to interesting information that will provide users with a competitive business edge. This is a departure from existing expectations that users will sift through the data themselves and find these patterns using manual analysis. Since manual techniques do not scale up with the amount of data, new methodologies need to be introduced as the amount of data we deal with in the enterprise increases.

Using an analogy of nuclear fusion, we introduce a research focus area we refer to as "data fusion." In this analogy, the "lighter" elements consist of data. The data can be in different formats: binary, text, audio, video, biometric etc. The data often reside in different repositories, both structured and unstructured. These data repositories are rarely integrated, physically or logically. The analogy to the nuclear fusion reaction is the processing, sometimes referred to as data-mining, that transforms the data into something more than data, into "heavier" elements. The output of the reaction is interesting (new, actionable, etc.) information, (interesting being defined in the business context) that provides a competitive edge.

Facilitating this process presents several challenges, all stemming from the *size* of the data:

1. Integrating data from disjointed repositories.
2. Integrating different data formats.
3. Gaining a competitive edge for the organization through discovery and presentation of interesting (in the business context) patterns.
4. Tracking the data life cycle from the entry of the data into the environment, through its transformations and interactions with other elements to its end-of-life.
5. Preserving privacy and security.

The first three challenges are aligned with the data federation, heterogeneity, and intelligence dimensions as outlined in [18]. Data federation refers to the distribution of data across different databases; data heterogeneity refers to the diversity in data formats, including unstructured data formats, and intelligence refers to the transformation and analysis of data to enable and support better business decision making.

ADAPTIVE, INTEGRATED BUSINESS PROCESSES

Our adaptive, integrated business processes research focuses on realizing the benefits of IT investments past and present while exploring how future processes, technologies, and systems will meet the need of our company and industry at large. Specifically, we strive to promote efficiency, flexibility, and performance in Intel's business planning and supply network processes, enable improved interaction and efficiency between Intel and other companies in our value chain, define business requirements for future IT architectures and applications, and create a sustainable competitive advantage by developing business processes around emerging IT capabilities [19][20]. We must comprehend the collaboration models and the architectural elements necessary to enable enterprises to expose their business processes through Service-Oriented Architectures (SOAs).

Advancing Existing Business Processes

Business has trended toward increasing globalization and product and service customization over the past decades, and the complexity of business planning and supply network management has increased commensurately. Widespread adoption of Enterprise Resource Planning (ERP) systems has enabled businesses to handle greater scale and complexity, yet approaching full potential in terms of performance and productivity will require organizational processes and supporting applications to evolve substantially over the coming years, even decades. Sizable investments in IT systems across all industries are slowly being transformed into more complete solutions and they are beginning to provide greater benefit as companies large and small rework organizational structures, management processes, and value-chain partnerships.

We look at today's processes in the context of case studies, learning how functional teams use information, policies, and tools to manage Intel's business. These case studies seek to understand how global, bottom-line performance is affected by such factors as information systems and analytical tools, local policies and incentives, and the flow of data and information through multiple decision points. We work with partner researchers in academia and industry to analyze the results of these studies using various modeling and simulation techniques and then develop innovative approaches and methods to drive improvement. Sample research topics derived from our case studies include using market mechanisms to replace traditional hierarchical planning systems, replacing point estimates with range forecasts for demand and supply planning, developing new models to forecast product transitions, and using new communication and planning tools to align product marketing and supply strategies. In all cases, we look to combinations of processes and IT solutions to provide better business outcomes while reducing the required human capital.

Developing New Processes Based on Proactive Infrastructure

Hyper-competitive global markets, rapid technology advances, and increased customer demands squeeze product lifecycles and market windows. To succeed, a company must be able to rapidly introduce new products at reduced cost and coordinate an efficient and effective response to various elements of risks across highly extended value chains with geographically dispersed internal teams and external partners. In response to these business drivers, we have expanded our research focus to

address those issues that will help deliver on adaptability and flexibility across the value chain while leveraging the proactive infrastructure.

For the most part Value Chain Integration (VCI) isn't about technology; it's about implementing business processes across an expanded enterprise. It involves bringing business relevance to the implementation of emerging Internet technologies with e-processes and evolving to an SOA. VCI aggregates business applications and business processes to a higher level of abstraction. The benefit is that it brings e-business and e-process relevance to a complex IT landscape composed of ongoing, flexible adaptation. VCI enables coordination across departmental, organizational, and enterprise boundaries from an overall business-level perspective.

This research agenda will build on our successful application of modeling and simulation to enable a universal inter-enterprise "vocabulary" to simplify the specification and sharing of business processes across the value chain and to accelerate instantiation of business solutions in the new business network. This work involves mapping processes and information flows in the value chain to support rapid/effective decision making with new usage models to synchronize and strategically optimize business processes. The goal is to align IT architecture with next-generation business processes in order to realize a competitive advantage within a proactive infrastructure.

Our research agenda includes building on Intel's leadership in driving RosettaNet by establishing the business process context for collaboration across the value chain. Besides driving a common process vocabulary, this agenda will entail architectural guidance for support of those collaboration processes in a federation of enterprises while maintaining context even with disparate interaction protocols.

Design New Processes in Response to New Technologies

New technologies will enable entirely new approaches to managing and optimizing supply networks and value-chain partnerships. By exploring new technologies and potential uses and then designing new processes to bridge the technologies to solutions, we will fundamentally alter business planning and execution. For instance, research in smart objects and RFID capabilities is demonstrating a number of new benefits in the areas of inventory management, production operations, quality control, and distribution. While most research on new technologies focuses on the technology itself, no business will be able to take advantage of any given technology until solution stacks based on new processes are developed. We are driving toward definition of these processes so that solution-stack development may follow.

COLLABORATIVE AND INTUITIVE COMPUTING APPLICATIONS

The top layer of the proactive enterprise capability stack is the applications layer with an emphasis on what we call Collaborative Computing and Intuitive Computing. Research in both these areas builds upon the three lower layers in the stack and investigates how technology can make people and enterprises more effective. Collaborative Computing looks at how technology can mediate collaboration while Intuitive Computing strives to support individual productivity.

Collaborative Computing

Large enterprises such as Intel's are faced with an increasingly geographically dispersed and increasingly mobile workforce across widely varying time zones, cultures, languages, and values. Today, we use information systems to improve the productivity of individuals or to automate tasks. However, these mainstream information systems do little to improve the ability of groups of people to work together on collective tasks such as collaborative problem analysis, idea synthesis, decision making, design, conflict resolution, and planning. Team productivity and performance has the potential to yield exponential results due to synergistic factors, knowledge creation and construction, diverse perspectives, and coping with complexity.

In the Collaborative Computing research focus area, we seek to invent new ways in which technology can mediate inter-team and cross-team collaboration.

More specifically, we are exploring emerging usage models such as asynchronous meetings; visibly engaging interfaces; use of visualization and expressivity; mobile collaboration; cross-team visibility and productivity; smart team and personal workspaces; and dynamic context, driven from business processes, events, and interests or preferences.

Our approach includes the following types of research activities:

- Surveying Intel's "virtuality" on five dimensions: time, space, business unit, media, and culture.
- Creating a baseline of related external research and current collaboration tool use at Intel.
- Identifying the "desired user experience."
- Designing and prototyping user-oriented solution concepts for Intel and similar organizations to explore emerging usage models.
- Defining an SOA for team collaboration.

- Specifying enabling IT infrastructure and platform dependencies.

Our baseline and survey work has validated the first-hand observations of team members and led to a definition of the desired user experience. By starting with the desired user experience as a goal, we avoided the trap of thinking in terms of predefined capabilities. Instead we worked from the experience to identify supporting capabilities. This led to the definition of an architecture that also included core team collaboration processes such as meeting and team management. We also designed a concept prototype to illustrate the desired user experience.

Team collaboration research and architectural concepts are explored further in [21].

Intuitive Computing

One of the principle objectives of any IT organization is to support and enhance the individual productivity of each employee. In addition to the traditional productivity applications such as word processing and spreadsheets there are numerous other opportunities in the proactive enterprise to support individual productivity. We review three here that we think align particularly well with the proactive enterprise.

As systems become increasingly complex, individuals find themselves spending more time diagnosing and correcting problems. This is especially true for managers, engineers, and technicians who oversee operations in networking infrastructure, IT managed server farms, communications systems, and manufacturing facilities, for example. It is also becoming increasingly so for other employees who will naturally first try to diagnose their own problems before turning to IT call centers and support Web sites. Enhancing the diagnostic tools of the former group improves operational efficiency, and providing better diagnostic methods for the latter group reduces the amount of lost work time and reduces call center costs.

For operations diagnosis, one promising approach is to model component failures and diagnostic processes using Bayes nets. Not only does this capture the knowledge of expert troubleshooters in a computational formalism, it streamlines diagnostics by presenting ranked lists of most likely component failures given evidence provided so far and recommends which diagnostic test would be the most discriminating if performed next. More on this approach is reported in another paper in this issue of the Intel Technology Journal [22].

For aiding general employees in their troubleshooting efforts, it is important to make it easier for employees to find the information they need to solve their own problems by, for example, allowing users to pose questions to an information repository using free-text. In [23], conditional probabilities are calculated through training sets to relate descriptive words in free-text queries to various troubleshooting procedures. When a user poses a query, key words are extracted and then a search is performed to maximize the probability that a troubleshooting procedure addresses the problem posed by the user. A list of the most suitable trouble shooting procedures is then presented to the user.

Another challenge is supporting productivity for mobile computing users. An interesting example is in Intel's semiconductor fabrication facilities, often called fabs for short, where engineers and technicians wear clean room gowns, complete with hoods and two layers of gloves. Because the work in these facilities typically requires interaction with machinery, traditional desktop or even notebook computer usage models don't function very well. Trials have taken place in Intel's fabs using Personal Digital Assistant (PDA) form-factors [24] which showed promise for enhancing communications among factory staff, reducing errors, lowering costs, and increasing productivity. Ethnographic and time-in-motion studies of Intel fab workers indicate that there are additional challenges to using PDAs in Intel fabs because of hands-free requirements and awkwardness of using a stylus with two layers of gloves. We are exploring the role of speech technologies to address these challenges. We have developed a prototype to augment a factory tool graphical user interface with speech commands and are looking into speech interfaces for preventive maintenance procedures, parts ordering, and note dictation.

The themes illustrated in these examples go to the heart of the Proactive Enterprise. We are striving to harness the recent advances in machine intelligence to make applications do more on behalf of the user as well as allow users to interact with these systems in natural intuitive ways.

SUMMARY AND CHALLENGES

The Proactive Enterprise covers multiple research disciplines and addresses business applications, IT services, and operational issues. The challenge for IT researchers or developers is to bridge the chasm between academic research and industry IT development in these areas. The demand for a Proactive Enterprise will only increase as IT organizations are required to deliver more complex applications, infrastructure systems, and IT services to workforces that are increasingly mobile and dispersed around the globe.

ACKNOWLEDGMENTS

The authors wish to thank Gene Meieran, Martin Curley, and Patrick Holmes for improving this paper through their insights and thorough feedback.

REFERENCES

[1] *Intel Information Technology Annual Performance Report*, Intel Corporation, 2003, http://www.intel.com/IT.

[2] *Proactive Computing*, Intel Research, http://www.intel.com/research/documents/proactivepdf.pdf.

[3] Tennenhouse, D., "Proactive Computing," *Communications of the ACM*, Volume 43, Issue 5 (May 2000), pp. 43-50, http://portal.acm.org/citation.cfm?id=332837*.

[4] Vicente, J., Hutchison, D., "7th IFIP/IEEE International Conference MMNS 2004," Management of Multimedia Networks and Services, October 2004, https://www.itsharenet.org/MMNS/*.

[5] Li, H., Vicente, J., Robak, R., "Automated Network Management" (internal) *Intel Technical Report*, July 2003.

[6] Dong, X., Hariri, S., Xue, L., Chen, C., Zhang, M., Rao, S., "AUTONOMIA: An Autonomic Computing Environment," in *Proceedings of the Performance, Computing, and Communications Conference*, April 2003, pp. 61-68.

[7] Sedayao, J. et al, "PlanetLab and its Applicability to the Proactive Enterprise," *Intel Technology Journal, Volume 8, Issue 4*, 2004.

[8] Kozuch, M. et al, "Enterprise Client Management with Internet Suspend/Resume," Intel Technology Journal, Volume 8, Issue 4, 2004.

[9] Krishnaswamy, D., Vicente, J., "Scalable Adaptive Wireless Multimedia in the Proactive Enterprise," Intel Technology Journal, Volume 8, Issue 4, 2004.

[10] Vicente, J., "Converged, Mesh Networks," (internal) *Information Technology Research presentation*, October 2004.

[11] Vicente, J., "Integrated Network Strategy and Architecture," (internal) *Intel Technical Report*, January 2003.

[12] L. Peterson, T. Anderson, D. Culler, and T. Roscoe, "A Blueprint for Introducing Disruptive Technology into the Internet," in *Proceedings of HotNets I*, Princeton, NJ, October 2002.

[13] R.H. Wouhaybi, J.B. Vicente, and A.T. Campbell, "Cognitive Peers," Under submission, August 2004.

[14] Hong Li, Ravi Sahita, Greg Kime, Jac Noel, and Satyendra Yadav, "Policy-Enabled Network Security with Adaptive Feedback Loop and Capability-Based Data Model," *Eurescom*, September 2003.

[15] S. Rungta, A. Raman, T. Kohlenberg, H. Li, M. Dave, and G. Kime, "Bringing Security Proactively Into the Enterprise," *Intel Technology Journal, Volume 8, Issue 4*, 2004.

[16] Trusted Platform Module: https://www.trustedcomputinggroup.org/downloads/specifications/*.

[17] Thangarathinam, T, Wyant, G., Gibson, J and Simpson, J. "Metadata Management: the Foundation for Enterprise Information Integration," *Intel Technology Journal*, Volume 8, Issue 4, 2004.

[18] Jhingran, A. D., Mattos, N. and Pirahesh H., "Information Integration: A research agenda," *IBM Systems Journal*, Vol. 41(4), 2002.

[19] Brown, G. W., Carpenter, R. E. "Successful Application of Service-Oriented Architecture Across the Enterprise and Beyond," *Intel Technology Journal, Volume 8, Issue 4*, 2004.

[20] Gillett, F. E., Schrek, G., Koetzle, L. and Fichera, R., "Organic IT 2004: Cut IT Costs, Speed up Business," 2004. Forrester Research. http://www.forrester.com/Research/Document/0,7211,34342,00.html*.

[21] Pickering, C., Wynn, E., "An Architecture and Business Process Framework for Team Collaboration," *Intel Technology Journal, Volume 8, Issue 4*, 2004.

[22] Agosta, J.M., Gardos, T.R., "Bayes Network "Smart" Diagnostics," *Intel Technology Journal, Volume 8, Issue 4, 2004.*

[23] Heckerman, D., Horvitz, E., "Inferring Informational Goals from Free-Text Queries: A Bayesian Approach," in *Proceedings of the 14th Conference. Uncertainty in AI*, 1998.

[24] Wireless in the Factory: Improving Productivity with Wireless Mobile Devices, *Intel Information Technology Whitepaper*, July 2003, http://www.intel.com/business/bss/infrastructure/mobility/wireless_factory.htm

AUTHORS' BIOGRAPHIES

George W. Brown joined Intel in 1994 and is currently a senior program manager within the Information Services and Technology group. He focuses specifically on methods and tools to ensure Intel reaches its goals in supply-chain management by identifying opportunities to apply IT in innovative ways to solve business problems and improve Intel business processes. He is also the past chairman of the Supply Chain Council and has represented

Intel in external research and benchmarking activities as chair of the SCC Research Strategy Committee. His e-mail is george.w.brown at intel.com.

Thomas Gardos has been with Intel since 1993 and is currently a senior researcher in Intel's IT Research team. His main interests are multimodal interfaces for mobile computing applications, Bayesian network-based diagnostic systems and digital video coding. Tom has thirteen patents in these areas. He received M.S. and Ph.D. degrees in Electrical Engineering from the Georgia Institute of Technology in 1991 and 1993, respectively, and a B.S.E.E. degree from the University of Delaware in 1985. His e-mail is thomas.r.gardos at intel.com.

Jay Hopman has been with Intel IT since 1993. His research interests include product diffusion, demand and supply forecasting and planning, and using market mechanisms to improve traditional business processes. Jay graduated from Purdue University in 1992 with a B.S. degree in Computer and Electrical Engineering, and he received an MBA degree (concentration in Strategic Analysis) from the University of California, Davis in 2000. His e-mail is jay.hopman at intel.com.

Hong Li is a senior researcher with Intel's Information Services and Technology Group, responsible for trustworthy and survivable systems research. She led the development of several IT security strategies and architectures. She is also active within the Intel and external research communities. She is a 2004 Santa Fe Institute Business Network Fellow. Hong holds a Ph.D. degree in Electrical Engineering from Penn State University. She is also a certified information systems security professional (CISSP). Her e-mail is hong.c.li at intel.com.

Sigal Louchheim leads the Data Fusion research focus area in the Information Services and Technology Group. Her main research interests are Interestingness (what is interesting) in Knowledge Discovery and Data Mining. Sigal received her Ph.D. degree in Computer Science in 2003, her M.Sc. degree in Computer Science in 1996, and her B.Sc degree in Mathematics and Computer Science in 1990. Her e-mail is sigal.louchheim at intel.com.

Cindy Pickering is an Information Technology principal engineer in Intel's IT Research Group. She focuses on global team collaboration, including people, processes, and technology. Her research combines needs assessment, new concept development, emerging usage models definition, and enabling architecture and technologies exploration. Cindy received a B.S.E.E. degree from Penn State University in 1981. She joined Intel in 1991 and has 23 years of industry experience. Her e-mail is cynthia.k.pickering at intel.com.

Jeff Sedayao is a staff engineer in the Planetary Services Strategic Research Project and in Intel's Information Services and Technology Group. He focuses on applying PlanetLab and PlanetLab-developed technologies to enterprise IT problems. Jeff has participated in IETF working groups, published papers on policy, network measurement, network and system administration and authored the O'Reilly and Associates book, *Cisco IOS Access Lists*. His e-mail is jeff.sedayao at intel.com.

John Vicente, an Intel principal engineer, is the director of Information Technology Research and chair of the IT Research Subcommittee. John joined Intel in 1993 and has 19 years of experience spanning R&D, architecture, and engineering in the field of IT networking and distributed systems. John has co-authored numerous publications in the field of networking and has patent applications filed in internetworking and software systems. He is currently a Ph.D. Candidate at Columbia University's COMET Group in New York City. John received his M.S.E.E. degree from the University of Southern California, Los Angeles, CA in 1991 and his B.S.E.E. degree from Northeastern University, Boston, MA in 1986. His e-mail is john.vicente at intel.com.

THIS PAGE INTENTIONALLY LEFT BLANK

PlanetLab and its Applicability to the Proactive Enterprise

Jeff Sedayao, Corporate Technology Group, Intel Corporation
John Vicente, Information Services and Technology Group, Intel Corporation
Rita Wouhaybi, Information Services and Technology Group, Intel Corporation
Hong Li, Information Services and Technology Group, Intel Corporation
Manish Dave, Information Services and Technology Group, Intel Corporation
Sanjay Rungta, Information Services and Technology Group, Intel Corporation
Stacy Purcell, Information Services and Technology Group, Intel Corporation

Index words: distributed systems, firewalls, PlanetLab, security

ABSTRACT

Current Information Technology (IT) models for security, content distribution, network and systems management, and provisioning do not match the level of complexity and flexibility required by today's enterprise. Private enterprises increasingly need to deal with internal security threats the same way as they deal with external security threats. Increasing mobility and globalization demand that applications be pervasive, i.e., working wherever an enterprise's employees operate. At the same time, the cost of maintaining applications, including deploying them and provisioning the proper resources, must be minimized, and enterprises need to be able to manage and deploy new services without impacting the existing operational services. In this paper, we suggest that PlanetLab [1], currently a world-wide research platform, is a potential IT delivery vehicle that can solve many of the challenges facing IT organizations, and it has the potential to become a commercially viable platform for enterprise organizations.

INTRODUCTION

As Information Technology (IT) can provide a significant advantage to a corporation's ability to compete in the growing Internet market, the speed to provision or configure network services (e.g., security enhancements, quality of service (QoS) or virtual private network (VPN) services) is as critical to an IT organization as bandwidth provisioning speed. This service flexibility can introduce a higher degree of network automation and control, allowing the administration and deployment of network services to be accelerated and operational procedures (e.g., change management and service-level management) to be reduced or automated. Additionally, the deployment of emerging Internet services has introduced a growing number of network devices (e.g., middle boxes [2]) along with an increase in vertical integration of traditional router or network device functionality. These devices increase already rising depreciation costs and necessitate the employment of more operations staff who must integrate these services into the operational environment. The need to deploy such services into the existing infrastructure ad hoc and without adding new devices or operating system upgrades is aimed at reducing the depreciation costs and enabling more rapid service provisioning. As a result, IT network infrastructure and supporting operations can be designed for agility, flexibility, and lower overall cost.

The Internet infrastructure was originally designed for simplicity, connectivity, performance, and reliable packet delivery. The "stupid network" [3] has been generally associated with a lack of control and transport intelligence within traditional network devices that performed primarily fast packet processing or route forwarding. While these principles should continue to drive the core transport design for packet delivery efficiency, the Internet has had to evolve a more intelligent system because of customer demand. This resulted in multi-service and overlay networks that have introduced enhanced services such as content-delivery, QoS, and caching into network appliances or application-aware devices. Despite these innovations, the vision of an Internet, where underlying network services and transport infrastructure support reliable and responsive delivery, innovation, and service flexibility, will require a fundamental change in the Internet service delivery philosophy. In today's fast-moving, Internet economy, the IT infrastructure needs to be able to deal quickly with ever-changing requirements and demands. Completion of IT service and project requests once took weeks, going

through a lengthy evaluation, design, implementation, test, and deployment cycle. Lengthy implementation times such as that are no longer acceptable. Today, the IT infrastructure needs to be proactive, anticipating requirements and delivering service whenever and wherever they are needed.

PlanetLab [1] is a planetary-scale, distributed research testbed spanning many networks and covering much of the world. Today, researchers throughout the world can use PlanetLab concurrently. They enjoy tremendous flexibility in what they can implement and experiment with, and they can isolate their particular applications or experiments. PlanetLab offers a global presence, distributed virtualization, overlay capabilities, and an open platform to enable innovative and rapid service delivery. In this paper, we compare PlanetLab technologies to traditional technologies. We discuss PlanetLab's relevance to IT and discuss how it can evolve to make it even more amenable to enterprise managers. We show how PlanetLab and its core architecture can solve IT problems more effectively, while enabling innovative services for the proactive infrastructure. In the first section, we discuss some of the key challenges facing enterprises. In the next section, we describe PlanetLab and some of the technology developed and implemented on it. We then show how those technologies and capabilities can solve major enterprise IT challenges and bring the vision of a proactive enterprise closer to reality. We finish by discussing our planned research and suggest how PlanetLab itself could be made more useful to enterprises.

PROACTIVE INFRASTUCTURE

While the proliferation of emerging network services continues to force the evolution of enterprise environments away from the "stupid network," several problems arise from today's current service provisioning model. First, the service provisioning granularity is static, coarse, and time-consuming. In the typical scenario, the introduction of new services into an enterprise or service provider environment is on the order of a network device or operating system upgrade or new device deployment. These new services generally reduce the useful life-span of a network device and translate into a higher depreciation cost model over time. Moreover, the provisioning cycles are on the order of months and years, and generally require new personnel and procedures to introduce the capability into the operational environment. This puts a strain on the IT organization or the service provider to adapt quickly to the changing business models and meet evolving customer needs. Second, as new services are introduced, the network management complexity and organizational burden (e.g., personnel training or operational overhead) are abrupt and magnified, rather than incremental and transitional. The reason for this relates (mostly) to our first issue, since new services are introduced as another layer of network infrastructure rather than as an evolving service component of the existing infrastructure. Third, there exists little integration and plug-n-play interoperability between network devices and/or alternative vendor solutions. Lastly, today's networks are becoming application-aware, but they are not designed to interface with the application developer or higher-layer application services to increase the end-to-end visibility. The issue here lies in the fact that traditional network devices are not designed to feed back or maintain the network state in a manner that is suitable to the application's specific needs.

To support a new Internet infrastructure paradigm, we detail three areas where the IT infrastructure needs to be more proactive.

Security. Traditional perimeter security, which strictly protects and isolates services and resources, is no longer sufficient to meet enterprise needs. Today's business environment calls for more frequent and intensive interactions between external and internal systems, rapidly growing usage of wireless and mobile devices by employees, and increased need for controlled sharing of information, services, and resources between business partners. The closed network control model has disadvantages in protecting the enterprise networks from distributed network attacks because of data inaccuracy, inability to perform overall impact analysis, and lack of data correlation from distributed sources in large networks. As more and more enterprises move towards relatively "open" perimeters (sometimes without realizing it as through unauthorized wireless access points and VPN connectivity) and distributed network environments in order to meet business demands, the associated provisioning and management cost will consequently increase, as will the complexity. The IT infrastructure needs to be able to provision security requests quickly and be pre-positioned and ready for such requests. The potential threat increases as enterprise networks become more and more like the open Internet. When there are security problems, the infrastructure needs to be able to track incidents before people get involved, when it becomes too late.

Planetary-Scale Services. The global nature of business demands that IT infrastructure scales globally. The quality of IT services needs to be uniformly good, whether enterprise users are in Asia, Europe, or North America. Moreover, users need to access content and data from different parts of the world with predictable and usable performance. To make that happen, applications and infrastructure need to be sufficiently adaptable and pervasive in order to react to the varying conditions of users. At the same time, a chief concern of a Chief

Information Officer (CIO) is maintaining a reasonable Total Cost of Ownership (TCO) for IT services. A global service must not only be pervasive and adaptable, it must also be cost effective.

Service Innovation and Provisioning. While an enterprise IT organization needs to react dynamically to changing loads and demands on its service, it must also have the ability to add new and innovative services. In many IT organizations, making changes to anything in the infrastructure is extremely difficult. This difficulty flows from the fact that so many IT services are mission critical, and a change-associated outage can have an impact ranging in the millions of dollars. This difficult and slow change causes software and infrastructure to become ossified and grow brittle. While the goal is to avoid outages, the consequences of such ossification can be severe:

- Software versions used can be so far behind that they cease to be supported by vendors.
- The time to propose, approve, and implement changes can stretch into months.
- The number of days where changes can be performed steadily shrinks.

More severe downtimes can result as versions and infrastructure age and fail. The IT infrastructure needs to anticipate the need for new applications in such a way that old applications are not affected by changes.

THE PLANETLAB VISION AND IMPLEMENTATION

PlanetLab was born out of the demands of professionals in the field of distributed system research, who wanted to research global-scale distributed services, without investing in separate testbeds. While many of the researchers had ideas for services and experiments that would work on a global scale, there was no truly global testbed to try out and validate those ideas. Another challenge was that the Internet has become so important for every day use that it was impossible to directly experiment with it. Intel joined with leading distributed system researchers and funded the first set of PlanetLab nodes spread across the world. PlanetLab was envisioned as a test ground for next-generation global-scale Internet services. It would be an overlay network–an application and service layer living over the Internet in the same way that the Internet was an overlay on top of the global telephony infrastructure. As the Internet becomes more ossified and harder to change, PlanetLab's design allows its nodes to host new services without having to get rid of or affect other existing services.

PlanetLab has evolved to become many things. It is a network and server infrastructure for testing global-scale services and experiments. It is a consortium of universities, corporations, and research institutions that run and make available a global testbed. It is also a set of technologies and standards for running distributed applications, as well as a platform deploying those applications and services. Finally, it is also a way of driving innovation through the use of overlays and overlay technologies, as well as an open platform encouraging cooperation.

To serve the potentially large number of distributed researchers that would use the PlanetLab infrastructure, the implementers of PlanetLab created the abstraction known as a "slice" [4]. A slice is piece of the PlanetLab infrastructure given to researchers, experimenters, and service implementers to use. To a person implementing a service on PlanetLab, a slice is a set of virtual machines on some set of PlanetLab nodes defined by that person. Each virtual machine appears to be a complete Linux* machine with root access. PlanetLab's virtualization, through the vserver package [5], happens at the system call level and allows us to scale to up to 1000 virtual machines per node [6].

A number of technologies have been tested and deployed on PlanetLab and are particularly germane to our discussion. Distributed Hash Tables (DHT) such as Chord [7] allow data to be stored and retrieved from nodes in a network without a single point of failure and with bounded access times. Overlay routing allows intelligent, application-aware routing on top of the Internet. Distributed query systems such as SOPHIA [8] and PIER [9] enable the timely querying of data on an Internet scale. Sensor interfaces [10] are a generic lightweight method of encapsulating data and making them easily available to applications.

On top of those technologies a number of applications have been deployed on PlanetLab. CoDeen [11] is a content distribution Web-caching system aimed at accelerating the browsing experience of those using the Internet, and it has thousands of users. Coral [12] is a way of distributing Web content so that any Web site deployed using Coral can more easily cope with flash mobs. Oceanstore [13] is a distributed file system built on top of the DHTs. PHI [14] is a system, designed to run on the PIER distributed query system, for examining the health of the Internet. It has been used to track attacks and the spread of worms.

* Other brands and names are the property of their respective owners.

There are a number of systems similar to PlanetLab. Most notably, the Grid has been compared to PlanetLab [15]. PlanetLab differs from the Grid in that it is more network centric; the Grid is more compute-centric. Many PlanetLab applications, such as network measurement [16] and content distribution rely on geographic and network dispersal to be effective, while rarely being CPU intensive. Emulab [17] is a network testbed that addresses many of the same concerns as PlanetLab. Emulab uses the FreeBSD jail [18] to isolate experiments in a type of virtual machine. PlanetLab differs from Emulab in that PlanetLab emphasizes the development of services and APIs and also aims to be a deployment platform for services. Also, while Emulab nodes talk primarily to each other, PlanetLab nodes are encouraged to and often do communicate with non-PlanetLab nodes, an essential quality for incremental deployment of new services on the Internet.

PLANETLAB: TOWARDS THE PROACTIVE ENTERPRISE

Enterprise Usage Models

At this point, we have discussed some of the requirements of the proactive enterprise and the properties that the IT infrastructure needs to bring it about. We have also talked about PlanetLab–its vision and implementation. In this section, we provide examples of how PlanetLab and its technology can be used to solve problems within the enterprise and bring the vision of a proactive enterprise closer to reality. We describe several types of applications that we envision in the context of enterprises. Some of these have minor advantages or eliminate inconveniences, while others offer breakthroughs in the way enterprises deal with their technical infrastructure and investments. In all of the categories, we envision servers running PlanetLab kernels while clients can run "light" versions or agents in the background. Note that by servers we mean active agents in the network, including routers.

Content Distribution

With the wide availability of PlanetLab nodes and available technology for adaptive content distribution, the PlanetLab platform seems an excellent candidate for the deployment of services offering content. Some of the examples include delivering streaming services at the application layer that can be managed and delivered in an efficient way, such as End System Multicast [19]. The distribution of load among servers for almost any traditional existing service, such as Web, news, mail, can be done using PlanetLab. Services such as CoDeen [11] and Coral [12] are targeted toward content distribution, from the viewpoints of both the consumer of information and the distributor, and they could be used effectively by enterprises.

Enterprises might find it particularly useful to run file servers on nodes closer to their branches as well as locations frequented by their employees. In this scenario, a client at an enterprise typically uses a file repository to ensure regular backups and high availability of his/her data. However, if the user is offsite, he must download the data locally and synchronize them at a later date. This can be very inconvenient, particularly if a crash happens during synchronization. An alternative is to continue to access the file repository, which could incur unacceptable delays. If the file repository runs on top of PlanetLab, then when the user logs in from an off-site, the client machine will detect the closest PlanetLab node and use that as its local server. Once the user goes back to the office, the files will be transferred to the main file repository. Using this method, the files are continuously on the file server, making it quite easy to backup while the access speed is not sacrificed. Distributed storage systems such as Oceanstore [13] and LODN [20] have been implemented on top of PlanetLab and may have applicability within the enterprise.

Backup, Recovery, and Resilience

The enterprise network needs an extensive level of resilience built into the infrastructure to support the business-critical applications that are required to keep the business running. While this level of resilience has become a standard for most enterprise-class applications, it is expensive and often times involves redundancy in all layers of the computing model in addition to backup and disaster recovery network/server capacity. This redundancy ends up being unused most of the time. The model for resilience seen on PlanetLab creates highly available applications built on individual components that are not always available, and doing this over an overlay on top of the available infrastructure. This model for redundancy has the potential to be drastically cheaper then implementing full redundancy at all of the layers of the network and computing stack and for every different service independently.

In addition, it is difficult to justify the cost associated with providing a resilient and highly available network for applications that are not mission critical. For example, a mission-critical application would be bill-ship-close cycle for a manufacturing or service company, while a non-mission-critical application would be availability of certain services and data for field sales personnel in order to get their job done. For this type of application and service, IT operations and maintenance will typically disrupt use. PlanetLab can help in minimizing downtime by offering the same services seamlessly through other nodes.

Security

As we discussed above, enterprises are realizing that traditional perimeter security with strictly protected (isolated) services and resources is no longer sufficient to meet the business needs, as mobility of users increases. An open, global-scale platform like PlanetLab is able to support the above changes and address related concerns by providing the following capabilities:

- Distributed and managed security services for subscription or peer-to-peer-style sharing of security services, which significantly lowers the provisioning and management cost, especially for small businesses.
- More efficient network monitoring and intrusion detection with global-scale data sharing and correlation, which reduces both false positive and false negative rates, and facilitates faster response to attacks.
- A network of Secure Overlay Services (SOS) [21], i.e., a network architecture that protects critical applications from distributed denial of services (DDoS) attacks while allowing distributed, legitimate users access to targeted applications.
- An open platform that can enable federated security service provisioning and management, which requires establishment of trust among distributed and policy domains (e.g., among business partners).

We envision PlanetLab deployed tools like PHI [14] and network telescopes [22] could be used to monitor enterprises for worms, viruses, and other security events and proactively move to retard their spread.

System and Network Management

A global enterprise needs to monitor the global end user and application perspective. Monitoring of the systems and services, particularly those exposed to and accessed through the Internet (in DMZ segments, i.e., DeMilitarized Zones, as border segments between enterprises and the Internet are known), is too often performed by internal systems located within the enterprise network. With the exception of a few highly critical services and applications, there is no monitoring performance specifically from an external end-user standpoint. Many organizations do not proactively discover failures that can impact end users (who can be inside and outside of the enterprise). Organizations also do not consistently test/probe Internet DMZs and its systems to validate their configuration. Commercial service and application providers offer monitoring services for availability and performance. The cost of these services depends upon the frequency and number of locations originating the probe. Besides availability, performance- and application-response time are also sold as separate services, and there is limited flexibility for implementing more generic monitoring or responses.

Figure 1 shows how we can monitor DMZ segments, servers, and applications. PlanetLab systems can be used to monitor these items and provide availability and perspectives from several points in the world. In addition, the enterprise will have more flexibility in configuring the monitoring parameters on PlanetLab to match its respective needs and requirements.

A similar approach can be taken with monitoring systems within an organization. The ScriptRoute [16] network measurement system could be deployed within an organization. ScriptRoute allows generic network measurement and monitoring scripts and routes to be sent to measurement systems. It includes facilities for making sure that measurement activity does not generate traffic harmful to some router implementations and that this activity does not inadvertently trigger intrusion detection systems.

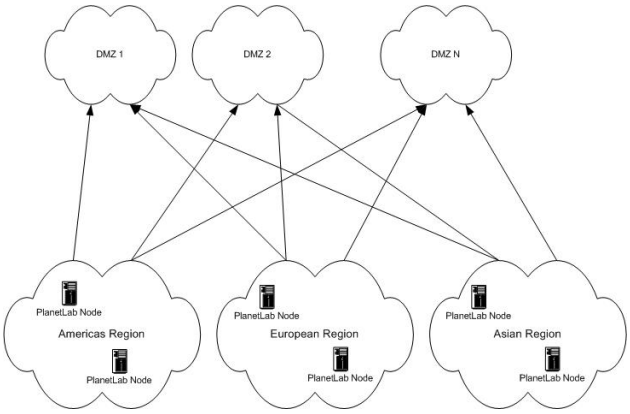

Figure 1: Monitoring DMZ segments from PlanetLab

Design Requirements of PlanetLab

In this paper we present a vision for using PlanetLab in the enterprise. We concentrate on the fields where PlanetLab can make a change or create a new environment that is suitable for the enterprise and thus offer a breakthrough. In all of these cases, we relied on the basic architecture and status of the current deployed PlanetLab. However, we also envision certain advances and changes for PlanetLab itself that we summarize here. Note that these changes are quite within the reach of PlanetLab, and the necessary techniques can be implemented and refined as the whole platform reaches maturity.

Better Slice Isolation

Even though the current implementation of the PlanetLab slice model [19] has proved to be extremely useful, it still lacks some additional features that would be essential in a corporate environment. The isolation should be more rigid since in some of the applications, we envision different

companies using different slices on the same machines. These companies might be competitors, and we would like to ensure privacy for each one of them. At a minimum, the slice mechanism should ensure that a slice cannot obtain or infer any information on another slice, including data present in any of the layers. We also envision encryption to be used on the data on a slice so that even administrators or personnel with physical access to the node cannot use their privileged position on the network to obtain any information. In addition, slices should be able to use ports even though they are reserved by another slice. This can be achieved with several techniques as explained in [24].

Limits and Credits for Slices

As more nodes join the network, the distribution of resources (e.g., computing, bandwidth) becomes more and more complicated. The current philosophy is to offer resources for slices as long as these exist. However, with academia as the main user space and audience, most nodes have low utilization with the exception of peak times ahead of publication deadlines. The usage model will be totally different when corporations join the network and use it actively. Due to the nature of their operation, corporations require some minimum guarantees. First, their applications should not be affected by fluctuating usage and should not suffer from any type of congestion. In other words, corporations expect a reasonable level of quality of service. Second, corporations require a mechanism that will ensure they get a fair share of the network as a whole, proportional to their level of participation as well as the amount of their investment. Corporations, in essence, tend to be much less forgiving than academia with respect to return on investment. For all of these reasons, PlanetLab as a platform should offer measures and indicators ensuring corporations that these conditions and guarantees are met and can be verified.

Management Interfaces

As we discussed in the previous sections, PlanetLab offers tremendous advantages for corporations. However, in its current state, the learning curve for implementing and/or deploying services on PlanetLab is quite steep. The community has contributed many tools and applications [25] to make the process friendlier, but the process is far from easy. We envision a common interface for managing all the resources and services running on top of PlanetLab. A common interface and platform has been an important but unachieved goal of administrators and vendors for quite some time. PlanetLab can facilitate reaching this goal and thus become an effective tool in the enterprise.

More Flexible Architecture

In the current architecture, authorization and authentication happens on PlanetLab Central. Different mechanisms and options should be available ranging from the current centralized state to a completely distributed mechanism. In some cases, corporate end-users can ask for authentication and authorization by proxy similar to the way roaming is implemented among ISPs or even wireless providers. In such a model, Enterprise *A* will create its local accounts on an internal central server that will manage its users while Enterprise *B* will manage its own set of accounts in parallel. Each of the two enterprises will have a master account that will act as the proxy for their respective accounts with the central server. This will ensure that account information and administration remains local to the enterprises. An example of such a mechanism is shown in Figure 2. Note that this is not the only possible solution, just a proposed example.

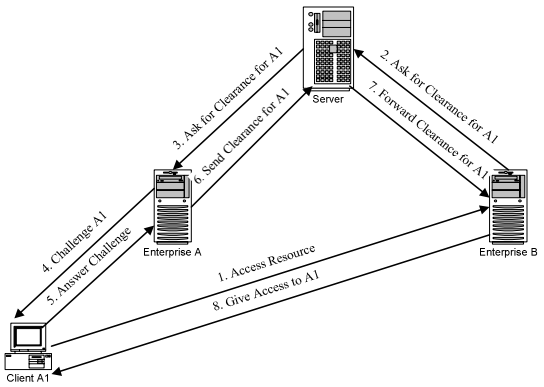

Figure 2: Authentication by proxy on PlanetLab

Our Research Agenda and Initial Results

One of the most difficult challenges to make PlanetLab useful to the enterprise is making PlanetLab services available to enterprise users. PlanetLab services are generally not available to most end-users within enterprises, and there has been a dearth of enterprise applications deployed on PlanetLab that have widespread use. One key reason for this situation is that PlanetLab applications assume direct network connectivity to PlanetLab nodes, an assumption that doesn't hold for most users within corporate environments. We are conducting research into how to make PlanetLab nodes available to users both inside of enterprises and on the public Internet. Our approach is to bring in PlanetLab nodes into the enterprise through various middlebox barriers such as NATs and firewalls. These PlanetLab nodes will have access to both the Internet and nodes inside an enterprise. We will study how PlanetLab applications have to adapt to this new split address space scheme.

Another approach of our research is to see how PlanetLab technology can be applied to IT problems. We are looking at how the external PlanetLab can be utilized as well as how the software and protocols built, tested, and deployed

on PlanetLab can be utilized by IT applications inside of an enterprise.

An additional research angle that we are considering is to create a PlanetLab that exists wholly internally within an enterprise and look at the issues necessary to making it work within the enterprise. We will look at which PlanetLab applications would be useful in this context. We also propose studying how PlanetLab applications change once they are within an enterprise, and how they may be provisioned externally.

Initial Results

We have had some initial results in using PlanetLab to monitor network and system connectivity. Since Intel uses Americas Region IP addresses outside of the Americas, we decided to study the phenomenon using the ScriptRoute [16] utility running on PlanetLab. We discovered that most of the deployed addresses had no problems, but for two networks, there were some routing anomalies associated with the address space choice. In addition, we have also developed a sensor interface for e-mail server performance data to be used for managing and monitoring those servers. We are currently developing a pilot implementation of this technology.

CONCLUSION

In this paper, we describe how enterprises in the global, highly competitive economy require a proactive IT infrastructure. We discuss the ideas, technology, and services of PlanetLab, and show how we can apply some of the technology and techniques of PlanetLab to realize our vision of the proactive enterprise. Our research agenda is to find ways to use PlanetLab to solve real-world IT problems and to remove obstacles to PlanetLab use by IT organizations.

ACKNOWLEDGMENTS

The authors wish to thank Robert Adams, Sally Hambridge, Robert Knauerhase, and Jim Brennan for their reviews and technical feedback to improve the quality of this paper.

REFERENCES

[1] L. Peterson, T. Anderson, D. Culler, and T. Roscoe, "A Blueprint for Introducing Disruptive Technology into the Internet," *Proceedings of HotNets I. Princeton*, NJ, October 2002.

[2] B. Carpenter and S. Brim, "Middleboxes: Taxonomy and Issues," *RFC 3234,* February 2002.

[3] D. Isenberg, "The Rise of the Stupid Network," *Computer Telephony*, August 1997, pp. 16-26.

[4] The PlanetLab Architecture Team, *Dynamic Slice Creation, The PlanetLab Architecture Team*, edited by Larry Peterson, October 2002. http://www.planet-lab.org/PDN/PDN-02-005/pdn-02-005.pdf*

[5] Linux VServers–http://www.linux-VServer.org*

[6] Andy Bavier, Mic Bowman, Brent Chun, Scott Karlin, Steve Muir, Larry Petersen, Timothy Roscoe, Tammo Spalink, Make Wawrzoniak, "Operation System Support for Planetary-Scale Network Service," *Proceedings of NSDI '04*: *First Symposium on Networked Systems Design and Implementation*, San Francisco, March 2004.

[7] I. Stoica, R. Morris, D. Kareger, F. Kaashoek, and H. Balakrishnan, "Chord: Scalable Peer-To-Peer lookupservice for internet applications," in *Proceedings of the 2001 ACM SIGCOMM Conference,* pp. 149-160, 2001.

[8] M. Wawrzoniak, L. Peterson, and T. Roscoe, "Sophia: An Information Plane for Network Systems," *PDN-03-014*, July 2003.

[9] R. Huebsch, J. Hellerstein, Nick Lanham, Boon Thau Loo, Scott Shenker, Ion Stoica, "Querying the Internet with PIER," in *Proceedings of the 29th VLDB Conference,* Berlin, Germany, 2003.

[10] T. Roscoe, L. Peterson, Scott Karlin, and M. Wawrzoniak, *PDN-03-010*, March 2003.

[11] L. Wang, K. Park, R. Pang, V. Pai, and L. Peterson, "Reliability and Security in the CoDeen Content Distribution Network," in *Proceedings of the USENIX 2004 Annual Technical Conference*, Boston, June 2004.

[12] M. Freedman, E. Freundenthal, and David Mazieres, "Democratizing Content Publication with Coral," in *Proceedings of NSDI '04: First Symposium on Networked Systems Design and Implementation*, San Francisco, March 2004.

[13] John Kubiatowicz, David Bindel, Yan Chen, Steven Czerwinski, Patrick Eaton, Dennis Geels, Ramakrishna Gummadi, Sean Rhea, Hakim Weatherspoon, Westley Weimer, Chris Wells, and Ben Zhao, "OceanStore: An Architecture for Global-Scale Persistent Storage," in *Proceedings of the Ninth international Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS 2000)*, November 2000.

[14] J. Hellerstein, P. Maniatis and T. Roscoe, "Design Considerations for Information Planes," Intel Research, *IRB-TR-04-017*, Aug. 30, 2004.

[15] Matei Ripeanu, Mic Bowman, Jeffrey Chase, Ian Foster, and Milan Milenkovic, "PDN-04-018: Comparing Globus and PlanetLab Resource Management Solutions," February 2004.

[16] N. Spring, D. Wetherall, and T. Anderson, "Scriptroute: A Public Internet Measurement Facility," *USENIX Symposium on Internet Technologies and Systems*, 2003.

[17] Brian White, Jay Lepreau, Leigh Stoller, Robert Ricci, Shashi Guruprasad, Mac Newbold, Mike Hibler, Chad Barb, and Abhijeet Joglekar, "An Integrated Experimental Environment for Distributed Systems and Network," in *Proceedings of the 5th Symposium on Operating Systems Design and Implementation*, Boston, MA, December 2002.

[18] P.H. Kamp and R.N.M Watson, "Jails: Confining the omnipotent root," in *Proceedings of the 2nd International SANME Conference*, May 2000.

[19] Y. Chu, A. Ganjam, T.S.E Ng, S. G. Rao, K. Sripanidkulchai, J. Zhan, and H. Zhang, "Early Experience with n Internet Broadcast System Based on Overlay Multicast," in *Proceedings of USENIX*, 2004.

[20] M. Beck, J. Gelas, D. Parr, J. Plank, S. Soltesz, "LoDN: Logistical Distribution Network," *Workshop on Advanced Collaborative Environments*, Nice, France, September 2004.

[21] Angelos D. Keromytis, Vishal Misra, and Dan Rubenstein, "SOS: Secure Overlay Services," *SIGCOMM'02,* August 2002.

[22] R. Pang, V. Yegneswaran, P. Barford, V. Paxson, L. Peterson, "Characteristics of Internet Background Radiation," to appear in *Proceedings of IMC,* October 2004.

[23] PlanetLab Implementation Team, *PlanetLab: Version 3.0*, August 2004.

[24] Jeff Sedayao and David Mazières "Port Use and Contention in PlanetLab," November 2003 (PDN ref).

[25] PlanetLab User Tools: https://wiki.planet-lab.org/bin/view/Planetlab/ContributedSoftware*

AUTHORS' BIOGRAPHIES

Jeff Sedayao is a staff engineer in the Planetary Services Strategic Research Project and in Intel's ITSG Research Group. He focuses on applying PlanetLab and PlanetLab-developed technologies to enterprise IT problems. Sedayao has participated in IETF working groups, published papers on policy, network measurement, network and system administration, and authored the O'Reilly and Associates book, *Cisco IOS Access Lists*. His e-mail address is jeff.sedayao at intel.com.

John Vicente, an Intel principal engineer, is the director of Information Technology Research and chair of the IT Research Subcommittee. John joined Intel in 1993 and has 19 years of experience spanning R&D, architecture, and engineering in the field of information technology, networking, and distributed systems. John has co-authored numerous publications in the field of networking and has patent applications filed in internetworking and software systems. He is currently a Ph.D. Candidate at Columbia University's COMET Group in New York City. John received his M.S.E.E. degree from the University of Southern California, Los Angeles, CA in 1991 and his B.S.E.E. degree from Northeastern University, Boston, MA in 1986. His e-mail address is john.vicente at intel.com.

Sanjay Rungta is a staff network engineer with Intel's Information Services and Technology Group. He received a B.S.E.E. degree from Western New England College and an M.S. degree from Purdue University in 1991 and 1993, respectively. He is lead architect and designer for the Local Area Networks for Intel. He has over 11 years of network engineering experience with three years of experience in Internet web hosting. His e-mail address is sanjay.rungta at intel.com.

Hong Li is a senior researcher with Intel's Information Services and Technology Group, responsible for trustworthy and survivable systems research. She led the development of several IT security strategies and architectures. She is also active within the Intel and external research communities. She is the 2004 Santa Fe Institute Business Network Fellow. Hong holds a Ph.D. degree in Electrical Engineering from Penn State University. Her e-mail address is hong.c.li at intel.com.

Rita H. Wouhaybi is a Ph.D. candidate at Columbia University. She holds an M.S. degree in Computer and Communications Engineering from the American University of Beirut. She is currently spending an internship at ISTG Research. Her research interests include peer-to-peer, overlays, network topologies, machine learning, and game theory, and she has publications in these fields. Her e-mail address is rita.h.wouhaybi at intel.com.

Manish Dave is a staff network engineer with Intel's Information Services and Technology Group. He is lead engineer and designer for the Internet Connectivity and external network connectivity for Intel. He has over ten years of network engineering experience and network security experience. His e-mail address is manish.dave at intel.com

Stacy Purcell graduated from the Georgia Institute of Technology with a BS-CS degree in 1992. He joined Intel Corporation in Folsom, CA immediately after graduating

where he has been employed in several roles including system administrator, network engineer, and manager over the course of the last 11 years. His e-mail is stacy.p.purcell at intel.com.

THIS PAGE INTENTIONALLY LEFT BLANK

Towards an Autonomic Framework: Self-Configuring Network Services and Developing Autonomic Applications

Brian Melcher, Information Services and Technology Group, Intel Corporation
Bradley Mitchell, Intel Communications Group, Intel Corporation

Index words: autonomics, Eclipse, network services, self-configurability, toolkit

ABSTRACT

Autonomic services in the enterprise are becoming more and more of a requirement in all types of networked environments. With an ever-increasing number, type, and complexity of network services available to individual computing systems comes an increasing complexity in establishing and maintaining the configurations of these services. Many network services are already self-configuring today, but this capability is not yet universally available for the broad spectrum of network services or networked environments. Without wide-reaching network services self-configurability, the benefits of reduced management complexity will remain unrealized.

We begin by examining existing network services configuration technologies and identifying incomplete or inconsistent capabilities for dynamically self-configuring these network services. We present for consideration the requirements of an architecture for dynamically self-configuring network services that drives enhanced yet simplified capabilities both to end users and to IT technicians and engineers in a corporate IT environment, as well as to roaming wireless users and home networks.

We then continue by examining the practical implementation of autonomic network service configuration. Purely autonomic systems cannot easily be built today due to a lack of comprehensive framework support. However, substantial pieces of autonomic technology exist in forms suitable for early adoption. Specifically, we focus on the IBM Autonomic Computing Toolkit*, an open set of Java*-class libraries, plug-ins, and tools created for the Eclipse development environment. The IBM Autonomic Computing Toolkit represents a modern framework for enterprise software integration. We examine the toolkit's standard interfaces and data formats to identify its applicability to network services configuration problems. We conclude with a summary of findings and recommendations for prospective enterprise developers and integrators of autonomic toolkits.

INTRODUCTION

Network services discovery is a significant aspect of today's network infrastructure. In today's network environments, network services configuration information is dispersed among a variety of information repositories, and the relationship between the storage and consumption of that configuration information is often managed through programs, procedures, or protocols specially developed for the specific environment at hand. As the number of end-user systems joining the network grows and the number and variety of network services grows, the complexity of and demand for a solution to manage this relationship between the configuration information and its consumption likewise grows.

The Need for Self-Configurability

Automating the management of network services configurations is becoming ever more a requirement across the entire spectrum of computing networks. The large networks in a corporate Information Technology (IT) or Internet Service Provider (ISP) environment are becoming too complex to manage configurations on an individual system-by-system basis. The sheer number of systems requires an increasing number of staff to manage them; proportional growth of staff-to-systems is undesirable, if not outright impractical. Additionally, individual hands-on system management increases the likelihood of errors being introduced into the environment. Mobile computing puts an increasing demand on reconfigurability as the system roams between

* Other brands and names are the property of their respective owners.

connectivity points. Home networks are becoming more commonplace as the number of computers in the home increases; yet, those same home networks must be easy to maintain in order for the typical home computing consumer to be able to manage them. In each of these environments, the end-user system has varied degrees of manageability control–that is, those who manage and control the network–and network services configurations may or may not have management control over the end-user system.

Self-configurability with respect to network services is the capability of a system to configure its own network-based services and applications in response to the needs of the user and the environment the system finds itself in. As the needs of the user change or the environment the system is in changes, that end-user system would recognize the change, understand the impact, and respond by reconfiguring itself accordingly. Flexibility of location, a wide range of administrative control, and the need for varied rates of dispersal of changed configuration information across the environment, all complicate the task of autonomic network services configuration behavior.

Structure of this Paper

In this paper we first review current-day network services configuration technology, identifying existing capabilities as well as incomplete or inconsistent capabilities in autonomic network services configuration behavior. We present for consideration an architecture that supports autonomic configuration of network services for a plethora of computing environments: a corporate IT environment, a mobile user hopping from one access point to another, and a home computing network whether or not it is connected to the Internet.

Second, we demonstrate development of autonomic applications using the IBM Autonomic Computing Toolkit in the Eclipse development environment. Created by the Eclipse Foundation, a consortium backed by Intel, IBM, and others, Eclipse provides a modern, extensible environment for software development. This toolkit specifically supports general-purpose problem determination, installation, and user access features that correspond to network service configuration tasks at a very high level. We examine details of the toolkit to develop practical recommendations for utilizing it in the IT, roaming user, and home network environment. Most recommendations apply generally to various other classes of autonomic problems.

NETWORK SERVICES CONSUMPTION PROBLEMS

Autonomic Network Services Solution Goals

We discuss the goals, considerations, and demands of self-configurability of network services in four types of network environments:

- A corporate IT environment.
- A wireless network provider with roaming mobiles.
- An Internet-connected home network.
- An isolated home Local Area Network (LAN).

We consider these four environments to span the range of network environments: these environments would push the limits of any solution for self-configurability of network services, and, therefore, any solution that applies to these networks would apply to network solutions in between.

In IT environments, one goal of self-configuring network services is to move away from individual system configuration management to policy management. This approach brings a higher level of abstraction to management by introducing a policy from which the configuration is derived, allowing the automation components of the infrastructure to apply these derived configurations to the individual systems across the environment. Policy management frees IT personnel from the role of sustain, maintain, and fix. Additionally, with the system itself deriving the configuration from a set of policies, this eliminates the human-error factor when configuring by hand.

More and more users are mobile, jumping from a wireless connectivity point to another as they travel from airport, to coffee shop, hotel, on-site at another corporation, or across a college campus. Eventually network services and connectivity for mobile computing should become as seamless as cell-phone usage going from one cell coverage area to another, or from one provider's region to another. In any mobile environment, user systems will come and go and network services will also come and go. As such, there can be no preconceptions or expectations by either party: the network and available services are foreign to the end-user system, and the end-user system is likewise foreign to the network and infrastructure.

In the home environment, multiple computers are becoming more commonplace. Home systems may be connected to just each other and otherwise isolated, or connected to the Internet. Either way, in this home environment usually the set of systems and services does not rapidly change, but there should be no expectations of establishing or sustaining a complex solution.

Usage Areas

The IT Environment

A corporate IT environment today consists of a mix of fixed and mobile, both wired and wireless, systems where IT personnel have end-to-end management control–that is, control of the configuration information repositories (the *back-end* systems), the system consuming said information (the *front-end* or *end-user* systems), and all the infrastructure in between. An additional characteristic is the technical expertise by the IT personnel for the entire breadth of the environment: the back-end, front-end, and infrastructure; and the standard images, sometimes referred to as *gold* images, deployed across the environment.

With this end-to-end system management (or, *tightly coupled* management) programs, procedures, and protocols specific to the environment can be put in place to manage the dispersal and consumption of any network services self-configuration information. In an IT environment where central administration of network services configuration is needed or desired, this manner of a gold image with pre-configuration or specialty-developed self-configuration utilities can more readily be introduced than it could be in other environments to completely address the problem. As a new service is established, a companion utility is developed and deployed in parallel with the service. Even with mobile users within the IT environment, such environment-specific utilities can still be employed to achieve network services self-configurability. Being mobile may require a more complex solution, but can be achieved.

In an IT environment, uses for network services self-configurability are varied: they range from day-to-day usage as mobile systems roam about the corporation (such as between buildings, sites, or sub-domains), to managing introduction of new services and retirement of existing services, to crises event management with distribution of anything from a security patch or anti-virus signatures data file, to a Web proxy configuration or mail relay blacklist across the entire environment.

Any solution applicable for an IT environment can require control of the end-user and/or back-end systems. It must also allow for policy management to drive individual end-user system configuration, and must allow for rapid policy and configuration changes to be introduced and dispersed across the environment.

The Roaming User

A roaming wireless user is in contrast to the tightly coupled end-user system in the IT environment. Here, the end-user system may not be under any degree of control by those administrating the infrastructure. Environment-specific utilities, as in the IT environment, could be introduced to provide self-configurability for that specific wireless neighborhood. However, it cannot be expected that this utility would be installed by a roaming wireless consumer. There is the initial question of trust and security of that utility: why would you trust a utility made available in a wireless hotspot in the middle of some airport or hotel. One question is whether it is a legitimate service or someone masquerading as a legitimate service in order to intercept the network connection. Further, and more applicable to the autonomic network services configuration problem here, as the user roams from one provider to another, the number of environment-specific utilities required on that mobile system increases, increasing the likelihood of conflicts or system instability.

With a roaming wireless user, the system being disconnected from the network must be considered. Any network discovery and self-configuration methods must recognize this disconnected state and allow for a stable and functional (as much as possible) computing system. Any solution must not require any degree of control on the end-user system, but can require establishment of back-end systems. A higher level policy management solution will greatly facilitate numerous systems coming and going, but the demand for an event-driven "push" of new configurations is less than with an IT environment.

Home Networks–Connected and Isolated

In the home network, both Internet-connected and an isolated LAN, there is not necessarily the IT know-how to create an environment-specific utility or even to deploy and configure a well-known solution. With computers now becoming ubiquitous in the home environment, the demand for autonomic network services configuration increases. In addition to the technical know-how, if a solution requires additional infrastructure it increases the physical cost of that solution. As such, an optimal solution would not require any additional infrastructure, services, or configuration. The solution should be transparent yet automatic, as if working magically with the existing collection of end-user class systems.

With the connected home network, there is a connection to the Internet at large through a service provider. The solution space for the connected home network is distinctive from the isolated home network in that it could be reliant upon the ISP to provide a certain level of autonomic network services configuration. The requirements for technical know-how and in-home infrastructure remain the same.

Solution Space Boundaries and Conditions

From the network usage areas described above we can create a series of checks by which we can evaluate the appropriateness of possible solutions:

- *No back-end infrastructure*. As indicated for the home network environment.
- *No technical know-how required*. Also as indicated for the home network environment.
- *No tightly coupled management*. For the wireless ISP and home networks.
- *Dynamic addition/removal of end-user systems*. For the IT environment and wireless service provider.
- *Dynamic addition/removal/change of services*. For the IT environment.
- *Universality*. Implemented by many OS and network equipment vendors.
- *Rapid change deployment*. For the IT environment.

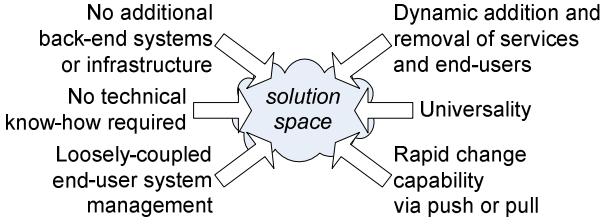

Figure 1: Network environment solution conditions

A wide range of network services should be considered for this problem of providing autonomic network services, for if the broadest scope is considered, a more robust solution will be forthcoming. These network services include hostname resolution, proxy (for FTP, Web, telnet), logging, Network Time Protocol (NTP), mail storage, mail relay, nearest printer, nearest patch distribution, and user authentication servers. Typically, what this configuration entails is a reference to another system, and then once the end-user system has been pointed to the service server, the end-user and server systems can communicate directly between each other.

The additional checks for varieties of network services should be as follows:

- *Nearest proxy server*. Can the nearest firewall/proxy server for FTP, Web and/or telnet be determined? This may be a list of systems, not just a single system, with each equally capable of providing service, but only one utilized for a transaction.
- *Nearest printer*. Can the nearest printer be determined? This is different than the *nearest proxy* check, as *nearest printer* is based upon geographic location whereas *nearest proxy* is based upon network location.
- *NTP server*. This is similar to the *nearest proxy server* and *mail relay server* checks: a list of equally capable systems is provided. This check differs in that all list members may be used for each transaction; however, only one is required. In the previous checks, only one system is used per transition.
- *Mail relay server*. Can the nearest mail relay server be located, a similar check as *nearest proxy*?
- *Mail server*. Can the mail server appropriate for this user be located? This condition is very different than the *mail relay server* check. A mail relay server is a system that receives and forwards e-mail (perhaps scanning for viruses or checking blacklists in the process); a mail server is the repository of a users e-mail inbox. Since a mail server provides a stateful service it requires locating the specific server with the appropriate state for the user, in this case, the state being the mail repository itself. The *mail relay* check differs in that it is stateless and readily interchangeable between peers.
- *Logging server*. This is much the same as the stateful *mail server* check in that a system may need to consistently report to the same logging server.
- *Nearest patch distribution server.* An authenticated patch distribution server must be located to provide quick and timely updates, particularly important in a tightly coupled environment.
- *Nearest user authentication server*. This is also for a tightly coupled environment, but for a stateful service in that the same server is employed time and again.

NETWORK SERVICES CONSUMPTION–STANDARDS, SOLUTIONS, AND CHALLENGES

Solution Methods

Self-configuration functionality can be implemented in a variety of methods. Functionality can be implemented locally, employing locally maintained policies, protocols, or programs wholly on the end-user system. Or, service configurations can be learned from the environment itself. These are the policies and protocols established in the infrastructure to facilitate the end-user system in learning, that is, updating itself, about the environment, both upon initial introduction to the environment as well as when the environment changes. Also, learning from the environment could come from peer systems already in the environment and successfully configured to operate within it. Learning from the environment can also stem from a global application of network resource configurations, specifications, policies, programs, procedures and/or protocols.

Intentional Hostname Naming Conventions

A long-standing approach to providing connectivity to network services, a pseudo-autoconfiguration method if you will, is the use of an intentional hostname naming convention. Services are provided by a system with a well-known hostname; for example, a mail relay is "mailhost" and a name service server is "ns."

This convention requires initial configuration on the end-user system to use this well-known hostname, but once configured, the act of reconfiguring network services occurs silently. And in actuality, on the end-user system there is no reconfiguration: the name resolution process provides the reconfiguration of the network service. Such a solution can be directly employed in IT and ISP environments with a managed back-end infrastructure.

With this convention the "short" hostname is used, not the Fully Qualified Domain Name (FQDN). For an IT environment using a gold image method of system installation, when that image is deployed across multiple domains, the service server name is resolved to the server within the local domain just like any other short hostname. A roaming mobile user would see the same effect when going from one service provider to another, providing there is consistency in the naming convention between service providers.

There are several shortcomings to this method. First, this convention is optional and therefore consistency between wireless ISPs cannot be guaranteed. Second, this does not provide an automated solution for a home network: it can facilitate a solution for an Internet-connected home network given the ISP's participation, but to the LAN internal to the home, this provides no solution.

Due to these shortcomings, this method would work best only in a tightly coupled environment such as an IT environment. But even here there are further shortcomings: this method doesn't meet several of our checks. Geographic information isn't available to address *nearest printer*. Further, *mail server* is not addressed due to how name resolution typically occurs: if there are multiple physical systems with the same hostname, name resolution may load-balance requests in a manner that returns the address of a different physical system for each name lookup query, rendering it useless for any stateful service that is distributed across a server farm.

DNS and DHCP

Some institution of autonomic network services comes from Domain Name System (DNS) and Dynamic Host Configuration Protocol (DHCP), both long-standing network name-service protocols. DNS provides hostname resolution through a distributed hierarchy of DNS servers. DHCP enables systems to dynamically configure their own IP address under the conditions and policies allowed by a DHCP server.

DNS

Use of DNS aliases is an instantiation of an intentional hostname naming convention and facilitates service portability. On the back-end the service proper can be relocated from one machine to another and, with the use of a DNS CNAME, end users are directed to the new service location as DNS is updated. This has commonly been done with hostnames such as "mail" and "www." Proposals recommend additional records to the DNS tables–beyond MX (mail exchange) to WKS (well-known services, which include FTP and www). However, this is still not seen as a long-term solution by even the Request for Comments (RFC) authors [3]. For the purposes of autonomic network services, we only need to communicate to the end-user systems within the immediate network. Additional DNS records would express this information externally as well, which may be an unacceptable information leakage; however, this can be remedied with a separate DNS zone available only internally. This internal zone increases the technical complexity and infrastructure required for the solution, but this separation of DNS zones is usually a requirement for IT and ISP environments segmented with firewalls from the Internet at large and should not limit its consideration as a solution for autonomic network services.

DHCP

DHCP provides dynamic network services through a software agent on the client and an infrastructure to service requests. DHCP is a well-established protocol and both client and server services are available on most operating systems today. Given this wide availability, DHCP could be well positioned to become the foundation for an autonomic network services configuration solution. Extensions to DHCP provide for additional network services, providing the DHCP client with DNS resolvers and a DNS search path [7]. Moreover, DHCP itself is an extensible protocol that could be utilized to provide pointers to additional network services configuration information.

While DHCP is an extensible protocol, these extensions only define packet content; Application Programming Interfaces (APIs) that utilize the additional DHCP-provided information are left undefined. This is a key reason DHCP is not a solution to the problem of autonomic network services. While this lack of APIs could be accommodated if the application itself implemented (at least partial) DHCP capability, this only transfers the problem and doesn't really resolve it. This method of introducing DHCP capability directly into the applications that utilize this DHCP-provided information

is a part of the environment-specific utilities discussed earlier, suitable only for tightly coupled environments.

Many of our solution space boundary checks can be met through careful policy construction, but this itself doesn't meet the *no technical know-how* check. With regard to *no back-end infrastructure* with the connected home network, DHCP is available in firewall/gateway appliances designed specifically for connecting up home networks, meaning this is not much of a limitation. However, the *dynamic addition and removal of services* check is not met with DHCP: an administrator must change the DHCP configuration policy when a service is added or removed.

Network Information Service (NIS/NIS+)

Network Information Service (NIS), and its security-enhanced follow-up NIS+, provide for database-like queries about information services. An NIS domain consists of a set of tables, and within those tables is a keyword-to-answer mapping. The NIS tables and mappings must be preconfigured, thus failing the *no technical know-how* check. Also, NIS does not lend itself to a dynamic environment of services coming and going, failing another check. Due to its one-to-one mapping nature, it also fails the *nearest printer* check, and it suffers the *mail server* problem in the same way as DNS. Further limiting NIS is that it is principally a UNIX*- and Linux*-only solution, failing the *universality* check.

NIS does have a unique characteristic. In DHCP, APIs do not exist to access new information types, but the APIs in NIS are generic enough to handle new information services: only the new table needs to be created.

Directory Service (LDAP)

Lightweight Directory Access Protocol (LDAP) [5] is a protocol for locating resources on a network through a hierarchical directory repository. Commonly used for authentication and as a replacement for NIS, LDAP and its information model are extensible, and as such can be considered for autonomic network services. As the information model is extensible, the type of information which can be included in the directory service can be used to meet practically all of the network services checks: network locality-based services of *nearest proxy server* and *mail relay server*; stateful services like *mail server* and *logging server*; and even geographical locality-based services of *nearest printer,* if the directory service information model contains the right data. LDAP includes APIs, allowing applications to interface with the information in the dynamic directory service. However, like DHCP and NIS, LDAP requires an infrastructure preconfigured to service requests, failing the *no technical know-how* and *no infrastructure* checks. Unlike NIS, LDAP can provide for dynamic responses and provides for service announcements, addressing the *mail server* and *logging server* checks. However, LDAP is still limited in that it does not provide for dynamic, spontaneous discovery, and as such does not lend itself as a solution.

Service Location Protocol (SLP)

Service Location Protocol (SLP) is a protocol that provides a framework for discovery and selection of network services, eliminating the need for many static network services configurations for network-based applications [4]. SLP is intended for an enterprise network with shared services (as opposed to a global network) fitting into the IT, ISP, and home environments.

With this protocol, end-user systems attempt to locate a service. The services themselves, as they come on-line, advertise their services and can communicate directly with end users or with a central directory agent. In this fashion it meets the *dynamic addition and removal of services* and *dynamic addition and removal of end-user systems* checks. APIs allow access to SLP data, and for non-SLP capable applications, SLP proxies can be employed. Like LDAP, SLP can provide dynamic responses and address many of the network services checks, again provided the directory agent is established with the right data.

DHCP options [6] and use of DNS Service Location Resource (SRV) records [10] have been proposed that would facilitate the discovery of the SLP directory agent, which would facilitate the use of SLP and increase its autonomic capabilities by reducing the *not tightly coupled* hurdle. SLP is a promising protocol and has been implemented in products from several vendors [15]. It does not, however, yet meet the *universality* check.

Limitations of Today's Solutions

Intentional hostname naming has existed for many years, and if it was a solution capable of providing autonomic network services for the broad spectrum of network services and networked environments, this would not even be a topic of continued research. The creation of DNS, and more specifically the MX and later the SRV records, facilitated service availability through intentional hostname naming. DHCP introduced dynamic hostname naming of end-user systems, but it is quite limited in facilitating network service discovery. DHCP, NIS, and LDAP all require an infrastructure and technical know-how to implement and sustain each services' own configuration. SLP also requires an infrastructure, but its implementation demands on technical know-how are less than those with LDAP; plus the demand can be reduced if

* Other brands and names are the property of their respective owners.

services register themselves, reducing the burden of activating the SLP Service Agent. SLP is not widely available: it is too soon to determine if an SLP-based solution is viable for autonomic network services configuration.

All these methods have limitations in one form or another, meaning that with any one implementation of these existing technologies and their implementation, a single universal architecture for autonomic network services has not yet arrived. However, a combined used of these protocols can provide a comprehensive solution. For instance, an end-user system using DHCP could acquire DNS name server information, and then with the information from DNS SRV records locate a directory service to subsequently locate configuration information for network services. Other combinations of protocols can also be applied to bridge the gaps that the participant applications have when standalone. These solutions certainly work, and, considerations of availability and reliability aside, they are complex architectures and, accordingly, do not meet the *no technical know-how* and *no back-end infrastructure* checks. Therefore, these protocol combinations are not beneficial to the unconnected home network. Further, even in the technical know-how rich IT and ISP environments, these solutions can more readily be employed where one protocol is built upon another, layer upon existing layer over time. To apply them to a new environment would incur high installation and maintenance costs.

TOWARDS AUTONOMIC NETWORK SERVICES

A Solution Architecture

To meet the *no back-end infrastructure* check suggests a solution in the direction of a peer-to-peer self-discovery mechanism. However, an IT environment requires a high-level policy control and rapid change capability, suggesting a client/server architecture. These two vectors seem contradictory, even mutually exclusive.

Consider a peer-to-peer self-discovery mechanism with a priority schema. By being peer-to-peer, there is no need for a back-end infrastructure, which satisfies the needs of a home environment. Then in the tightly coupled IT environment, we introduce a system we call a *services broker*, upon which policies are managed directly by IT personnel. In this priority schema, the services broker has a higher priority than the class of end-user systems: the services broker will "shout" while end-user systems "whisper." If so desired, the priority of end-user systems could be set to nearly zero (with priority defined as the higher the number, the higher the priority). Or, it could be set to zero (or "off"), in essence transitioning the original peer-to-peer architecture into a traditional client/server architecture–but all within the same solution implementation.

This is not, however, a traditional client/server architecture. Due to the priority-schema peer-to-peer structure this architecture has an ability for dynamics, fault tolerance, and self-healing built right into it. The outage of a services broker can be tolerated by end-user systems being able to utilize a configuration from any one of the collection of services brokers. In an "end user at zero" model, the end-user system cannot reference any of its peers, necessitating that end-user systems seek out other services brokers in the environment. In an "end user at near-zero" model, in an outage of all services brokers, any end-user system can listen to the "whisper" of another peer and discover what limited services may still be available.

In either a "zero" or "near-zero" end-user model, we gain an additional key benefit: the ability to rapidly introduce a new service or service configuration, a crucial feature to address crisis and denial-of-service situations. One can "seed" configuration solutions simply by introducing a system with a newer configuration (the "fix") into the environment. Even in a total services broker outage, an IT administrator can configure the new service on his or her own end-user class system and temporarily elevate the priority of that configuration, spreading the new configuration by "whispering" louder than other end-user systems.

Such a priority-based schema could be implemented onto an LDAP or SLP infrastructure, even maintaining compatibility with existing infrastructure deployments. SLP is considered a state-of-the-art directory services protocol and a strong choice for future management of network services [16], but not likely to displace existing LDAP infrastructures.

With these considerations we have presented the following as requirements for an architecture for autonomic network services:

- Priority-based peer-to-peer structure, allowing for a traditional peer-to-peer architecture that can transition into a client/server architecture.
- Service consumer self-discoverability of service providers and services brokers.
- Rapid introduction of new services and services configurations.

SOFTWARE TOOLKITS FOR AUTONOMIC NETWORK SERVICES

To fully address issues of autonomic network services, an autonomic framework is necessary. To reach this goal, "we need to rethink the structure of the system and the application software (and the tools that help build them)" [12].

While some features of autonomic services are best implemented in computer hardware, typical autonomic solutions for the enterprise also demand software support. Administration of traditional client/server and n-tier systems can benefit substantially from having autonomic services implemented in the operating system or higher-level software applications. A fully autonomic-aware Web database system, for example, should be capable of identifying and repairing certain classes of database integrity errors to minimize impact to future user sessions.

Unfortunately, to make enterprise systems fully autonomic requires significant effort. Modern enterprise architectures tend to feature numerous applications and components supplied by different vendors. These systems increasingly distribute the software, hardware, and data storage functions geographically across networks. Additionally, individual hardware and software components have grown exponentially in complexity over the past few years as processing speeds and development tools have improved.

Software toolkits deliver required autonomic capabilities utilizing standard APIs, data formats, and network protocols. A toolkit provides reusable, standards-based software to reduce these efforts of software development and integration. Examples of functions ideally supported in an autonomic software toolkit include the following:

- Event logging APIs and data formats.
- Issue alerting mechanisms.
- Network discovery mechanisms.
- Data migration and conversion utilities.
- Interfaces for extensibility and integration with third-party software.

In the following sections we explore functionalities of the IBM Autonomic Computing Toolkit. Though not a complete enterprise solution, this toolkit provides a practical framework and reference implementation for incorporating autonomic capabilities into software systems.

THE IBM AUTONOMIC COMPUTING TOOLKIT

The IBM Autonomic Computing Toolkit comprises class libraries, plug-ins, and tools for the Eclipse development environment. To support both development and execution, the toolkit depends on specific versions of the Java Runtime Environment* (JRE). In the following sections we describe each software component of the toolkit.

General Concepts

The toolkit is based on the dual concepts of Managed Resources and Autonomic Managers. Managed Resources can represent end-user computers, other network nodes, or individual software components running on a device. Resources monitor their environment and are capable of detecting and reporting events to an Autonomic Manager. Managed Resources also take administrative actions in response to Autonomic Manager requests.

Autonomic Managers oversee the operation of Managed Resources. Managers implement administrative policy and business logic to facilitate and coordinate optimal operation of resources. All direct communication between managers and resources is handled via Java interfaces. Although the toolkit is architected to accommodate distributed managers and resources, the current implementation of manageability interface APIs supports only limited forms of communication on the local device.

Common Base Events

The IBM Autonomic Computing Toolkit defines a standard data format called the Common Base Event. Common Base Events are Extensible Markup Language (XML) structures ("blobs") that define a standard data format for communicating events. Common Base Events provide a convenient mechanism to centralize and correlate events from disparate applications.

Each instance of a Common Base Event may define the software component that generated the event, a location of the event (such as short or fully qualified hostnames), the time of the event, and a description of the situation or scenario leading up to the event.

Generic Log Adapter

One way to generate Common Base Events is through the Generic Log Adapter (GLA). Runtime support in Autonomic Managers utilizes the GLA to convert data from existing log files of legacy applications to the Common Base Event format. For each type of log file to

* Other brands and names are the property of their respective owners.

be supported by the GLA, a developer must implement parsing, formatting, output, and other objects tied together by a configuration file. The toolkit provides an Eclipse plug-in called the Adapter Rule Editor to simplify creation of these configuration files.

Log/Trace Analyzer

The Log/Trace Analyzer is a simple implementation of an Autonomic Manager. Administrators use this analysis tool to graphically view and correlate event log files.

Associated with the Log/Trace Analyzer in the toolkit is support for "symptom databases." These databases consist of XML files that encode possible resolution actions for Common Base Events. A toolkit plug-in allows administrators to build symptom databases according to their policies.

Resource Model Builder

The Resource Model Builder generates data models of monitored resources. Resource models define event types, polling intervals, thresholds, and actions to take when thresholds are crossed. These models employ industry-standard Common Information Model (CIM) classes for holding resource properties. The Resource Model Builder also supports CIM and Windows Management Instrumentation (WMI) standard Managed Object Format (MOF) files in addition to custom scripts.

Automated Management Engine

The Automated Management Engine (AME) hosts deployed resource models. This engine executes resource model scripts within a control loop. It also stores operational data in an embedded local database. AME contains a CIM Object Manager (CIMOM) extensible via Engine APIs.

Integrated Solutions Console

The Integrated Solutions Console (ISC) is a Web-based console user interface. ISC implements centralized management of autonomic capabilities using a WebSphere Application Server as the supporting infrastructure. An Eclipse plug-in supports development of add-on console components. The console supports user interaction in environments where full enterprise management console integration is not in place.

APPLICATIONS IN NETWORK SERVICE CONFIGURATION

In the following sections, we describe several approaches for integrating the IBM Autonomic Computing Toolkit into existing network service configuration technologies to address the usage models described earlier.

Toolkit Integration for Client-Server Environments

At a conceptual level, the IBM Autonomic Computing Toolkit can be integrated in a straightforward fashion with IT client/server network architectures. Each network device can be modeled as a managed resource, and dedicated server or gateway nodes can be configured as Autonomic Managers (clustered as necessary). CIM/WMI support allows easier integration with legacy enterprise management data stores. Specific versions of Java Virtual Machines (JVM*s) can be targeted for enterprise deployment across multiple platforms as needed.

However, the toolkit's current implementation prohibits this design approach, as manager-resource interfaces enable only local (intra-device) communication. Until supported extensions for inter-device communication are available in future versions of the toolkit, workarounds or extensions for the IT environment must be designed. One possible workaround involves adding remote monitoring support to Managed Resources. This approach entails extending the current toolkit architecture with a new "Remote Managed Resource" component as shown in Figure 2.

* Other brands and names are the property of their respective owners.

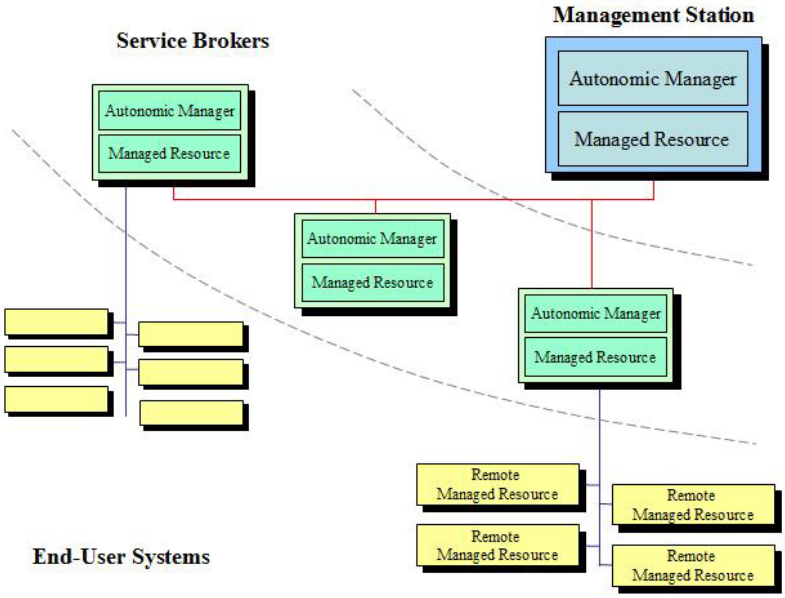

Figure 2: Extended Eclipse autonomic toolkit architecture

In this proposed architecture, Remote Managed Resources are lightweight implementations of toolkit managed resources. These remote resources incorporate the following functions:

- Monitoring and control capability for software and/or hardware on the local device.
- Support for discovery of toolkit Managed Resources and Autonomic Managers.
- Retrieval of control instructions remotely from managers.

To interact with ordinary toolkit resources, remote resources initiate all requests to parent nodes, called toolkit Service Brokers. Toolkit service brokers are simply Autonomic Managers with extensions to support remote resources. They represent a single-service implementation of the general-purpose "services broker" functionality discussed earlier. To pass event logs or alerts to a service broker, for example, remote resources push Common Base Events and other data up to a previously discovered broker. Likewise, to obtain configurations and instructions, a remote resource requests and pulls data from the manager.

These one-way discovery and communication processes can be implemented in IT environments through XML-RPC (Remote Procedure Call) or similar HTTP-oriented protocols. Protocol conventions for decoding configurations and actions must be established between managers and resources. This architecture enables management of remote resources by establishing data-driven monitoring and configuration procedures.

Toolkit Integration for Peer-to-Peer Environments

Attempting to fit the IBM Autonomic Computing Toolkit into a peer-to-peer architecture for network services poses several new challenges. Peer-to-peer architectures utilizing the toolkit demand that any device be capable of taking on Autonomic Manager responsibilities when needed. It follows that each peer must be granted access to the resource models, symptom databases, and event logs deployed in that environment. Besides the increased configuration burden, this architecture also places increased computational demands on nodes that may not be fully equipped to handle the additional overhead.

One method for implementing peer-to-peer toolkit support entails periodic synchronization of all Autonomic Managers. This approach requires a method to verify that each Autonomic Manager configuration is up-to-date as well as a mechanism to trigger propagation of the manager's policy data when needed. The toolkit's architecture must be extended to provide such support. As noted earlier, however, the resource overhead incurred by running an Autonomic Manager may exceed the capability of some peer-to-peer devices. Alternatively, these devices can be configured as lightweight Remote Managed Resources.

Toolkit Integration with Network Services

As described above, developers can, with effort, integrate the IBM Autonomic Computing Toolkit into various client/server and peer-to-peer networks. In IT environments, Autonomic Manager functionality fits

logically as an add-on to existing DHCP, DNS, LDAP and other server platforms. Deploying Remote or full Managed Resource technology to the client-side likewise can be centrally administered using SLP or another available discovery protocol to tie the system together.

On home networks, the difficulty of enabling autonomic services increases. As home "routers" and "entertainment gateways" continue to expand in popularity, these devices become the natural host for Autonomic Manager functions. However, these devices do not generally support a "push" model for deployment, and many home administrators will refuse to install toolkit components on their managed devices manually. Additionally, networked printers, game consoles, and low-end home computers will generally not meet minimum system requirements to serve as toolkit Managed Resources. Given suitable policy setup on the router, the toolkit offers a promising solution for the myriad configuration problems that afflict home networks today. For example, an Autonomic Manager could detect and heal mis-configured workgroup names, wireless encryption settings, and other security settings in addition to basic DHCP setup.

Mobile and roaming wireless users also benefit from features of the IBM Autonomic Computing Toolkit. By pre-installing a laptop or other mobile device with Remote Managed Resource components, the device can discover Autonomic Managers as needed to update DHCP and other network service configurations. During times when the mobile device is not connected to the network, the system can rely on configuration/policy information cached in the local Autonomic Manager. Wireless Internet Service Providers (WISPs) may also leverage the toolkit to build universal sign-on and automated billing services.

AUTONOMIC COMPUTING AND ENTERPRISE MANAGEMENT FRAMEWORKS

Autonomic functions represent the latest step in the natural evolution of enterprise system, network, and storage management capabilities. In the 1980s, Simple Network Management Protocol (SNMP) introduced basic monitoring and control functions for enterprise IP networks. SNMP supports basic configuration, logging, and alerting mechanisms. Unfortunately, SNMP did not define standard resource models for managed devices, hindering its adoption in cross-platform networks.

In the 1990s, Intel, as part of the Desktop/Distributed Management Task Force (DMTF), actively developed and promoted the Desktop Management Interface (DMI) standard. DMI offers a more sophisticated monitoring, configuration, and control framework that enhances support for end-user usage models. Specifically, DMI includes predefined objects for modeling PC client resources and system events.

Subsequently, the DMTF also developed specifications for CIM, which eases the burden of software development associated with DMI. WMI in turn incorporates CIM into standard management software for Windows* platforms. Finally, in the late 1990s, Intel and other companies jointly developed the Intelligent Platform Management Interface (IPMI). Whereas SNMP, DMI, and CIM/WMI function on top of an operating system and a network protocol stack, IPMI provides management capability at a lower level, monitoring platform hardware attributes like voltages, fan speeds, and temperatures, and working independently from the operating system.

Autonomic technologies complement each of these other management standards. Specifically, the IBM Autonomic Computing Toolkit leverages the CIM Object Model and can work with MOF files as indicated earlier. Through the Generic Log Adapter, the toolkit also can aggregate data generated by other software components tied to these standards. In general, autonomic services operate at the next higher level of the management solution stack. It follows that the ability of autonomic solutions to support self-configuring and self-healing depends on the availability of these standard management technologies as well as the extent of instrumentation deployed.

IBM AUTONOMIC COMPUTING TOOLKIT SUMMARY

The IBM Autonomic Computing Toolkit enables developers to add self-configuring and other autonomic capabilities to their software. Capabilities center on the ability to model and monitor resources, implement policies for configuring and controlling those resources, and manage log data using standard cross-application and cross-platform mechanisms. The toolkit attempts to leverage and retain compatibility with existing enterprise management standards. It is a framework best utilized as a reference implementation by early adopters to assess the impact of autonomic technology on their environment.

SUMMARY

In this paper we presented existing network services configuration technologies, identifying current capabilities and shortcomings for dynamically self-configuring network services. We proposed requirements for additions to these existing technologies to address these shortcomings to provide a fully capable autonomic

* Other brands and names are the property of their respective owners.

network service for several types of networked environments. We also described the IBM Autonomic Computing Toolkit, an application development suite that provides software developers with a technology to develop autonomic applications, including dynamically self-configuring network services.

ACKNOWLEDGMENTS

We thank Shu-min Chang, Susan Harris, Marian Lacey, Ray Mendonsa, and Tim Verrall for their valuable technical and editorial contributions to this paper.

REFERENCES

[1] Gulbrandsen, A. and Vixie, P., "A DNS RR for Specifying the Location of Services (DNS SRV)," *IETF* RFC 2052*, Oct 1996.

[2] Droms, R., "Dynamic Host Configuration Protocol," *IETF* RFC 2131*, March 1997.

[3] Hamilton, M. and Wright, R., "Use of DNS Aliases for Network Services," *IETF* RFC 2219*, Oct 1997.

[4] Viezades, J., Guttman, E., Perkins, C., and Kaplan, S., "Service Location Protocol," *IETF* RFC 2165*, June 1997.

[5] Wahl, M., Howes, T., Kille, S., "Lightweight Directory Access Protocol (v3)," *IETF* RFC 2251*, Dec. 1997.

[6] Perkins, C. and Guttman, E., "DHCP Options for Service Location Protocol," *IETF* RFC 2610*, June 1999.

[7] Smith, C., "The Name Service Search Option for DHCP," *IETF* RFC 2937*, Sept. 2000.

[8] T'Joens, Y., Hublet, C., and De Schrijver, P., "DHCP Reconfigure Extension," *IETF* RFC 3203*, Dec. 2001.

[9] Klensin, J., "Role of the Domain Name System (DNS)," *IETF* RFC 3467*, Feb. 2003.

[10] Zhao, W., Schulzrinne, H., Guttman, E., Bisdikian, C., and Jerome, W., "Remote Service Discovery in the Service Location Protocol (SLP)," *IETF* RFC 3832*, July 2004.

[11] Ganek, A.G. and Corbi, T.A., "The dawning of the autonomic computing era," *IBM Systems Journal*, vol. 42, no. 1, 2003, pp. 5-19*.

[12] Bantz, D.F., Bisdikian, C., Challener, D., Karidis, J.P., Mastrianni, S., Mohindra, A., Shea, D.G., and Vanover, M., "Autonomic personal computation," *IBM Systems Journal*, vol. 42, no.1, 2003, pp. 165-176*.

[13] Balakrishnan, H. "Resource Discovery Using an Intentional Naming System," Nov. 1999, http://nms.lcs.mit.edu/talks/stanford-netseminar/*.

[14] Konstantinou, A. V., Florissi, D., and Yemini, Y, "Towards Self-Configuring Networks," *DARPA Active Networks Conference and Exposition*, May 2002, http://www1.cs.columbia.edu/dcc/nestor/nestor-dance-2002.pdf*.

[15] Alex, H., Kumar, M., and Shirazi, B., "Service Discovery in Wireless and Mobile Networks," http://crewman.uta.edu/psi/download/Mohan_Shirazi/ServiceDiscovery.pdf*.

[16] Perkins, C., "SLP White Paper Topic," May 1997, http://playground.sun.com/srvloc/slp_white_paper.html*.

[17] Jacob, B., Lanyon-Hogg, R., Nadgir, D. and Yassin, A., *A Practical Guide to the IBM Autonomic Computing Toolkit*, IBM International Technical Support Organization, http://www.redbooks.ibm.com/redbooks/SG246635/*.

AUTHORS' BIOGRAPHIES

Brian Melcher, GCUX, RHCE, is a senior UNIX/Linux security engineer in Intel's Information Services and Technology Division. His technical interests include UNIX and Linux security and operating system internals. Brian received a Master's degree from the University of Arizona and a Bachelor's degree from the University of Illinois at Urbana-Champaign. His e-mail is brian.a.melcher at intel.com.

Bradley Mitchell is a validation manager and senior software engineer in Intel's Flash Products Group. His technical interests include automation software, high-performance computing, and wireless networking. He holds one patent with two additional patents pending. Bradley received a Master's degree from the University of Illinois at Urbana-Champaign and a Bachelor's degree from M.I.T. His e-mail is bradley.mitchell at intel.com.

Scalable Adaptive Wireless Networks for Multimedia in the Proactive Enterprise

Dilip Krishnaswamy, Intel Communications Group, Intel Corporation
John Vicente, Information Services and Technology Group, Intel Corporation

Index words: wireless networks, Local Area Networks, multimedia, ad-hoc networks, Proactive Enterprise, Scalable Systems, Distributed Network Information Base, cross-layer optimization

ABSTRACT

This paper presents a scalable and adaptive system-level approach to wireless multimedia in the emerging, Proactive Enterprise computing environment. An overview of wireless networks, cross-layer optimization techniques, and advances in wireless LAN technologies is presented. A Distributed Network Information Base with Service Agents at each node is proposed to enable network-wide, proactive adaptation with adaptive routing and end-to-end Quality of Service (QoS) management. The paper suggests that a combination of technological advancements in emerging wireless networks, node-level cross-layer optimizations, and the proposed distributed cross-node system-level architecture are all required to efficiently scale and adapt wireless multimedia in the Proactive Enterprise.

INTRODUCTION

Interactive multimedia on wireless networks requires high bandwidth due to the data rates and payload size [1]. One frequently cited wireless scenario has users conducting real-time, multimedia videoconferencing sessions over a wide-area wireless Internet connection. The wireless client is highly mobile within an enterprise campus or dense city limits and is sharing a media-rich presentation with multiple parties, dispersed in different connectivity scenarios (e.g., home, hotel, or corporate office). The user is experiencing a high degree of QoS supporting the multimedia session and presentation delivery, regardless of his movement or locality. In order for this scenario to occur, multiple effects including time-varying channel conditions, local or remote congestion conditions, and end-to-end QoS requirements must be matched with an adaptive application capable of offsetting the limitations of the network by managing reliability, latency, and throughput degradations while hiding the user from the underlying complexity, degraded QoS, and mobility issues in a wireless environment.

The convergence of wireless networks and the Internet is forging new developments in communication systems and networking services. The rapid evolution of wireless communication technologies is influencing the development of applications, as well as the network services and resources, that will be necessary to deliver traditional voice, data, and emerging media-converged applications, as newer wireless technologies get deployed in the corporate enterprise. Managing service quality and scale across alternative traffic types has long been a challenge in terms of policy-based provisioning and Service-Level Agreement (SLA) management even within the wired infrastructure. These challenges are primarily network administration overhead, limited resources (e.g., network bandwidth) and end-to-end dependencies due to application QoS or mission-critical requirements. With wireless connectivity eventually being more pervasive than physical connectivity supporting the virtual enterprise, these challenges will increase and more than likely create a higher burden on total cost of ownership for Information Technology (IT) managers, based on pre-existing tools and methodologies for provisioning and managing real-time or media-rich applications.

In this paper, we chose to focus on Wireless Local Area Networks (WLANs). WLAN technologies (802.11/Wi-Fi) are targeted to work well within a range of the order of 100 meters. MIMO-based WLAN technologies can use multiple antennas to increase throughput, with the additional capability to trade off increased range for increased throughput. Technologies such as Multiple Input Multiple Output (MIMO) can be applied to other wireless technologies as well. Although WLAN (802.11a/b/g/n) protocols are expected to be the predominant technology for wireless access in an

enterprise, future systems in use in the enterprise may have reconfigurable radio technologies and/or multiple radios. WiMAX (802.16a/d/e*) technologies are expected to have a range in the order of several kilometers (< 50 km). UltraWideBand (UWB*) (802.15.3) and Bluetooth* (801.15.1) technologies would be used for short distances (about 10 meters) with UWB technology enabling high data rate wireless transmissions over short distances. Cellular technologies are primarily optimized for voice traffic, and in general, are geared towards supporting long-range but significantly lower data throughput compared to the Wi-Fi-based technologies. Network management and administration techniques used in cellular networks may be applicable in WiMAX and techniques used in wired networks such as Ethernet* (802.3) could also be applied to wireless networks. One could conceive handoff of a mobile user from/to a WLAN to/from other wireless networks such as a lower data rate cellular network depending on the service availability, user mobility, channel conditions, and location constraints. Seamless transfer of multimedia sessions or Voice over Internet Protocol (VoIP) calls between such networks will be an interesting and challenging task in the enterprise in the foreseeable future.

With devices supporting voice, data, and multimedia communications, the unification of voice and data applications will require a higher degree of complexity to untangle end-to-end service dependencies and manage this from within the core network or traditional distributed systems management tools. Policy-Based Management (PBM) has been positioned as a viable technology to provide greater control and management of underlying network resources via the creation and distribution of high-level policies, integrated with the enabling mechanisms of the network infrastructure. As defined in [1], we define PBM as a "unified regulation of access to network resources and services based on administrative criteria." PBM has been shown to be effective in provisioning QoS, security, and virtual private networks. Nevertheless, PBM is based on a traditional perspective of human administration, similar to many traditional network management tools. In the context of wireless multimedia systems, these systems cannot scale with the time and space varying dynamics imposed by wireless and real-time multimedia systems. In this paper, we argue that, without a systemic shift in how we automate, or in an autonomic fashion, provision and manage rich, multimedia-based services in the wireless enterprise, it is unlikely that a large-scale deployment[1] of multimedia in the wireless enterprise will be achievable. Thus, we propose to move away from the traditional, in-network or administrative means to adapt, scale, and manage multimedia over wireless environments.

In the first section of this paper, we explore options for improved Physical (PHY) and Medium-Access-Control (MAC) facilities to enable capacity improvements in physical and data link layers of the wireless LAN, and we discuss application and transport layer considerations. We present node-level, cross-layer optimizations in the next section and explore opportunities to manage end-to-end state and policies across the layers of the Open Systems Interconnection (OSI) protocol stack. A Distributed Network Information Base is then proposed with Service Agents to enable proactive adaptation for end-to-end QoS management and adaptive routing. Finally, we propose integrating state and policy at all layers of the OSI stack with local, global, and end-to-end services by cohesively integrating in-network facilities with node-level, cross-layer optimizations to achieve multimedia adaptation and scale in the wireless environment.

WIRELESS LAN CONSIDERATIONS

WLANs offer several challenges with regard to multimedia streaming [3, 4, 5]. Dynamic variation in channel conditions due to noise, interference, and path loss effects impact data throughput and packet loss. Dynamic changes in the number of users in the network with their varying data rate requirements resulting in a varying degree of contention and collision in the network impact the amount of bandwidth per user or per flow. Real-time adaptation at the MAC layer is required to adapt to varying conditions. The choice of the transport layer such as Transport Control Protocol (TCP) or User Datagram Protocol (UDP) [1, 5] is also of concern. Multimedia applications have the ability to scale [1] and adapt to varying wireless network conditions, which must also be considered and exploited.

Physical Layer Considerations

At the PHY layer, various modulation and coding schemes are available to a wireless station for transmitting data. Modulation schemes that allow more bits per symbol, help in increasing data rates; however they have symbols closer to each other (in the constellation diagram), and small errors could result in erroneous decoding. Varying code rates can be employed

* Other brands and names are the property of their respective owners.

* Bluetooth is a trademark owned by its proprietor and used by Intel Corporation under license.

[1] Specifically we are referring to many "point-to-point" or many "point-to-many points" scenarios.

within each modulation scheme to adapt to changing channel conditions by allowing more bits for coding (i.e., lower code rate *k/n*) for more robust transmission as conditions deteriorate. As the code rate decreases, the effective data rate is reduced, and hence the achievable throughput is also reduced. We use the term "mode *m*" to refer to a specific choice of a modulation and coding scheme. The probability, $P_e^m(L)$, of error in a packet of length L bytes (also referred to as the physical layer packet error rate or PER), for a given mode *m*, as a function of the bit error probability p_b is given by equation (1), where the inequality represents the fact that one can recover from bit errors in a packet, due to the coding scheme used at the packet level.

$$P_e^m(L) \le 1-(1-p_b^m)^{8L} \quad (1)$$

The effective PHY layer throughput can be then expressed as $T_{PHYm}(x) = A(1-P_e^m(L))$. For a given mode *m*, $T_{PHYm}(x)$ and $P_e^m(L)$ can be approximated with sigmoid functions [6] of the form

$$T_{PHYm}(x) = A/(1+e^{-\lambda(x-\delta)}) \quad (2)$$

$$P_e^m(L) = 1/(1+e^{\lambda(x-\delta)}) \quad (3)$$

where x is the SINR in dB, and $y = T_{PHYm}(x)$ is the throughput in Mbps. Link adaptation schemes are used so that a user adapts and operates in a region on the throughput curves where the PER is low. When P_e is small, $log(P_e) = -\lambda(x-\delta)$. Game theoretic formulations could be used to develop optimization techniques for multimedia adaptation by using such sigmoid mathematical modeling as described in [6, 7]. PHY-level throughput versus Signal to Interference+Noise Ratio (SINR) curves are shown in Figure 1 for 802.11a/g networks. The max PHY-layer throughput of 54 Mbps is obtained in the 64 QAM, rate ¾ mode.

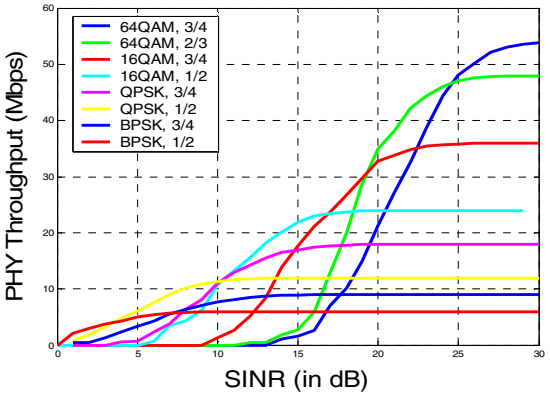

Figure 1: PHY throughput for 802.11a/g networks

MIMO for Improved PHY-Layer Capacity

Multiple transmitting antennas [8] can help increase the data rates in the same channel. Multiple receiving antennas can help in efficient recovery of the transmitted data. Multiple antennas can also be used to increase range and reliability of data transmitted in a channel without an increase in data rates. Alternatively, one could increase data rates using multiple antennas by transmitting data in multiple channels simultaneously.

With two transmit antennas (typically in a 2x3 MIMO configuration with 2Tx antennas and 3Rx antennas), one can expect a throughput of 108 Mbps [5] in a 20 MHz channel. Using a wider 40 MHz channel, as opposed to a 20 MHz channel can increase the throughput to 216 Mbps. With four transmit antennas, the throughput can increase further to 432 Mbps. Using additional Orthogonal Frequency Division Multiplexing (OFDM) sub-channels and/or using newer coding schemes such as Low-Density Parity Check (LDPC) codes approaching the Shannon limit, could increase the throughput to over 500 Mbps in future WLAN systems. These improvements in overall PHY-layer capacity are being pursued in the 802.11n standard.

MAC Performance Considerations

The effective throughput at the top of the MAC is further reduced due to a number of factors [5, 7, 9, 10] such as the number of current users in the network medium, user requirements, priorities, retry-limits, and link adaptation schemes used, channel conditions based on noise and interference, backoff counter depths, backoff stages, protocol timing, and header overheads, and also, the amount of additional time that the medium is unused/idle. The overall throughput is also affected by the transport mechanism used (TCP/UDP/UDP-lite), and by whether there is additional application-layer redundancy such as Forward Error Correction (FEC) across packets being used. In general, the overall throughput as a function of the SINR continues to assume a sigmoidal form with a reduced maximum asymptotic value for the throughput.

MAC Throughput Considerations

The throughput at the top of the MAC is affected by the following:

1) The PHY-layer throughput. (The PHY-layer throughput depends on the PER at the physical layer and the transmission duration for each packet, which in turn is a function of the modulation and coding scheme used during transmission.)
2) Protocol timing overheads such as interframe spacings, and acknowledgement time.

3) Time spent in the random backoff counter (the value range increases exponentially with transmission failures).
4) Time during which the medium is busy with other users transmitting.
5) Unused idle time in the network.

For example, in a 54 Mbps PHY mode in an 802.11a WLAN, including protocol overheads, the ideal MAC throughput is approximately 30 Mbps, and one may obtain a throughput of only 24 Mbps in a typical WLAN. Since only one user transmits at a given time using the Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) protocol, the available bandwidth for a user can be considerably reduced depending on the data rate requirements of other users in the same network.

Packet Errors and Retransmissions

Retransmissions are attempted at the MAC layer in the absence of an acknowledgement. Packets may be in error if the acknowledgement is not received from the destination within a specified duration after transmission. Either the packet may be in error when it reaches the destination or the return acknowledgement might itself be in error. In both cases, a packet is retransmitted by the MAC, as long as the retry-limit is not reached. During retransmissions, link adaptation may be performed to attempt sending a packet in a more robust modulation and coding scheme. Queue backup in the transmit queue at the MAC layer can impact performance. Jitter and delay requirements may impact how many retransmissions should be attempted before further packet retransmission attempts are discarded. If a reliable transport mechanism such as TCP is used, then retransmissions at the MAC layer should be preferred as MAC-level retransmissions have significantly less overhead compared to retransmissions attempted from the transport layer. Additional overhead at the MAC layer occurs due to collision with other users in the CSMA/CA protocol. To avoid collision, an exponentially increasing backoff counter is used, which can cause increased overhead.

Link Adaptation

The desired behavior for link adaptation to determine the optimal choice of the modulation and coding scheme to use is depicted in the "hysteresis loop" in Figure 2. Here OOP represents the Optimal Operating Point [6] for a given choice of the modulation and coding scheme. (In this figure the throughput functions are zoomed in for two modes.) The throughput improvement becomes reduced beyond the KNEE [6] (point of maximum curvature on the sigmoid), and, beyond the OOP, the throughput only improves marginally. The Adaptation Switching Point (ASP) denotes where it can become advantageous to switch to a different modulation and coding scheme. As the network conditions change, the OOP will dynamically vary for each wireless station. It can be difficult to distinguish between foreign interference, collisions, and noise. The effectiveness of the adaptation is limited by the rate at which the channel changes (including activity of other devices) and how fast the algorithm adapts (limited by the integration period over the metrics used for adaptation).

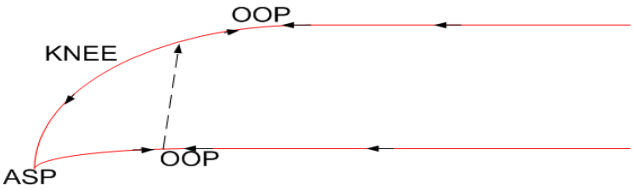

Figure 2: Hysteresis Loop for adaptation

In general, more robust modes are used (less bits per symbol or more coding) when channel/link conditions are degraded. However, these more robust modes take longer to transmit the same amount of data, which reduces throughput.

802.11e Protocol Improvements

The 802.11e protocol is designed to optimize QoS in a wireless network. Two service classes are suggested: QoSNoAck and QoSAck. The QoSNoAck mechanism may not be very useful as it is typically desirable to attempt retransmissions at the MAC layer. There are four access categories (Voice, Video, Best Effort, and Background) and eight user priorities (two for each access category). The access point configures the MAC-level parameters such as the contention window size (CWmin/max), the interframe spacings (AIFS), and the transmission opportunity (TxOP) duration. Reducing the contention window size allows a transmission to be attempted sooner with reduced backoff. Reducing interframe spacings reduces protocol overhead. Increasing the transmission duration allows more blocks of data to be sent in sequence with the same TxOP duration. A block acknowledgement can indicate the correctly received blocks so the blocks in error can be retransmitted during the TxOP duration. The Wi-Fi Multimedia (WMM specification) implements the Enhanced Distributed Channel Access (EDCA) contention-based access mechanism suggested in 802.11e with the access priorities. The Wi-Fi Multimedia–Scheduled Access (WMM-SA) specification provides support for HCCA-based centralized scheduling in addition to EDCA. The Hybrid-coordination-function Controlled Channel Access (HCCA) mechanism allows for bandwidth to be dedicated to a specific transmission in a contention-free period. All stations admitted with dedicated bandwidth in the contention-free period complete their transmissions and then the rest of the stations contend for access using

EDCA during the contention period. Multimedia applications that need guaranteed time on the medium may use the HCCA mechanism to access the medium. The overall 802.11e specification is expected to include the WMM and WMM-SA specifications, and to support additional features such as Block Acknowledgements, Automatic Power-Save Delivery (APSD) capabilities, and Direct Link Setup (DLS) for peer-to-peer (P2P) communication.

802.11n MAC Enhancements

MAC-layer enhancements are being suggested in the 802.11n standard, to supplement the improved capacity with MIMO in the PHY layer. One key enhancement is to enable aggregation of several MAC-level protocol-data units into a single PHY-layer protocol data unit. This enables a longer packet to be sent relative to the protocol timing overheads associated with the WLAN transmission. Block acknowledgements specifying the correctly received portions of the aggregated message are used. Receiver feedback to a sender for improved link adaptation at the sender is enabled. Reverse direction data flow can be used by a receiver to use available transmission time to transmit data to a sender in conjunction with an acknowledgement. These optimizations can further help to improve application performance and scalability in a wireless environment.

Application/Transport Layer Considerations

UDP is cited as the preferred transport mechanism [1, 4, 5] for video and audio streaming. TCP can incur increased delay and jitter with TCP's congestion-avoidance mechanism and re-transmission [1, 5]. Application-layer FEC [3, 4] between packets over UDP could be used to compensate for lost packets at the MAC layer, with error-concealment strategies [1] used at a receiver to mitigate the effect of packet losses. One needs to exploit scalability in multimedia representation and identify the most important information to communicate given the available conditions. Application and MAC-PHY cross-layer optimizations [4, 7] and joint source/channel coding [3, 11] can help in adapting to optimally transfer the most relevant information over the wireless channel in response to current channel conditions.

The video quality can be represented, in general, using the rate-distortion model $D_e = \theta/(R_e - R_0) + D_0$, that was suggested in [3]. However, for the range of operation and the SINR values required to make decisions, the video quality Q (i.e., PSNR) could be approximated by a linear equation (or piece-wise linear) [7] of the PER P_e of the form $Q = -\mu (P_e - \rho)$. The choice of the adaptation mechanisms at the multimedia algorithm level can influence the value of μ. If P_e is small, then we can assume $\log (P_e) = -\lambda(x - \delta)$. Therefore $dQ/dx = \mu \lambda P_e$. Thus, one can study the variation in video quality as a function of the varying channel conditions and establish correlation between conditions at the physical layer to perceived application performance. This suggests that one could consider direct cross-layer interactions between various layers in the protocol stack to proactively optimize multimedia performance in a wireless environment.

NODE-LEVEL PROTOCOL-STACK CROSS-LAYER OPTIMIZATIONS

Cross-layer optimizations in ad-hoc wireless networks [12, 13] have been proposed for direct cooperation between layers in a protocol stack to achieve optimal performance. Recent research in the area of streaming wireless multimedia has focused on cross-layer optimizations [4, 7] in the protocol stack at each node. Such optimizations include link adaptation at the MAC layer, retry-limit adaptation at the MAC to compensate for packet errors, application-layer FEC to compensate for packet losses at the MAC, traffic reshaping to handle varying bit rates, dynamic resizing of buffers, management of arrival and departure rates into queues, reducing end-to-end delay and jitter to meet real-time requirements, adaptive modulation schemes to use more robust modes for base layers, joint source-channel coding, channel reassignment under worsening conditions, and the use of more robust modulation and coding schemes for interference tolerance. This node-level adaptation is done quickly with interaction between key protocol layers such as the application layer, the MAC layer, and the PHY layer based on the knowledge of the current conditions in the network. The scalability inherent in multimedia representation helps in adapting to such dynamically varying constraints. Additional optimization could use light-weight communication information exchange to indicate degradation in battery availability for mobile devices, and in memory availability with other simultaneous tasks needing to be supported in the system. Such information can be useful for the two end points to adaptively reduce processing requirements in terms of frame size and frame rates and hence minimize energy utilization. This would extend the duration of a multimedia application such as a video conferencing session.

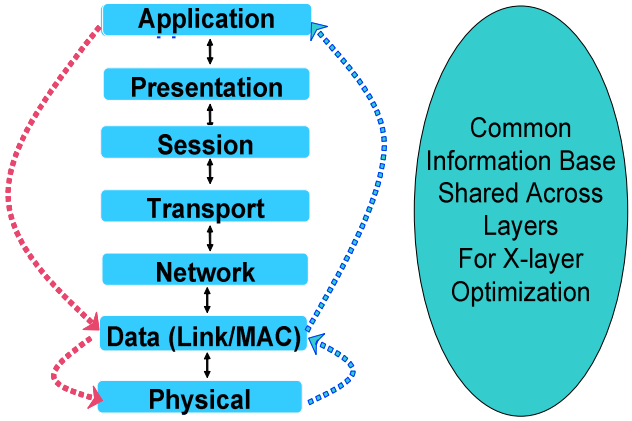

Figure 3: Node-level application-MAC/PHY x-layer optimization

Figure 3 shows cross-layer optimization between the application and MAC/PHY layers with direct exchange of information between the layers, and adaptation feedback between the layers. A common information base can be used to share information between layers. The action to be taken is determined by the layer that owns the action, based on feedback from the various layers.

With direct interaction between the layers, each layer can adapt proactively with the most current information in the layers. If channel conditions degrade, and if such information is made available to the application layer much sooner compared to information propagation through the protocol stack, then the application can proactively adapt based on current conditions in the network. Likewise decisions at the MAC layer can be proactively taken to prioritize scheduling or selective transmission of information over the channel, based on current channel conditions.

SYSTEMS ARCHITECTURE FOR WIRELESS MULTIMEDIA SCALABILITY AND ADAPTATION

To support a horizontal (end-to-end) and vertical (node-level) systems orientation to scale and adaptation of wireless multimedia, we believe the end-to-end principle [14] and in-network control must converge. The ability of applications to adapt to positive or negative changes in wireless conditions, by either leveraging in-network services or binding alternative node-level, inter-layer optimizations, will give applications greater flexibility in managing real-time or media adaptation within a wireless environment. We therefore call for tighter layer integration and automation of application control and bandwidth management. The application's adaptation flexibility will depend on its ability to detect or respond on a much faster time-scale (e.g., to support fast handoff), thus requiring the application to cooperate with the transport layer's congestion control loop. Alternatively, the necessity to manage the wireless channel bandwidth will depend on the clocking rate and control mechanisms being used by the application or session layer to control the incoming rate of the flow; sharing knowledge or policies between these layers can further increase their cooperative effectiveness. Additionally, we view existing layered or component services (e.g., application-layer FEC, TCP congestion/flow control, 802.11e, IP Differentiated Services) coordinating functions of congestion management, service differentiation, bandwidth management, and reliability to achieve a higher degree of local system optimization.

We propose broadcasting or multicasting system-level information in smaller-scale, tightly coupled networks, or direct exchange of information between nodes with information propagation when required. Such system-wide proactive adaptation and management of resources real-time can ensure that the system is made more aware and resilient to dynamically varying constraints, ensuring that the impact of worsening conditions is minimized or that the best option is taken during improved network conditions. Further, we propose using direct peer-to-peer transmission after establishing contact between peers, if the direct peer-to-peer wireless link is good and if direct communication between peers is enabled. Alternatively, one may have to use multiple hops over short distances to reduce packet errors due to path loss over large distances in wireless links. However, hops over the same frequency channel can cause contention, which can reduce available bandwidth. In multihop networks, one has to be concerned with both exposed nodes (in the sender's range but out of the range of the destination) and hidden nodes (out of the sender's range but within the range of the destination). The network configuration between two endpoints may vary dynamically depending on varying network conditions, mobility, or other constraints in the system. QoS policy or state information exchange through intermediate nodes is also necessary to meet end-to-end requirements.

In large wireless ad-hoc networks, routing tables with link information can grow significantly in size [15]. This information could include link quality information as perceived at the MAC/PHY layer at a node. To achieve scalability, nodes should store only local information about nearby links (such as information about links that are only one hop or two hops away). This information stored in a distributed fashion can be propagated through nodes on request, to understand end-to-end performance on a communication path between two endpoints in the network. The response time of adaptation mechanisms at the nodes based on these dynamically varying conditions will determine how effective and proactive the adaptation mechanisms are, to ensure that any variation in the quality

of the received multimedia transmissions is imperceptible to the user.

Systems Approach

The adaptive techniques at a single node are not sufficient to address overall scalability issues in the wireless or hybrid wireless and wired enterprise networks. To address end-to-end QoS, we must extend the cross-layer optimization techniques to address constraints and issues in the network and across the enterprise to provide additional system-level feedback and more intelligent adaptation in the protocol stack at respective nodes in the end-to-end path. Optimizing overall end-to-end QoS requires knowledge or state of key elements of the network system or the communicating session. This knowledge must also be shared effectively between these elements. The optimization across the network will be required due to variations in link conditions and user mobility constraints in the wireless environment. Current conditions at the MAC and PHY layers can be propagated to the network layer at each node, and joint optimization between the network and MAC layers can be used, in conjunction with network-wide information, to optimize routing and end-to-end QoS dynamically in the network as depicted in Figure 4.

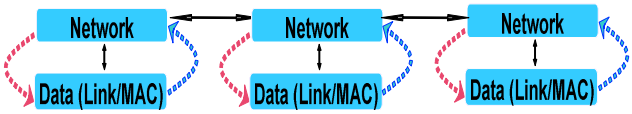

Figure 4: Joint optimizations between the IP layer and the Data (Link/MAC) layer across nodes for adaptive routing and end-to-end QoS management

Network-wide adaptation can be proactively achieved, with fast exchange of information about individual links between two endpoints to assess end-to-end performance in the network. The PHY-layer conditions on each link in a wireless network can dynamically vary due to several factors such as network congestion with other users, interfering signals and noise, and the path loss associated with the link. The optimal path between two endpoints in a network would be a function of the quality of each of the links on the path between the end points. With dynamically varying constraints in a wireless network, a statically configured optimal path may soon become a less optimal one. Alternate paths for communication of information can be proactively established to quickly switch to the best alternate path, as conditions vary over time in the system.

We propose a Distributed Network Information Base (DNIB) with query management for retrieval of end-system-level or network node-level knowledge or information. A DNIB Service Agent (DNIB-SA), as illustrated in Figure 5, will be available at each node to respond to requests for information from a neighboring node in the network. The DNIB-SA will respond with information available from the DNIB content locally available at the node, and, if necessary, the agent will forward the request to one or more neighboring node(s), until end-to-end information between two end points involved in the communication is obtained. The DNIB content, locally available at a node, could consist of information about a node and recent information about neighboring nodes (based on the depth–number of hops–of the routing table maintained at the node). Request forwarding is done selectively to nodes that are more likely to be candidates for the communication path (based on link quality, link utilization, and node-location information, if available) between two end points. During request forwarding, a list of visited nodes is maintained to avoid loops, and information gathered is propagated back to the end points.

The DNIB-SA should monitor conditions on links being actively used in its routing table. Network conditions on active links in an active route can vary due to user mobility or due to changes in link quality or link utilizations. For an active critical weak link on which conditions may be degrading, the DNIB-SA will quickly propagate information about dynamic variation in network conditions to the nodes on the active route, to eventually propagate the information to the end points on the active route. This will enable proactive end-to-end QoS management between endpoints involved in communication, with real-time adaptation. Additionally, a subset of DNIB-SAs corresponding to nodes in an active route could dynamically reconfigure an alternate route in a sub-network. Local routing tables will get modified to reflect the new active end-to-end route. End-to-end QoS information for the newly chosen route will be propagated to the endpoints.

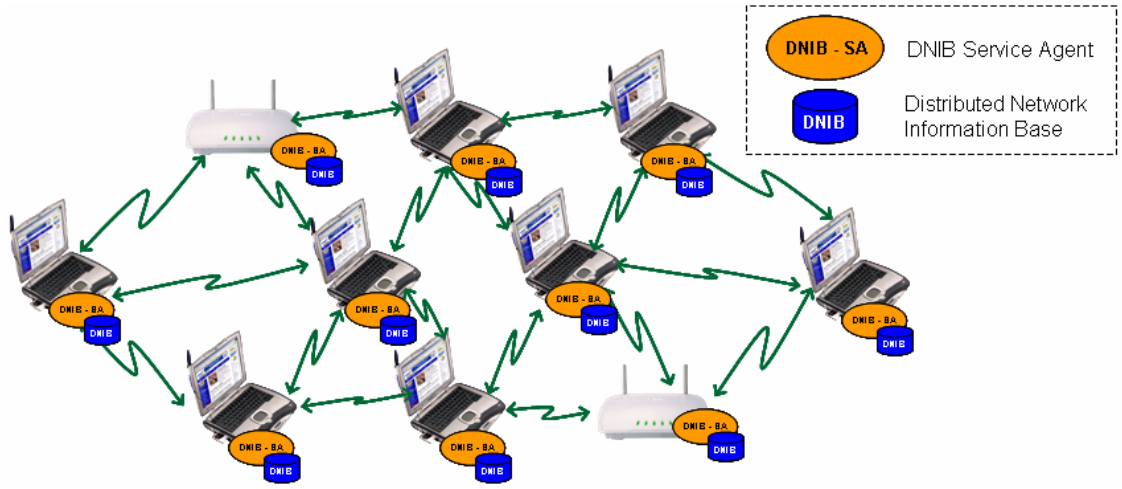

Figure 5: DNIB Service Agent distributed configuration

To summarize, DNIB-SAs have the ability to perform the following tasks.

1) Respond to requests from neighboring nodes.
2) Monitor conditions on active links in active routes and forward information about dynamic variations in link quality to nodes in the local routing table at a node, with information to be propagated quickly to endpoints in a communication path to enable real-time adaptation.
3) Change policies that can relate to routing, QoS, or security as needed. In addition, depending on local, global or end-end situations, the policies could relate to changing flow transport parameters or changing monitoring requirements.
4) Change network configuration and enable localized route modifications when required by collaborating with DNIB-SAs in a sub-network to configure an alternate route when conditions on a link degrade.
5) Propagate end-to-end information to endpoints for a newly configured route.

Finally, we suggest partitioning the control system across three layers of resource control hierarchy. We believe that this is necessary to achieve autonomous control on a global level, a local level, and a flow level; each of which is directed at a different set of objectives (e.g., global usage efficiency, local access maximization and fairness, and end-to-end flow control and adaptation), but overlapped on their influence on the wireless channel resource.

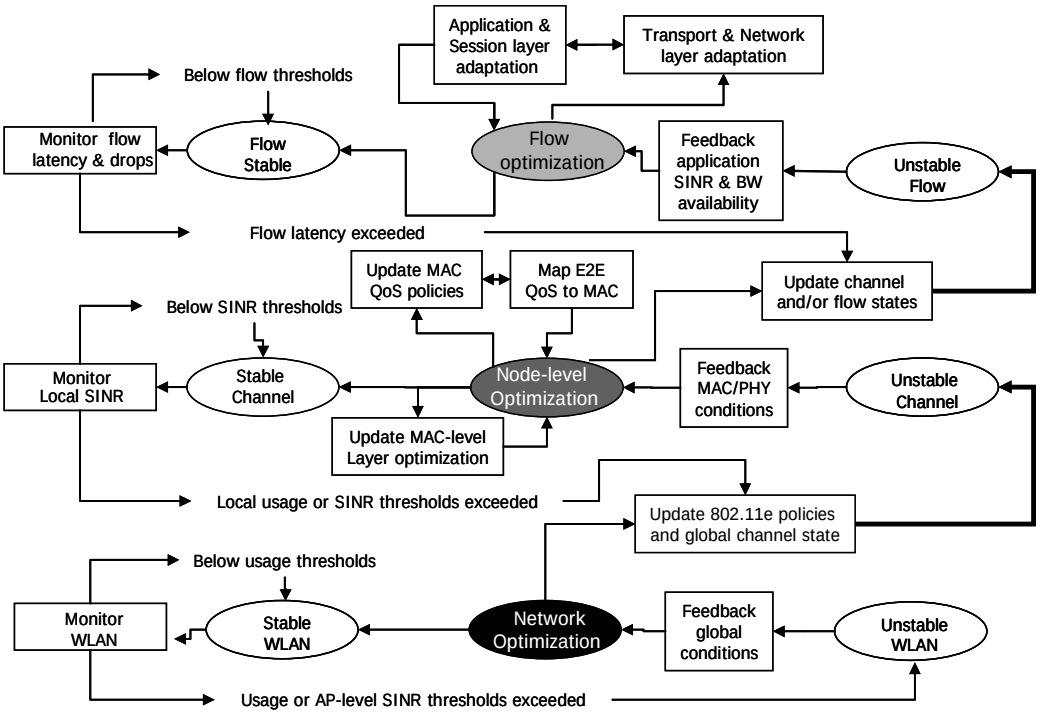

Figure 6: Distributed State Machine

As illustrated in Figure 6, three autonomous levels of feedback-based control are supported in the proposed Distributed State Machine (DSM) [16]. At each level, a stable and unstable state exists, while an operational state is centered between them to represent the control state. Also at each level, a monitoring procedure checks against stability thresholds to determine the possibility of instability and the need to enter into a control state, invoking alternative algorithms, which manage the particular level of concern. At the global and local level, policy changes will cause the state machine to enter into an unstable local channel state or unstable flow state, respectively. Multiple instances of the state machine procedure will run; one per wireless node and one for each of the flows running within the wireless environment. The flow procedure is essentially part of the normal transport process supporting both congestion and reliability control for each session flow. However, we expose it here as a necessary integration aspect of our systems framework. Also shown at the local and flow level is a procedure to update the respective layers of the stack on specific bandwidth availabilities and SINR state, respectively. This multitier DSM approach does not suggest these three (i.e., state machines) threads as being independently managed or controlled. Instead, they must be cooperative and autonomic; policies and states for their operation are exposed or exchanged for the intended stability and balance of the system operation. To achieve this, we consider the cross-layer optimization framework, where layered services cooperate through exposed interfaces for binding purposes, dynamic configuration, or state management. Furthermore, we can have greater control over the stability and efficiency of the system by enforcing policy controls at different timescales as needed to react, maintain, or be proactive, as warranted by the wireless device, the application flow, or the wireless channel.

To summarize, an overall systems approach is proposed with three major elements:

1) Adaptive capabilities to reconcile network delivery requirements and state management on a local and end-to-end level.

2) Distributed information base at a cross-node level and network-wide level, for adaptive scalable routing and dynamic end-to-end QoS management. Information propagation is suggested using point-to-point, broadcasting, or multicasting communication mechanisms.

3) A three-level DSM supporting closed-loop control and management is suggested with interactions between the three levels (flow, node, and network levels).

CONCLUSION

In conclusion, an alternative approach to traditional methods is recommended for provisioning and managing a scaleable media rich, mobile, and dynamic proactive enterprise. This paper proposes a system-wide, cross-

node, cross-layer optimization architecture to scale the number of multimedia users, audio/video streams, and varying multimedia bandwidth requirements, while cognizant of competing mission-critical traffic and aggregate demands on network resources. A Distributed Network Information Base with Service Agents at each node is proposed for adaptive routing and dynamic end-to-end QoS management. Combining node-level, cross-layer optimization architectural enhancements, end-to-end flow, and network-wide feedback and optimizations, we have proposed a closed loop, multitier resource control and management system architecture to enable self-manageable multimedia adaptation and scalability.

ACKNOWLEDGMENTS

We are grateful to Tom Gardos and Gene Matter for their feedback and comments regarding this paper.

REFERENCES

[1] *Compressed Video over Networks*, Editors Ming-Ting Sun and Amy R. Reibman, Marcel Dekker Publishers, 1st edition, 2001.

[2] R. Rajan, D. Verma, S. Kamat, E. Felstaine, S. Herzog, "A Policy Framework for Integrated and Differentiated Services in the Internet," *IEEE Network Magazine,* Sept./Oct. 1999.

[3] K. Stuhlmuller, N. Farber, M. Link, B. Girod, "Analysis of Video Transmission over Lossy Channels," *IEEE Journal on Selected Areas in Communication*, Vol. 18, No 6, June 2000.

[4] M. van der Schaar, S. Krishnamachari, S. Choi, X. Xu, "Adaptive cross-layer protection strategies for robust scalable video transmission over 802.11 WLANs," *IEEE Journal on Selected Areas in Communication*, Vol. 21, Issue 10, Dec. 2003, pp. 1752–1763.

[5] D. Krishnaswamy, R. Stacey, R. van Alstine, W. Chimitt, "Performance Considerations for Efficient Multimedia Streaming in Wireless Local-Area-Networks," 49th Annual SPIE conference on Applications of Digital Image Processing, August 2004.

[6] D. Krishnaswamy, "Game Theoretic Formulations for network-assisted resource management in wireless networks," *IEEE Vehicular Technology Conference*, pp. 1312–1316, Sept. 2002.

[7] D. Krishnaswamy, M. van der Schaar, "Adaptive Modulated Scalable Video Transmission over Wireless Networks with a Game-Theoretic Approach," *IEEE Multimedia Signal Processing Workshop*, September 2004.

[8] G. J. Foschini, M. J. Gans, "On Limits of Wireless Communications in a Fading Environment when Using Multiple Antennas," *Wireless Personal Communications 6*: 311–335, 1998.

[9] G. Bianchi, "Performance analysis of the IEEE 802.11 Distributed Coordination Function," *IEEE Journal on Selected Areas in Communications*, Vol. 18, Issue 3, March 2000, pp. 535–547.

[10] D. Qiao, S. Choi, K.G. Shin, "Goodput analysis and link adaptation for IEEE 802.11a wireless LANs," *IEEE Transactions on Mobile Computing*, Vol. 1, Issue 4, Oct-Dec. 2002, pp. 278–292.

[11] L. Cheng, W. Zhang, L. Chen, "Rate-Distortion Optimized Unequal Loss Protection for FGS Compressed Video," *IEEE Trans on Broadcasting*, Vol. 50, No, 2, June 2004, pp. 126–131.

[12] M. Conti, G. Maselli, G. Turi, S. Giordano, "Cross-Layering in Mobile Ad-hoc Network Design," *IEEE Computer*, Feb. 2004, pp. 48-51.

[13] G. Carneiro, J. Ruela, M. Ricardo, "Cross-layer design in 4G Wireless Terminals," *IEEE Wireless Communications*, April 2004, pp. 7-13.

[14] J.H. Saltzer, D.P. Reed, and D.D. Clark, "End-to-End Arguments in System Design," http://www.reed.com/Papers/EndtoEnd.html*

[15] C. Santivanez, B. McDonald, I. Stavrakakis, R. Ramanathan," On the Scalability of Ad-hoc Routing Protocols," *IEEE Infocom*, New York, June 2002.

[16] J. Vicente, A. Campbell, "IP Radio-Resource Control System," *IFIP-IEEE Conference on Management of Multimedia Networks and Services*, 2002.

AUTHORS' BIOGRAPHIES

Dilip Krishnaswamy received his B.Tech. degree in Electronics and Communication Engineering in 1991 from the Indian Institute of Technology, Madras, his M.S. degree in Computer Science in 1993 from Syracuse University where he was a University Fellow, and his Ph.D. degree in Electrical and Computer Engineering in 1997 from the University of Illinois at Urbana-Champaign. He received the best paper award for the 1997 IEEE VLSI Test Symposium. He is currently staff architect in the Communications Architecture Lab in the Intel Communications Group at Intel Corporation. He was the architect for the Intel® PXA800F Cellular processor. He teaches courses related to computer architecture and parallel computer architecture at the University of California, Davis. He is the vice-chair of the IEEE Communications Society emerging technical committee on Communications Design and Development. His e-mail is dilip.krishnaswamy at intel.com.

John Vicente, an Intel principal engineer, is the director of Information Technology Research and chair of the IT Research Subcommittee. John joined Intel in 1993 and has 19 years of experience spanning R&D, architecture, and engineering in the field of information technology networking and distributed systems. John has co-authored numerous publications in the field of networking and has patent applications filed in internetworking and software systems. He is currently a Ph.D. Candidate at Columbia University's COMET Group in New York City. John received his M.S.E.E. degree from the University of Southern California, Los Angeles, CA in 1991 and his B.S.E.E. degree from Northeastern University, Boston, MA in 1986. His e-mail is john.vicente at intel.com.

THIS PAGE INTENTIONALLY LEFT BLANK

Bringing Security Proactively Into the Enterprise

Sanjay Rungta, Information Services and Technology Group, Intel Corporation
Anant Raman, Technology and Manufacturing Group, Intel Corporation
Toby Kohlenberg, Information Services and Technology Group, Intel Corporation
Hong Li. Information Services and Technology Group, Intel Corporation
Manish Dave, Information Services and Technology Group, Intel Corporation
Greg Kime, Information Services and Technology Group, Intel Corporation

Index words: firewalls, packet filtering, port security, super-VLAN, sub-VLAN, distribution layer, access layer, patching, hardened systems, policy-enabled network

ABSTRACT

Prevailing network architectures are designed for openness, collaboration, and sharing. The majority of viruses and worms use the network to spread rapidly through the enterprise network, enabling these cyber threats to reach their targets effortlessly. The most common solution available today for cyber security is hardening of systems via "patching" or keeping the operating systems, applications, and anti-virus software current. This option is reactive and time/labor intensive because security updates are available only after exploits are known and already in use. The currency of software does nothing to prevent cyber attacks from reaching their targets. We believe that policy-enabled network security complemented by system hardening, provides a proactive and more comprehensive strategy to deal with security by reducing the likelihood of cyber threats entering the network and by controlling their spread. Typical enterprise network architectures are developed to bring scalability, extensibility, and availability to the Intranet. Security capabilities have not been part of the enterprise network architecture and are typically implemented in reactive fashion. Additionally, current security capabilities require manual and labor-intensive efforts that negatively impact costs and take time to implement. Firstly, we propose a change to the enterprise network architecture by integrating security components such as packet filtering, stateful inspection, port-based access control, and super/sub Virtual Local Area Networks (VLANs). Secondly, we propose a fundamental change in the implementation of the enterprise network architecture by using a security management system referred to as Policy-Enabled Network Security (PENS) that leverages the new security capabilities in an integrated and proactive manner and reduces unstructured manual, labor-intensive, and error-prone activities.

INTRODUCTION

One of the major security problems in the enterprise network is the "permit all" capabilities of all Local Area Networks (LANs). While the openness was a catalyst to the growth of computer networks, it also presented and continues to allow computers with security issues to freely connect to the network and potentially infect other computers. LANs do not have the capability to determine who is connected to the network and therefore, in an enterprise, cannot separate out company employees from visitors. Moreover, the enterprise network is now required to support flexible connectivity to portable computers such as laptops for company employees as well as visitors such as field service personnel, consultants, customers, partners, etc.

The enterprise network is also increasingly becoming more diverse and complex with the explosive growth in the usage of wireless technology. New business models such as supplier support of equipment over the Internet, remote operations of equipment over the network (with safety precautions), and company guests accessing their Intranets over the Internet, have blurred the lines between the Intranet and Internet. The added flexibilities in the workplace cause a greater security risk than ever before: the source of an infection nowadays can be any host accessing the enterprise. Figure 1 demonstrates the typical enterprise network.

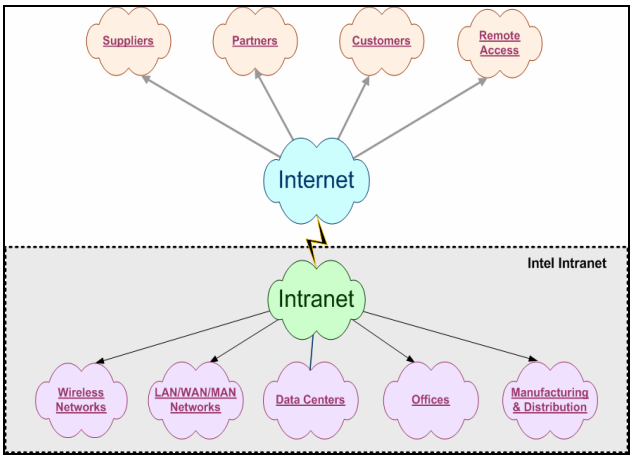

Figure 1: Network connectivity in the enterprise

Unhardened Computer Systems

An enterprise network typically has several operating system and software standards such as UNIX*, Microsoft*, Apple*, OpenVMS*, etc. Most commercial off-the-shelf operating systems provide a general-purpose computing platform for different types of applications and configurations. These systems tend to have more services than are typically needed, and most people tend to keep these services running. Vendors provide frequent software updates and bug fixes. These computer systems, commonly known as unhardened computer systems, have now become the major source of security issues. They can easily become infected and in turn, infect other computers accessible over the LAN.

Inadequate Policies and Management

Traditionally, enterprise policies have been limited to perimeter security resulting in the usage of controls to reduce risks within the enterprise network. With the increase in interactions across the perimeter as we just described, the challenge is to adapt security policies to reflect these new threats. These policies and corresponding operational capabilities have to be able to support complex protection profiles to protect each participant in the entire enterprise network. Each participant contributes to the business and is a challenge to the security of the business based on his/her level of interaction with the environment. This seemingly paradoxical situation increases the complexity of managing security policies and of enforcing them. Traditional processes and methods for implementing security are manual, labor intensive, and error prone.

Intranet No Longer a "Safe Haven"

In the aftermath of virus attacks such as SoBig, Nimda, FunLove, SQL Slammer, etc., as well as a lack of a policy management system, the Intranet, as it exists today, is no longer secure. Enterprises are struggling to keep up the frenetic pace of updating software because the hackers are ready with the next set of attacks before the systems are updated with fixes for the previous attack. Keeping thousands of computers updated is a very reactive process because it can only secure us from problems with known and available solutions.

Solving the Problem

We started our research by reviewing our current network design methods, security capabilities, and management practices and concluded that a change in the network architecture is the most effective way to bring security to the enterprise network proactively. A product-level approach is also possible, but it would require us to find products that can provide enterprise-level security capability for 100,000 to 500,000 computers, 100,000 to 150,000 LANs, and greater than 10,000 subnets. This would take several person years and would not guarantee us a solution to our security problems. This approach is not feasible and is therefore, not addressed in this paper.

ADDING SECURITY TO THE ARCHITECTURE

Today's enterprise networks are based on a three-tiered architecture called the Hierarchical Model for Internetwork Design [1]. A slightly modified version of this architecture is shown in Figure 2 with the three layers: core, distribution, and access. The core and access layers correspond to Data Link Layer of the Open System Interconnections (OSI) model [2], and the distribution layer corresponds to the Network Layer of the OSI Model, respectively. The core layer separates the enterprise infrastructure backbone (WAN, Internet, etc.) from the end-user areas with workstations, application servers, databases, and other services.

* Other brands and names are the property of their respective owners.

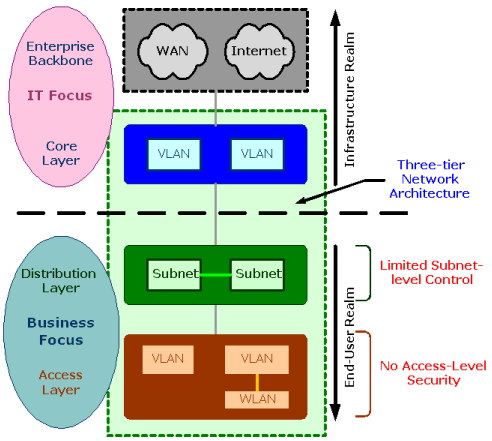

Figure 2: The enterprise network architecture

As pointed out in Figure 2, there are no elements of security in this model as any LAN device can connect and freely communicate with other LAN devices. Additionally, a TCP/IP device can connect and freely communicate with other TCP/IP devices anywhere in the enterprise network.

The Security Capabilities

The two key security additions to the network architecture are as follows:

1. *Access layer capabilities*. These focus on knowledge of the device connecting to the LAN or knowledge of which LAN devices can communicate with each other.
2. *Distribution layer additions*. These focus on subnet-to-subnet-level application-level control.

Access Layer Security Capabilities

There are three forms of access layer security.

MAC Address Filtering

The simplest of the access layer security capabilities is based on knowing the MAC addresses of all LAN devices and unknown MAC addresses not being permitted to join the VLAN. Most enterprise-level Layer 2 switches support MAC address filtering. This approach is clearly not scalable for the enterprise networks but suitable for very small areas. Additionally, with the prevalence of notebook computers with removable network interfaces, a known MAC address does not provide sufficient access layer authentication.

Port-Based Access Control (802.1x)

The IEEE 802.1X standard [3] offers both wired and wireless devices a method to authenticate the device and the user of the LAN device before connecting to the VLAN. Based on the Extended Authentication Protocol (EAP), the 802.1X standard routes the EAP network traffic to a RADIUS server [4], the authentication server. Only authenticated users and devices are allowed to connect to the VLAN. All other devices are not allowed to connect to the enterprise. Device authentication can include local security-level checks such as operating system updates and virus signatures. While IEEE 802.1X is more scalable than MAC address filtering and takes network port security to a new level, it is considerably more complex and expensive because it requires the support for the standard from the Layer 2 switches and the operating systems. Legacy operating systems have very limited support for IEEE 802.1X.

Super and Sub VLANs

RFC 3069 [5] introduced the notion of super-VLANs and sub-VLANs to realize the optimization of IP addressing in a switched environment. Each sub-VLAN has its own broadcast domain while using the default gateway IP address from the super-VLAN. A leading network vendor has added two security capabilities in the sub-VLAN and super-VLAN space, called secondary and primary VLANs, respectively [6]. First, the secondary VLANs can be either in an isolated mode where the members cannot communicate with each other or the community mode where the members can communicate with each other in a peer-to-peer fashion, with the primary VLAN providing the default IP gateway access. Super-VLANs and sub-VLANs extend port-level security by creating communities that are allowed to communicate with each other and denying communications to all others. Figure 3 shows all three of the access layer security capabilities.

Distribution Layer Security Capabilities

Packet filtering is the common term for distribution layer security. Packet filtering provides subnet-to-subnet-level network access control. Firewall devices, routers, and Layer 3 devices use the network layer information to allow or deny access to TCP/IP devices. There are two types of packet filtering:

1. Static packet filtering
2. Dynamic/stateful packet filtering

Static Packet Filtering

Static packet filtering provides us the ability to control the source and destination of the network traffic with application access through TCP/IP ports. Static packet filtering rules are applied at network and transport layer headers only. Most routers and Layer 3 switches provide static packet filtering. Typical packet filtering includes “permit” and “deny” rules. Static packet filtering also includes the shielding of internal IP addresses through Network Address Translation (NAT) [7].

Dynamic/Stateful Packet Filtering

Dynamic packet filtering adds additional intelligence to the static packet filtering. Static packet filtering allows the entire dynamic port range (e.g., greater than 1023) during a client-server session. Dynamic packet filtering, on the other hand, knows to look into the data and find the client port. The dynamic packet filtering then allows only that client port in real time for that session, as opposed to static packet filtering, which opens the range of client ports (e.g., > 1023). Dynamic packet filtering prevents security attacks based on hijacking of established sessions. Moreover, dynamic packet filtering can add time sensitiveness by opening and closing ports on an as-needed basis. Various types of dynamic packet filtering are available today. Stateful inspection [8], for example, matches the HTTP protocol-to-protocol headers along with send-receive pairs and data types as specific in the header. Circuit filtering [9] permits inspection of sessions, as opposed to connections or packets, with each session containing a number of connections. Circuit-level filtering takes into account secondary connections such as the data part of the FTP protocol and streaming media. Application filters [10] or "proxies" are application-specific with access to details of the application-level commands. HTTP proxies can be set to deny the GET command but to allow the PUT command instead. Figure 3 shows the distribution layer security capabilities in addition to the access layer security capabilities.

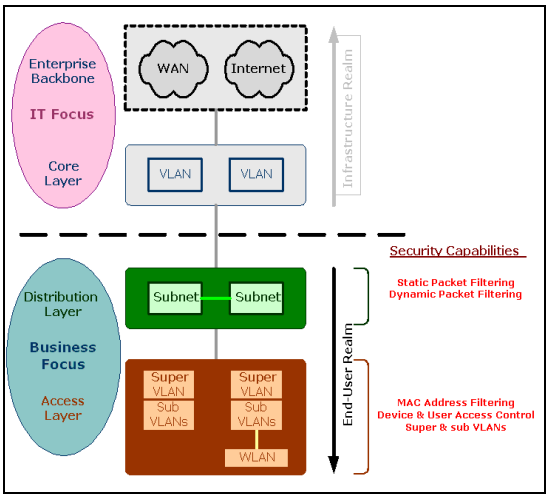

Figure 3: Distributed layer and access layer security

Policy-Enabled Network Security (PENS) Management

PENS [11] is an architecture developed by Intel Information Services and Technology Group (ISTG) that enables a common security policy specification across a heterogeneous enterprise network, and that allows correlation of abnormal events observed by other network and security-monitoring solutions such as intrusion-detection systems, vulnerability scanning tools, and incident alert services. In this architecture, policies are dynamically linked to the threat environment via an adaptive feedback loop.

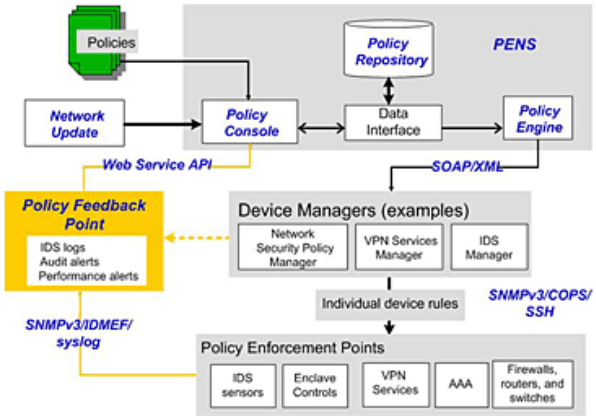

Figure 4: PENS architecture

As shown in Figure 4, the components of this policy-enabled management system include policy server, Policy Enforcement Points (PEPs), and Policy Feedback Points (PFP). As indicated, the main distinction of the PENS architecture from the standard policy-based network management architecture [12] is the introduction of PFP along with the adaptive feedback loop. A PFP collects data on intrusions, security alerts, and other abnormal network behaviors from a variety of systems (e.g., intrusion-detection systems, system performance logs, audit alerts, vulnerability scanners, etc.) and sends such data as feedback to the Policy Decision Point (PDP). The PDP then correlates the feedback data and determines if any policy updates are needed.

The main features for the above PENS architecture are as follows:

- A centralized view of the network from the management console, with real-time updates of the network.
- Automated management with an adaptive feedback loop that can dynamically modify the implemented security controls to address changing security threats.
- The ability to have common policies pushed from a central location to various network devices (PEPs) from different vendors.

Here the PENS server may act as manager of managers where the vendor-specific rules are managed by device managers and are therefore transparent to the PENS administrator.

Resulting Enterprise Network Security Architecture

Figure 5 provides an integrated view of the network and security architecture. Access layer security gives us the ability to authenticate devices and users connecting to the LAN. Device checking will give us the ability to ensure that unhardened systems considered insecure will not be permitted to connect to the enterprise network. Likewise, users unknown to the enterprise will not be permitted to connect to the enterprise network. Additionally, authenticated devices and users will be confined to the required communities of users and systems and denied access to everything else.

Distribution layer security provides subnet-to-subnet-level access control, enabling application-level control over the enterprise network.

The use of super-VLANs and sub-VLANs converts the three-tiered network architecture into a four-tiered network architecture integrated with security, making the Intranet more secure.

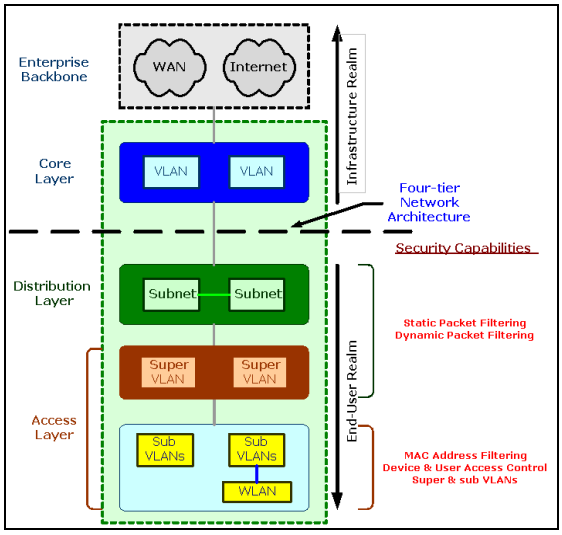

Figure 5: Four-tiered network architecture with security

With the addition of the required key security components to the Enterprise Network Architecture, we now turn our focus to the implementation challenges facing us, among which are significant resource requirements and consistent and comprehensive management of security policies. The following section addresses these challenges in detail.

Integrating Security with PENS

Some of the different ways of integrating the security capabilities proposed earlier with PENS are outlined below.

Location-Based Authentication

We can apply specific security policies to specific networks. For example, the authentication requirements for user access to applications in an office environment will be different from access to applications hosted in an Internet DMZ.

Behavior-Based Re-Authentication

It is possible to alter the authentication time limit after initial authentication is complete. This type of policy allows system administrators to control the authentication time limits when end users misbehave.

PENS can modify the authentication policies of a network and force a re-authentication, particularly when the network is under a cyber attack. It is also possible to specify a different policy for re-authentication dynamically.

Network Status-Based Re-Authentication Requirements

In the case of a serious cyber attack, the policies for infected networks could force all users to log off and disallow reconnection until a compliance scan shows that appropriate security updates are completed.

Authentication-Based User-Specific Policy Delivery

This scenario requires different authentication methods for different users followed by delivering different policies for the users. For example, company visitors have minimal authentication to our network and are allowed access to the Internet with no access to the Intranet. In another scenario a person from the security incident response team may be required to provide strong authentication but would then be able to connect to any network and have full access without being required to run a security scan.

IMPLEMENTATION CHALLENGES

The cost in human resources and technology to add security to network design, based on the new architecture, will not be trivial. The extent of the investment will be determined by the size of the enterprise (number of people, number of computers, number of subnets, number of locations, span of business group across geography, etc.). It is essential to have accurate knowledge of the security weakness in order to control the cost of implementation. The major issues are as follows:

1. Technology or technical issues
2. Business issues
3. Integration with legacy systems

Technical Issues

The availability of sub-VLAN and super-VLANs technology and 802.1x technology is limited. Not all

vendors provide this capability and in addition, not all product lines have the capabilities implemented fully.

Packet filtering requires the identification of all network communications paths. It is imperative to comprehend who needs to communicate with whom and with what applications. Understanding and documenting the various protocol interactions of applications is a very tedious but necessary process in successful implementation of packet filtering. Packet filtering rules can be easily created and managed with the knowledge of sources, destinations, protocols, and ports.

The complexity of the communications model leads to the next issue: the size of the packet filtering rule set. It is important to keep the packet filtering rule size fairly compact and simple. Larger rule sets could lead to over-utilization of the network devices (such as routers and switches) leading to performance degradation of the enterprise network [13].

Packet filtering additionally brings new operational challenges to network management. The Internet Message Control Protocol (ICMP) is considered insecure [14]. Therefore, utility programs such as "ping" and "traceroute" have limited use in this environment. Network operations will require the development of alternate network troubleshooting and debugging methods.

Business Issues

The use of new security capabilities require a significant investment in hardware. Port-level security along with super-/sub-VLAN technology requires the edge network equipment be switched as opposed to the older concentrators. Additionally, only limited products support these capabilities.

The next challenge is to group computers based upon the similarity of their communication requirements. As described before, this requires intimate knowledge of all the applications along with their source and destination TCP/IP parameters. One of the many complexities network engineers face is the accommodation of high availability key infrastructure services such as authentication, DNS, Web, database, and middleware within packet filtering rules.

Some of the operational challenges are outlined here:

1. Identifying and separating the servers and workstations of each business unit in the enterprise network because each business unit may have different security requirements.
2. The identification and subsequent separation of mission-critical services from regular services. This is particularly important because mission-critical services from different business units are required to run continuously on the network even if there is a large enterprise-level cyber attack, such as an SQL Slammer [15].
3. Operational knowledge of the impact of these mission-critical services on the functioning of the enterprise. For example, turning off the enterprise shipping application due to a virus attack may not impact the engineering business unit but may impact the distribution and manufacturing business units.

One of the challenges is keeping the training of IT professionals up to date on new security capabilities. Keeping the operations personnel trained is vital to maintaining service levels. In addition to trouble shooting and diagnostics skills for the various products, IT support personnel need to have working knowledge of the many applications and systems in the enterprise. Deeper knowledge of the system design and components is critical to satisfactory customer service.

Managing Legacy Environments

One of the biggest integration challenges in an enterprise has to be the co-existence with legacy applications and their infrastructures. To minimize the business impacts, changes are typically incremental and are managed through a migration process with well-defined phases. The new security capabilities will present numerous challenges as they will have to be integrated into the technology and customer impact areas of the existing business such as downtime and interruptions.

OUR RECOMMENDATIONS

We favor the divide and conquer approach by dividing the enterprise into several business units, such as Corporate IT, e-Business, Manufacturing, Engineering Design, etc. Each business unit has domains of end users, applications, servers, workstations, and networks, which are typically separate for each business unit. The business units may leverage one or more corporate IT services for authentication, Internet access, DNS, DHCP, NTP, etc. We have compiled the following prerequisites for adding security to the enterprise:

1. Identification of the all networks for the business units.
2. A list of all hardware and software running on the network along with the security configurations.
3. Identification of the mission-critical services in the business unit and the impact of unavailability of network services on the business unit.
4. The communication model for each of the applications running within the business unit including end-user information.

5. Working knowledge of security weakness in the business units.

We recommend that each business unit be in a security enclave [16]. By definition, a security enclave has well-defined boundaries and comprehensive security policies for network operations, based on the requirements of the business unit. For example, one business unit may require the use of corporate authentication whereas another business unit could use local user accounts and groups. In another example, a business unit may require conformance to corporate minimum security specifications for its servers and workstations, whereas another business unit may have legacy computer systems that may never be upgraded. Business-unit-to-business-unit communications can be published and managed through the appropriate configuration of their enclaves.

The usage of enclaves solves operational issue number 1 as described in the Business Issues section. Using PENS with enclaves, as described in Figure 6, solves operational issue number 3 described in the Business Issues section.

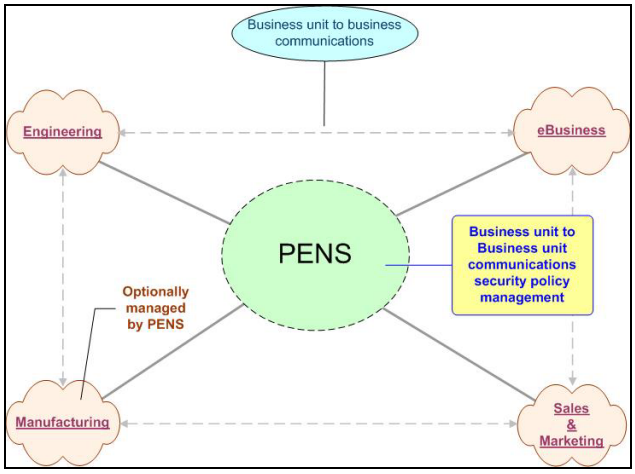

Figure 6: Using PENS to manage enclave-level communications

Not all elements of security described in this paper are required for every network environment. For example, 802.1x port security is a better fit for dynamic environments such as office users with remote access or wireless access. MAC address-based port security is more suitable for data centers and factory environments than 802.1x-based port security. Likewise, we also recommend super-VLAN and sub-VLAN technology for static environments such as data centers and factories but not for office environments. Table 1 lists our recommendations.

Table 1: Security recommendations

Environment	Security Model
Dynamic end-user such as office networks	802.1x
Static environment such as business critical systems (e-Business, Shop Floor)	Sub/super VLAN, static MAC address, packet filtering
Lab/development environment	Packet filtering
Partner, supplier, customer	Packet filtering
Remote access	802.1x and dynamic packet filtering

Lastly, we recommend allowing hardened systems to connect to the enterprise. There are several standards for defining hardened systems [17, 18]. Hardened systems contribute to security by controlling it at the source. If a service is not needed and turned off, it cannot be exploited over the network.

CONCLUSION

Until recently, security implementations have been reactive, their modus operendi being to keep current with the latest software updates and virus signatures. We have made security a proactive activity with integration of key capabilities into the network architecture so that networks and systems are delivered with security capabilities in place from the very outset. We have implemented several such enterprise networks in business-critical environments with no virus/worm infections in those areas in 18 months. The approach has been successful in letting our business units perform what is required of them while eliminating unauthorized and random access and probing. Adding new securities to an enterprise has the potential to increase operational costs. We are now focusing on putting together proper adaptive and dynamic policies that enable us to manage the security of the enterprise cost effectively without compromising the security of the enterprise.

ACKNOWLEDGMENTS

We thank Gary Morris, Shane Milburn, Kevin Heine, Amit P. Shah, Colm O'Halloran, Clark Mason, Mark Sokol, Anand Rajavelu, Praveen Sampat, and Richard Phillips for the various security capabilities that have been implemented in manufacturing. We also extend thanks to Ravi Sahita and Satyendra Yadav for their contribution to the development of policy-enabled architecture; and to Jonathan P. Clemens, Greg Tao, and Sridhar Mahankali

for reviewing this paper. We also appreciate the contribution of the ITJ editorial staff.

REFERENCES

[1] "Internetwork Design Guide," Cisco Systems, http://www.cisco.com/univercd/cc/td/doc/cisintwk/idg4/*

[2] "Open System Interconnection Reference Model," Cisco Systems, http://www.cisco.com/univercd/cc/td/doc/cisintwk/ito_doc/introint.htm*

[3] "Port-Based network Access Control," http://standards.ieee.org/getieee802/download/802.1X-2001.pdf*

[4] "RADIUS Protocol Security and Best Practices," Microsoft Corporation, http://www.microsoft.com/windows2000/techinfo/administration/radius.asp*

[5] "RFC 3069–VLAN Aggregation for Efficient IP Address Allocation; Internet RFC/STD/FYI/BCP Archives," http://www.faqs.org/rfcs/rfc3069.html*

[6] "Cisco–Securing Networks with Private VLANs and VLAN Access Control Lists," Cisco Systems, http://www.cisco.com/warp/public/473/90.shtml*

[7] "RFC 1631–The IP Network Address Translator (NAT); Internet RFC/STD/FYI/BCP Archives," http://computer.howstuffworks.com/framed.htm?parent=nat.htm&url=http://www.faqs.org/rfcs/rfc1631.html*

[8] "Stateful Inspection–Webopedia.com," http://networking.webopedia.com/TERM/S/stateful_inspection.html*

[9] "Computer Security Dictionary: Packet Filtering (screening); ITsecurity.com," http://www.itsecurity.com/dictionary/packfilt.htm*

[10] *Building Internet Firewalls*, Chapman & Zwicky, O'Reilly & Associates, Sebastopol, 1995.

[11] Hong Li, Ravi Sahita, Greg Kime, Jac Noel, and Satyendra Yadav, "Policy-Enabled Network Security with Adaptive Feedback Loop and Capability-Based Data Model," *Eurescom 2003*, September 2003.

[12] D. Verma, "Policy-Based Networking, Architecture and Algorithms," *New Riders*, November 2000.

[13] "Understanding ACL Merger Algorithms and hardware Resources on Cisco Catalyst 6500," Cisco Systems, http://www.cisco.com/en/US/customer/products/hw/switches/ps708/products_white_paper09186a00800c9470.shtml*

[14] "A summary of DoS/DDoS Prevention, Monitoring, and Mitigation Techniques in Service Provider Environment," Michael Glenn, *SANS Institute,* 2003. http://www.sans.org/rr/papers/70/1212.pdf*

[15] "CERT Advisory CA-2003-04 MS SQL Server Worm," http://www.cert.org/advisories/CA-2003-04.html*

[16] "DoD Electronic Business Architecture," United States Department of Defense, http://www.amc.army.mil/amc/ci/matrix/documents/dod/jta31e.pdf*

[17] "Center for Internet Security," http://www.cisecurity.org/*

[18] "The SANS Institute (SANS)," http://www.sans.org*

AUTHORS' BIOGRAPHIES

Sanjay Rungta is a staff network engineer with Intel's Information Services and Technology Group. He received his B.S.E.E. degree from Western New England College and his M.S. degree from Purdue University in 1991 and 1993, respectively. He is lead architect and designer for the Local Area Network for Intel. He has over 11 years of network engineering experience with three years of experience in Internet web hosting. He holds one United States patent in the area of Network Engineering. His e-mail is sanjay.rungta at intel.com.

Anant Raman is a staff engineer in Components Automation Systems with Intel's Technology and Manufacturing Group. He received his B.Tech degree from the Indian Institute of Technology, Bombay in 1981, a Master of Science in Mechanical Engineering in 1984, and a Master of Computer Science in 1991 from Arizona State University. He is the lead architect and designer for the e-Diagnostics infrastructure and manufacturing security initiatives within Intel. He also chairs the ISMT e-Diagnostics security sub team that is responsible for developing e-Diagnostics security guidelines for the industry. He holds three United States patents in the areas of software, network engineering, and security. His e-mail is anant.raman at intel.com.

Hong Li is a senior researcher with Intel's Information Services and Technology Group, responsible for trustworthy and survivable systems research. She led the development of several IT security strategies and architectures. She is also active within the Intel and external research communities. She is a 2004 Santa Fe Institute Business Network Fellow. Hong holds a Ph.D. degree in Electrical Engineering from Penn State University. She is also a certified information systems security professional (CISSP). Her e-mail is hong.c.li at intel.com.

Manish Dave is a staff network engineer with Intel's Information Services and Technology Group. He is lead engineer and designer for the Internet Connectivity and external network connectivity for Intel. He has over ten years of network engineering experience and network security experience. His e-mail is manish.dave at intel.com.

Greg Kime is a security architect with Intel's Information Services and Technology Group. He is responsible for the development and fostering of long-term security strategy and architecture. His focus is on wired, wireless, and platform security architectures along with process and governance development in a program called Enclaves. He is a CISSP. His e-mail address is greg.kime at intel.com.

Toby Kohlenberg is a senior information security technologist for Intel's ISTG Risk Management group. He has extensive experience in penetration testing, incident response, architecture design and review, and IDS, among others. In the last couple of years he has been responsible for developing security architectures for Intel's deployment of secure WLANs and Windows* 2000/Active Directory, and for the overall IDS strategy including the Security Operations Center. He is a handler for the Internet Storm Center and a co-author of the book *Snort 2.1* from Syngress. He currently is responsible for providing information security consulting to Intel product groups as well as evaluating new and emerging technologies. He currently holds the CISSP GIAC Certified Firewall Analyst (GCFW), Certified Incident Handler (GCIH), and (GIAC Certified Intrusion Analyst (GCIA) certifications. His e-mail is toby.kohlenberg at intel.com.

* Other brands and names are the property of their respective owners.

THIS PAGE INTENTIONALLY LEFT BLANK

Enterprise Client Management with Internet Suspend/Resume

Michael A. Kozuch, Corporate Technology Group, Intel Corporation
Casey J. Helfrich, Corporate Technology Group, Intel Corporation
David O'Hallaron, Carnegie Mellon University
Mahadev Satyanarayanan, Carnegie Mellon University and Intel Corporation

Index words: virtual machine, distributed storage, migration, management

ABSTRACT

Internet Suspend/Resume (ISR) is an exciting new model for managing client machines in the enterprise. ISR provides the administrative benefits of central management without sacrificing the performance benefits of thick-client, personal computing. This capability is made possible through the novel combination of two well-understood technologies: virtual machines and distributed storage management.

With ISR, a user's entire personal computing environment, including the operating system, applications, data files, customizations, and current computing state, is maintained in centralized storage. By leveraging virtual machine technology, this computing environment may be transported through the network and rapidly instantiated on any ISR-enabled client machine. The central management may include automatic backup, virus scanning, and maintenance.

Further, the ISR software stack is naturally partitioned into two parts: the ISR base and the user environment. The ISR base, which comprises the virtual machine monitor and management tools, runs directly on the physical hardware and is centrally managed by the enterprise Information Technology (IT) department. The user environment is the familiar software stack, which comprises the operating system and applications; it may be managed by the end-user, the IT department, or both. This separation enables the user environment to rapidly migrate from physical platform to physical platform to recover from hardware or software failures, for example. This separation also enables the IT department to protect the enterprise network by quarantining badly behaving user environments.

INTRODUCTION

Over the past two decades, the advent of the Personal Computer (PC) has transformed the computing industry. Part of the PC's success in the marketplace is due to its *personal* nature. Individual users have their own hardware resources, manage their own software resources, and are able to customize their computing environment to suit their needs. In an enterprise environment, however, this aspect of personal computing often imposes a maintenance burden on the Information Technology (IT) department that manages the hundreds or thousands of machines that constitute the computing environment of the enterprise.

Internet Suspend/Resume (ISR) is a new technology that improves the manageability of computing environments without sacrificing the personal aspect of modern desktop computing. Moreover, ISR enables centralized-style maintenance of personal computing state while still preserving the performance benefits of thick-client computing.

Personal computing state, in the context of ISR, refers to the user's *entire* computing environment–including the operating system, applications, data files, customizations, and current execution state. The ISR system collects each user's computing environment into a set of files, called a *parcel*, and it maintains these parcels on network servers.

During normal operation, ISR is virtually invisible to the end user. In the context of a corporate campus, for example, as employees prepare to go home in the evening, they *suspend* the operation of their computers by clicking an icon on their desktop. This operation is very similar to closing the lid on a laptop in that the current execution state of the computer is collected. However, an ISR suspend operation also updates the master copy of the parcel, which is stored on a centrally managed ISR server.

Once the user's parcel is updated on the server, the corporate IT department is able to perform common maintenance tasks such as generating a backup copy of the user's environment, or scanning the parcel for virus signatures.

When the employee returns to work the next day, he or she is able to *resume* the execution of his/her computing environment. This operation will instantiate the user's computing environment on that client machine. Execution will resume with precisely the same state that existed at the time of suspend: the correct applications will be open, the user's data files will be open, and the cursor will be in the expected location.

Two properties are essential to the usefulness of ISR as a management tool. First, the IT department administers the personal computing state of every user according to a centralized style. Consequently, the IT department is able to perform traditional centralized maintenance tasks such as backup copy creation on the user parcels while still providing the high-performance end-user experience associated with thick-client computing. Second, the IT department, not the user, administers the low-level ISR software that runs on all client machines.

On an ISR client, the user's environment does not control the hardware directly. Instead, the ISR client program instantiates and supports the user's software. Because the function of this software is limited to supporting the user, or *guest*, software, it is relatively small and simple. Certainly, this program is much less complex that the guest software it supports.

Because the ISR client program is small and simple, it should be less error-prone and easier to maintain than the massive modern operating systems that IT departments currently manage on client machines. Further, the ISR client program does not have to be customized to the guest software that it supports. A single instance of the client program can support many different guest operating systems, for example. Therefore, the ISR client program represents a uniform computing base to the IT department.

Most importantly, the user never modifies the ISR client program. The ISR client architecture effectively divides the client software stack into two parts: the ISR client program and the user environment. The IT department exclusively manages the ISR client program. However, the user environment, which includes the traditional operating system and applications, may be managed either by the user or by the IT department, according to IT policy, just as in non-ISR computing systems.

We expect this division to be the right compromise between user control and IT control in many situations. This organization enables users to modify their software environment while still providing system administrators with a stable, uniform management platform.

ISR DESIGN

When designing the ISR architecture, we developed the following system requirements.

- The user's entire computing environment must be easily managed by the IT department.
- In the event of hardware failure, the user must be able to resume the environment on a new hardware platform that is not necessarily identical to the original.
- The system must enable *checkpointing* of the user's environment, and the user must be able to restore a previously saved checkpoint rapidly.
- Common use scenarios, for example when the user uses the same client platform every day, must perform well and must not impose undue load on the ISR server or network infrastructure.
- The system must support mobile platforms.
- The performance of the system must be comparable to the performance of a traditional (non-ISR) system.

Considering the above requirements, we developed a system design that combines virtual machine technology with network data transport. This combination of two old (at least, by computer science standards) technologies forms a new mechanism for managing personal computing state.

Virtual Machine Technology

Virtual machine technology is a well-understood field of computer science dating to the 1960s [4]. Historically, the term, *Virtual Machine (VM)*, referred to abstract software containers that mimic the operation of physical machines. Low-level software, called the *Virtual Machine Monitor (VMM)*, controls the operation of the VM, which includes virtual versions of the devices found in a physical machine (e.g., processors, memory, and disks). Guest software running within a VM container behaves as if it were running on a physical machine, and assuming that the VMM is sufficiently complete, the guest software will be unable to detect that it is interacting with a software container rather than physical hardware.

In the ISR system, VM technology performs several functions. First, the VM abstraction provides a natural interface through which the ISR system can collect the state of the user's environment. Because the VMM manages the entire operation of the VM, the VMM must also maintain the complete state of the VM. At suspend time, the ISR system simply stores a description of the

current VM state into the user's parcel. In particular, the ISR system collects this state at the virtual device level (processor state, memory state, disk state, peripheral state, etc.). Because the state of the environment is captured at the virtual hardware interface rather than some layer within the guest environment, the ISR system need not modify the guest software, nor does it need to be aware of which guest software is included in the user's environment.

The second role performed by the use of VM technology is isolating users' environments from differences in hardware platforms. For example, suppose that a user suspends her environment before leaving work at the end of the day and that, during the night, the user's hard drive crashes. Upon realizing this after arriving in the morning, the user could simply retrieve another ISR client platform from the IT department and resume her parcel on that machine. Resuming on anonymous hardware is possible because (1) the user's entire state is maintained on the server, and (2) the VM interface exported by the ISR client software is identical to the old VM interface, even if the underlying hardware platform is not. The VMM hides differences between the crash and recovery machines, and the user's software is unable to detect that the physical machine has changed.

VM technology provides a third benefit in that the VMM can support operations that manage the client hardware in a context outside that of the user's environment. For example, the VMM might manage the physical network interfaces in the client platform, and when guest software initiates traffic through the virtual network interface, the VMM controls the transfer of that traffic to the physical network device. To detect the infection of guest environments by certain computer worms and viruses, the corporate IT department could provide software in the VMM's virtual network component that monitors network activity originating in the user environment. If that activity resembles the traffic patterns corresponding to known worms or viruses, the VMM could filter the problematic network traffic at the client and outside the context of the infected environment. Naturally, the IT department could also alert the user and/or remotely administer the guest environment.

Network Data Transport

While VM technology satisfies our requirements for state encapsulation, hardware isolation, and client administration, it must be paired with network data transport in order to realize the full promise of ISR. Once the user's environment is captured into a parcel, for example, that parcel must be transmitted to a server for safekeeping.

ISR does not impose any specific requirements on the network transport protocol; any mechanism that efficiently transfers data packets from the client to the server and back should suffice. Example mechanisms might include distributed file systems, http-based transport, and session-based mechanisms such as File Transfer Protocol (FTP) and secure shell (ssh). In fact, while our early work [1] relied on a distributed file system, we have recently included support for an http-based system.

When installing an ISR system, the IT department may freely choose the data transport mechanism. However, the ISR installation team should carefully consider several factors before choosing a particular protocol. For example, most ISR files are tens to hundreds of kilobytes in size, and a few are tens of megabytes in size. Will the protocol efficiently transport medium to large files through the network? Are there any peculiarities (e.g., high latencies or asymmetric links) regarding the intended network that may affect the performance of certain protocols? Does the network include firewall, proxy, or Network Address Translation (NAT) machines?

Another factor that the ISR installation team may wish to consider is the physical location of the ISR servers. Because the ISR system maintains the entire state of each user's environment, it provides an attractive approach for enterprise disaster recovery. Site disasters, such as an office building fire, can be devastating to a business when the information maintained on personal computers is lost in the disaster. An ISR system could help an enterprise recover from a disaster, even when many personal computers are lost, provided that the ISR servers survive the disaster–by being replicated and/or physically separated from the clients they serve. In the same way that ISR can help a single user quickly recover from a hardware failure by resuming on an anonymous platform, ISR can help a hundred users quickly recover from a catastrophe by resuming their environments on a hundred anonymous platforms.

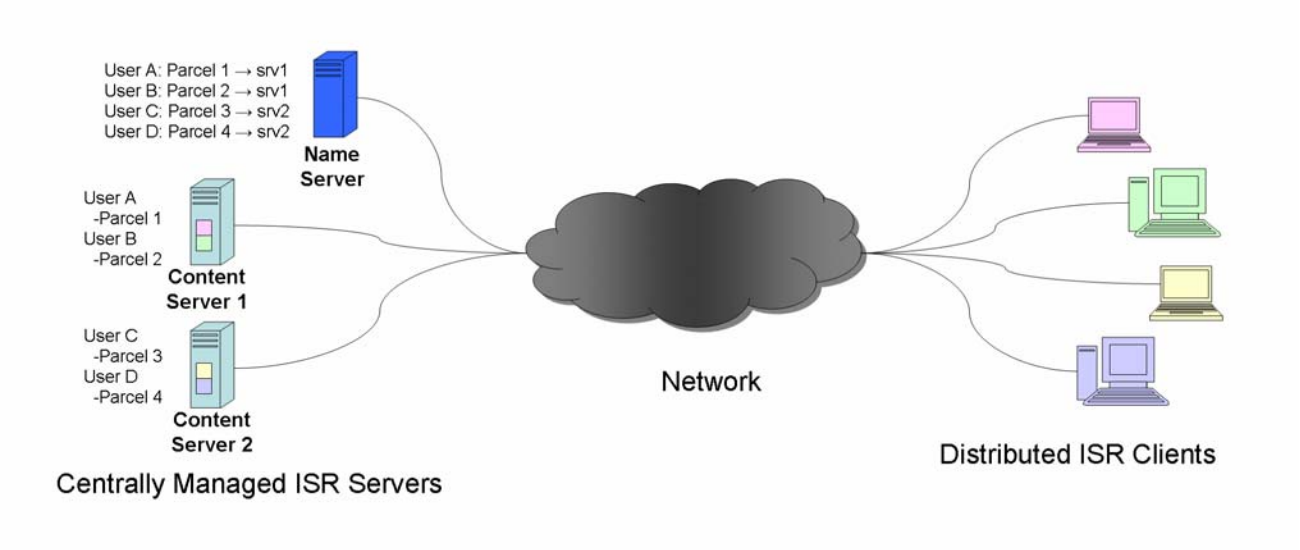

Figure 1: The ISR Network Architecture

SYSTEM ARCHITECTURE

Regardless of the network protocol employed, an ISR system, which is depicted in Figure 1, consists of three logical network entities: clients, nameservers, and content servers. An ISR client is the machine that the user interacts with directly and can be either a desktop or a laptop machine. Content servers manage the data associated with user parcels, and name servers provide a lookup facility through which ISR software can discover which content server is associated with a given user and parcel name. A typical enterprise installation of ISR would include many clients, several content servers, but possibly only one name server.

Client Architecture

Thus far, we have used the terms *resume* and *suspend* loosely to refer to the general operation of the system. In practice, the system includes several more operations and the terms resume and suspend refer specifically to starting and stopping the VM on the client. For example, before a particular environment may be resumed, the client must *check-out* the corresponding parcel from the content server, and after the environment is suspended, the client must *check-in* the parcel. From the user's perspective, the following are the essential ISR operations.

- *checkout* prepares the parcel to be run on the client by obtaining the necessary authentication tokens, decryption keys, software locks, and critical parcel data.
- *resume* resumes the VM and makes the user's environment available.
- *suspend* stops the execution of the VM and saves its current state to the parcel on the client.
- *checkin* saves the current state of the parcel to the content server.
- *discard* deletes the current state of the environment on the client without saving it to the server. This is occasionally useful if the user realizes that something unfortunate occurred during this resume session, such as virus contamination.
- *hoard* caches the entire state of the user's parcel on this client in order to prepare for a planned or possible network disruption.
- *ls* lists the state of all the user's parcels.
- *stat* prints information regarding the state of a particular parcel.

Each parcel may only execute on one client at a time. Hence, the operations checkout and check-in associate a given parcel with a particular client. The resume and suspend operations control the execution of the VM on which the user's environment runs. The discard and hoard operations control the caching of parcel data on a particular client, and the ls and stat operations provide feedback to the user regarding the state of the system.

These operations move the user's parcel through various states as shown in Figure 2. At a given client, the parcel is initially not present. When the user executes the checkout

operation, the parcel's critical data are stored on the client and, consequently, the parcel is logically present, but unmodified. Once the user executes the resume operation, the parcel enters the running state and the user may interact with the reconstructed environment. Upon suspend, the environment is halted and the new state of the parcel is stored locally on the client. Note that the modified state is not propagated to the content server until the user issues a checkin command. Consequently, the user may continue using the environment by resuming the current state, or the user may return to the last saved state by issuing a discard command.

The hoard command simply populates the client-side parcel cache. By design, the system is organized in a manner that permits the user to start a hoard operation while in either the unmodified or modified state.

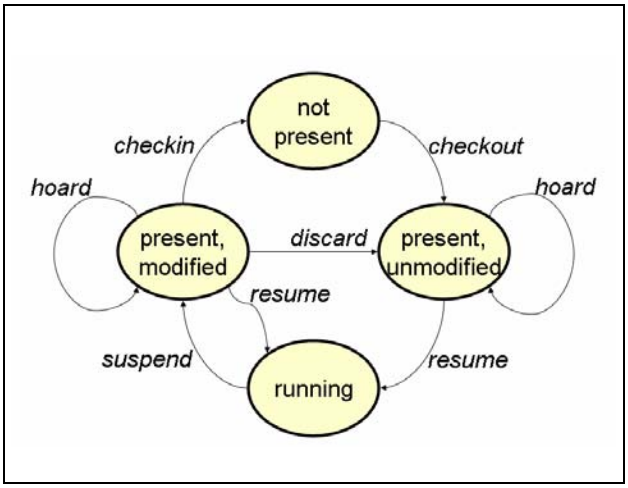

Figure 2: Client-side parcel state transition diagram

In our current implementation, these user commands may either be executed by the user on the command line or through the graphical user interface shown in Figure 3.

Figure 3: Prototype ISR graphical user interface

These operations are implemented through the orchestration of the three components that constitute the ISR client software depicted in Figure 4. The VMM supports the operation of the user environment, which includes the guest operating system and guest applications. The state files that compose the user's parcel are stored in the parcel cache and managed by the parcel cache manager. The VMM accesses these files on demand while the user's environment is running. The ISR client manager orchestrates the movement of data between the parcel cache manager and the content server during the operations listed above.

For example, during the checkout operation, the client manager first fetches a configuration file describing the location and organization of the user's parcel. With that information, the client manager begins to fetch the associated data over the network and primes the parcel cache manager before invoking the VMM to instantiate the user environment.

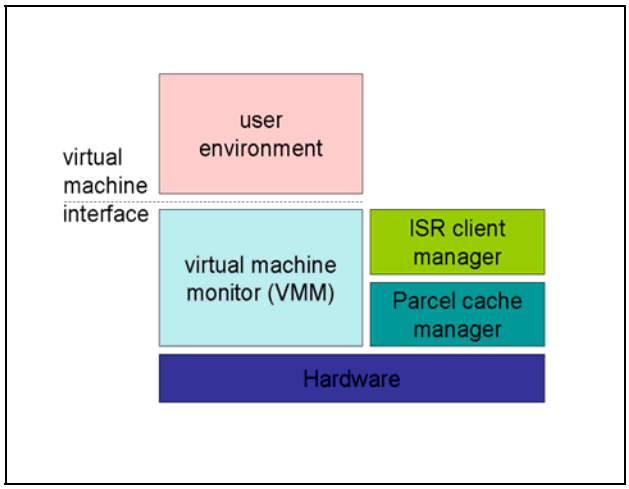

Figure 4: ISR client program

The parcel cache manager is responsible for organizing all parcel data on the client. The ISR design comprehends on-demand fetching of parcel data. That is, the client manager is not required to fetch the entire parcel before invoking the VMM to restore the user's environment. Instead, the client manager need only fetch the critical files needed to resume the VM. As the environment executes, the VMM submits data requests to the parcel cache manager. If the cache manager finds that a requested file does not exist in the cache, it fetches the file from the content server and installs it in the cache before supplying the requested data to the VMM.

In our ISR prototype, the parcel cache manager organizes parcel data into a traditional file hierarchy as depicted in Figure 5. This figure shows an example of the files that will be present on a client machine when a parcel is checked out and in use. The first line in Figure 4 describes

the root of the directory used for this parcel. The directory contains two sub-directories: *last* and *cache*. The last directory contains read-only data describing the state of the parcel at the time of checkout. As the user's environment is running, changes to the environment and files fetched to satisfy demand misses are stored in the cache directory.

In particular, as the client uses the parcel, the *cache/disk* directory becomes populated, as parts of the disk are demand fetched. The client can also explicitly populate this directory with the hoard command. Until this directory is fully populated, the user should ensure that the client machine remains connected to the network because the client must be able to satisfy demand misses from the content server. However, once the disk directory is fully populated, the client can operate in disconnected mode and requires no further access to the network until the user wishes to check-in the parcel.

The hoard command enables the user to prepare a client for disconnection. This function is particularly useful for preparing a laptop before removing it from the network. The hoard command and the discard command both operate on the client's cache directory. Hoard causes the cache manager to populate the cache fully. Discard causes the cache manager to delete the cache without checking it in to the content server, thereby irrevocably destroying the data.

```
/home/[username]/.isr/[isr username]/[parcelname]/
        |-->/last/memory
        |-->/last/meta
        |-->/last/keyring (old)
        |-->/cache/disk/
        |-->/cache/memory
        |-->/cache/meta
        |-->/cache/keyring (current)
```

Figure 5: Client parcel cache example

Name Server Architecture

Each session on a client typically begins with a checkout operation, which, in turn, begins by contacting the ISR name server to determine the location of the user's parcels. The role of a name server is similar to that of DNS servers on the Internet. The name server is the only machine that a user needs to identify to the ISR client software, as the name server contains information that describes the location of all other content.

Each username within a name server must be unique because a <*name server, username*>-tuple identifies a single user in the ISR system. Further, the system must ensure that each parcel name is used only once for each user as a <*name server, username, parcel name*>-tuple uniquely identifies a parcel.

When the user initiates a checkout operation, the ISR client software executes an authentication sequence with the name server and fetches the metadata associated with that parcel. The metadata identifies the content server(s) responsible for maintaining the parcel. The client software uses the retrieved metadata to contact the appropriate content server and begin fetching data.

In some implementations, such as ours, the parcel metadata may also include encryption keys required for decrypting the content on the content server. Our current implementation of the name server is a hardened Linux* computer running sshd, the secure shell (ssh) daemon. The client program contacts the machine via ssh and downloads a single file called *parcel.cfg*. Parcel.cfg is a small protected text file that contains all of the necessary configuration data describing a parcel, including the protocol used to access the content, the path/URL of that content, and the master decryption key, called the *keyroot*. The parcel.cfg file can also be extended to contain any protocol-specific information.

Content Server Architecture

As mentioned previously, the ISR system imposes very few requirements on the operation of the content server. This component can be implemented as a distributed file system server, database server, or other distributed storage system such as a distributed hash table-based service.

Essentially, this component need only be able to deliver data files as ISR clients request them. However, many implementations will naturally organize their content as described in this section. Understanding the structure of a parcel is essential to understanding the structure of the content server. An ISR parcel contains three logical types of data: memory, disk, and metadata.

The memory data corresponds to the state of the physical memory in the VM. For example, if the user's VM includes 512 MB of RAM, this file describes the 512 MB state of physical memory. In practice, however, we have observed the size of the memory file to be typically half the size of the virtual machine RAM once it is compressed, encrypted, and written to disk. As a special case, when the VM has been powered off, this file can be eliminated as the contents of the virtual RAM can be reset during the next resume operation.

The disk data represents the contents of the disk(s) in the VMs. Disk data can potentially be very large–several gigabytes to several tens of gigabytes, and because of this, much of the research in ISR has been focused on the

* Other brands and names are the property of their respective owners.

management of disk data. To enable quick access to portions of the disk and to support on-demand fetch, the disk state is divided into small files that are compressed individually. Each of these files represents a contiguous range of sectors on the virtual disk called a *chunk*. Each of these files is individually compressed and encrypted whenever they are not in use by the virtualization layer on the client. Carefully encrypting these files is particularly important because they may contain information of unknown sensitivity including protected guest files, the guest swap file, and the virtual disk metadata. The memory image must also be carefully encrypted for similar reasons.

The metadata associated with a parcel is a collection of very small configuration files used by the virtualization layer, various logs, and one file that contains the decryption keys for the disk data, called the *keyring*. For each of the disk chunks, the keyring contains two 20-byte entries. The first entry is the disk chunk's decryption *key* and the second entry is the disk chunk's lookup *tag*.

Keyring

The encryption technique that ISR employs for encoding disk chunks is convergent encryption [3]. Convergent encryption provides the useful property that two users who have the same file can both encrypt the file to the same cipher text without having to exchange any encryption keys.

This result is achieved by deriving the encryption key from the data, itself. If both users employ a cryptographic hash of the data as the encryption key, each will encrypt the original plaintext to the same cipher text. Each disk chunk's key and cipher text are derived in this manner:

$$key = hash(chunk)$$

$$cipher\ text = encrypt_{key}(chunk)$$

The motivation for using this encryption technique is the observation that many users' parcels may exhibit significant data similarity. For example, the parcels of users whose environments include the same operating systems and/or applications may have very similar contents. If these users are able to satisfy disk chunk requests amongst themselves, they will possibly (a) reduce the load on the server, (b) reduce the load on the network, and (c) reduce the average latency of chunk requests experienced by the users.

To enable inter-user data exchange, the users must not only encrypt their data in a common manner, they must also maintain a means for addressing those blocks. Hence, the keyring file also includes a tag for each chunk such that the tag is a hash of the cipher text:

$$tag = hash(cipher\ text)$$

The tag now uniquely identifies a disk chunk. An ISR client can request disk chunk files using this tag and decrypt those files using the associated key.

As an example, suppose that user A is working on a client machine that does not contain a fully populated cache for user A's parcel. User B uses the same operating system and applications as user A, and user B's client machine is in an adjacent office. Before resuming the VM, A's client has fetched the keyring associated with A's parcel. At runtime, if A's cache manager determines that a particular chunk file is needed, it can lookup the tag associated with that chunk in the keyring and request the file by tag value from B's client (by using, for example, the techniques reported in [2]). If B's client returns the file, A's client can decrypt it using the associated key.

Note that while clients do advertise the chunks being sought using the tag associated with that chunk, the tag does not reveal any information regarding the contents of the file and cannot be used to decrypt the file. Only the keys can be used to decrypt the chunk files, and these are never exchanged between clients. Instead they are stored in encrypted form in the keyring on the content server and only decrypted on the client. The encryption key for the keyring is the only secret a user will need to unlock access to his system. This secret is part of the parcel.cfg file that the client obtains from the name server during a checkout operation.

While initially intended as a mechanism for maintaining the association between chunk files and their tags and keys, the keyring has proven to be a surprisingly useful tool. For example, the keyring file allows for quick calculation of disk changes by comparing a newer keyring with an older one. Because each keyring includes a single entry for each disk chunk, by comparing the two keyring files entry-by-entry ISR software can determine which of the disk chunks has been modified. This technique is used in our implementation of the checkin operation to determine which disk chunks need to be uploaded from the client to the content server.

Content Server Structure

The structure of the content server will vary slightly, depending on which protocol it services. Figure 6 depicts an organization that should be useful for http- or distributed-file system-based content servers.

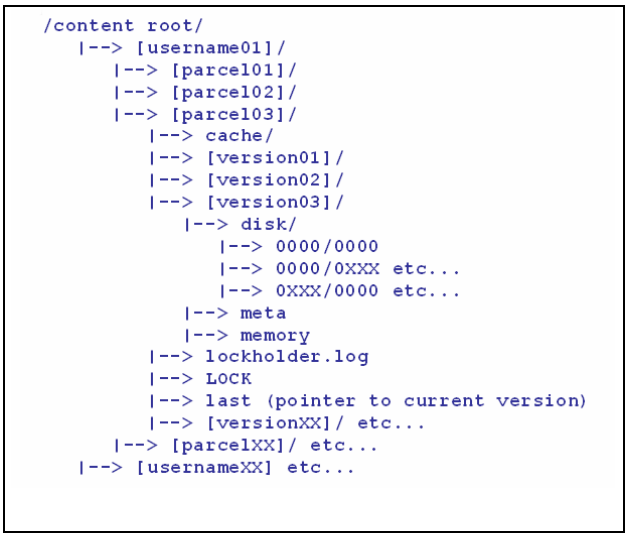

Figure 6: Content server data organization

Figure 6 depicts a content server that provides service for several users and permits multiple parcels per user. Each parcel can have many versions; each version represents the state of the machine at a checkin time during the life of the parcel. Each version contains enough information to recover the state of the parcel at that time. The version pointed to by "last" contains a fully populated parcel, while the previous versions only contain the delta information describing what virtual machine state changed between it and the subsequent version.

Parcel Versioning and Rollback

Rollback is an operation that enables a user to revert the state of his/her environment to the state associated with a previous checkin point. For example, if a user realizes that she introduced a computer virus into her environment on Wednesday, the user could employ the rollback operation to restore the state of the parcel to the state corresponding to some version prior to Wednesday's.

To support this feature, the ISR content server maintains a version of the user's parcel for every possible rollback point. Naturally, naively maintaining many copies of the parcel would require an excessive allocation of storage space. Instead, we can capitalize on the fact that a parcel typically does not change much from one version to the next. Consequently, each version directory only contains the data that changed between that version and the next. We expect the delta-encoding format to provide a space savings of between one and two orders of magnitude.

The number of versions maintained per parcel and the timing (weekly, monthly, etc.) of those versions are policy decisions set by the content server administrators. Fortunately, the delta-encoding format of the versions enables the system administrator to collapse several versions when it becomes necessary to reclaim space. Consequently, the policy may be dynamic; when the available disk space on the content server falls below some threshold, a reclamation process may iterate through all the parcels collapsing versions until sufficient free disk space becomes available.

Data Transport Protocol

As mentioned previously, we have experimented with two different content server protocols. The first protocol requires that the ISR content server data appears to reside in the local file system of the host. The data could either be truly local (on an attached portable hard-drive, for example), or they could appear to be local (in a mounted distributed file system, for example). In either case, the ISR client software accesses the data through simple file operations such as open, close, read, and write.

In the second protocol, the disk data are logically remote, and the ISR client software must fetch the data explicitly from a remote server via http or ssh. When a section of disk is requested and it is not yet cached on the client, the chunk files are explicitly requested from the content server and cached on the client to service future access requests.

The system currently supports both approaches. During a checkout operation, the client fetches the parcel.cfg file that describes which transport protocol is used for this parcel. Further, we have defined an abstraction layer in the client software so that the client can switch between the various transport protocols by calling into dynamically loaded libraries. This mechanism also supports the development of new transport mechanisms.

CHALLENGES AND SOLUTIONS

In developing the ISR architecture and initial implementation, we encountered a number of challenges. In this section, we list several of these challenges and our current solutions.

Large Environment State

By far, the greatest challenge that ISR presents is the enormous volume of data associated with a parcel. In particular, the virtual RAM is typically on the order of a hundred megabytes and the virtual disk drive is on the order of gigabytes. A naïve approach for data movement, transferring the complete image between client and server during every checkout and checkin operation, would be impractical even over fast corporate networks. Transferring ten gigabytes over a 100 Mbps network requires at least 800 seconds (15 minutes).

We have already described several of the techniques employed to reduce the impact of the state size. First and foremost, we have organized the parcel so that the disk data may be fetched on demand. This reduces the volume of data that must be fetched when performing a checkout

operation on a client with an empty cache to the size of the memory image rather than the disk image. Further, we employ standard compression to reduce the footprint of the memory image. If the compressed memory image is 100 MB, fetching the image over a 100 Mbps network will require at least eight seconds. In practice, the startup sequence including authentication, download, decryption, and decompression requires approximately 30 seconds.

Each of the disk chunks may be fetched separately, and consequently, each is compressed and encrypted separately. In our implementation, we have chosen a chunk size of 128 KB. The compression ratio can range from 0% to 100%. Portions of the disk that are empty, because they have not been used by the guest operating system, compress to 152 bytes while portions that contain nearly random data are uncompressible.

The chunk size is a parameter chosen by the system administrator for each parcel. The chunk size is essentially the cache line size for the parcel cache on the client. Choosing a larger chunk size will typically reduce the number of misses observed by the cache but will also increase the network bandwidth consumed by the client. Our experiments indicate that chunk sizes in the range of 64 KB to 256 KB are reasonable.

To compensate for potentially poor network performance, we rely heavily on client-side caching. Every chunk fetched by the client software is placed in the client parcel cache. Further, because this cache is disk-resident and disk space is relatively inexpensive, we assume that evictions will be relatively rare events. We also define the hoard operation to provide good performance in cases where the use of a particular client is known *a priori*. This operation can also be issued remotely. If a user knows that she is going to a remote site with low network bandwidth, a hoard operation can be issued remotely to ensure that data are prefetched into the remote client cache while the user is in transit.

For situations in which the network conditions are poor, and the location could not have been predicted, we have developed an optional performance enhancement to ISR called Lookaside caching (LKA) [5]. This technique relies on the user carrying a portable storage device such as a flash memory device to maintain compressed and encrypted copies of parcel data. When the user attempts to access his environment from an ISR client, the ISR client software can fetch necessary data from the device rather than from the content server. However, we require the client to verify the validity of the data on the portable device with the content server before using it. Again, we can rely on the keyring to provide this validation. Because the keyring contains the hash for every disk block, the client can simply hash data blocks on the portable device and compare the hashes to the keyring tag to determine if the portable device contains a correct copy of the data. In this way, the system never relies on the portable device; it can be lost, forgotten, or contain old data. The content server is always considered to contain the authoritative version of the user's parcel. Further, because the data on the portable device are encrypted, other users cannot make use of the portable device if it is lost or stolen.

Finally, we rely on delta encoding during both checkin and rollback operations as well as for reducing the amount of space occupied on disk by our parcel versioning system. For example, to represent a virtual machine with 256 MB of RAM and a 10 GB hard drive with Microsoft Windows* XP* and Microsoft Office* XP* installed occupies approximately 2.4 GB of server side storage after compression. For each rollback point that a user chooses to keep, delta encoding often reduces the server storage required to approximately 100 MB per version.

Heterogeneous Clients

Another potential impediment to widespread ISR adoption is the diversity of client machines found in the typical enterprise. Fortunately, the virtualization layer in the client software stack handles heterogeneous clients relatively easily.

In fact, virtualization is able to convert this challenge into a feature. If an enterprise adopts the ISR architecture, migration from one platform to the next becomes much easier. Currently, to migrate a user from one platform to a newer platform with a faster processor, either the IT department or the user must construct a new environment on the new machine and carefully consider which files and programs to install on the new machine before destroying the old environment. With ISR, simply suspend execution on the old machine, checkin the parcel, replace the hardware, checkout the parcel, and resume.

The one requirement that ISR imposes on the virtualization layer is that all client software provide the same VM interface. The VMMs on different clients need not be identical, but they must export the same VM interface. If the VMMs do not export the same interface, guest software may become confused when it resumes on a virtual machine that is different from the suspend site VM.

On the surface, differences in attached peripherals on various clients would appear to be a problem, but with the advent of solid support for plug-n-play devices in modern operating systems, this issue is largely resolved. Modern operating systems are able to handle the addition or

* Other brands and names are the property of their respective owners.

removal of plug-n-play devices (a USB printer, for example) cleanly and transparently. For example, suppose that the suspend site client is connected to a USB printer, but that no printer is connected to the client at resume time, the guest operating system will simply detect that the device has been unplugged and respond accordingly.

Similarly, modern operating systems are typically able to respond well to changing network conditions. When a modern laptop is disconnected from one network and reconnected to another network, the operating system is typically able to reconfigure the network stack automatically to compensate. In the same way, if a guest operating system within a VM detects that the network conditions changed between suspend and resume, the operating system will reconfigure networking to compensate.

Naturally, persistent network connections do prove to be a problem in the ISR system across suspend-resume cycles, but no more so than in general laptop usage. Persistent network connections are typically disrupted during laptop suspend-resume cycles due to time-out, address migration issues, or both. A suspend-resume cycle in the ISR system will cause similar results.

OVER-THE-WIRE MOBILITY

While this paper primarily discusses ISR in the context of enterprise management, the same infrastructure also supports over-the-wire mobility [1]. In particular, the same technology that enables a user to resume his or her environment on an anonymous hardware platform after a hardware failure enables the user to resume on a different hardware platform in the absence of failures. The user simply suspends on one client and resumes on another.

This capability is potentially very interesting in a number of situations. For example, some environments require a fluid office allocation; when employees arrive for work, they are given an office assignment out of a pool of available offices. ISR supports a clean mechanism for customizing the computers in those offices. The user simply sits down at the client machine and resumes his or her environment.

ISR could also be used to support employees who work at home some of the time and at work some of the time. Rather than carrying a laptop to and from work, the employee can leverage the network to transport her work environment from work to home and vice versa. This scenario is a particularly good application of prefetching as the employee may follow certain patterns such as arriving at work every day at 8:00 and leaving at 5:00. Such predictability greatly improves the effectiveness of the client parcel caches.

CURRENT PROJECT STATUS

We have built a prototype implementation of ISR and have been using the prototype internally for experimentation for some time. The prototype has become sufficiently robust that we plan to conduct a test deployment of the system with external volunteer users.

Through this test deployment, we hope to evaluate the following questions quantitatively:

- *How do users use the system?* How many times does the average user checkin, suspend, and discard? Are users doing something unforeseen and/or interesting with the system?
- *What are the hardware requirements for the system?* What is the load on the server? What is the load on the network? How much disk space is consumed per checkin?
- *How satisfied are users with the system?* Does the system fulfill an interesting need? Will users continue to use it? Will users recommend it to their friends?
- *Does ISR improve enterprise management?* Does system administration time increase or decrease in the context of ISR?

To answer these questions, we have instrumented both the server-side and client-side ISR code. The instrumentation will gather various statistics such as average number of bytes sent per checkin and frequency of checkin. The users will be fully aware that this information is being gathered, but we will try to protect their privacy by making the collected data anonymous and not observing activity *within* the guest environment.

We hope to start with approximately ten users and to increase that number to approximately 100 at the peak of the trial. Additionally, we hope to attract tolerant users in the early stages as we work out any bugs that remain at the beginning of the deployment. In the later stages, we plan to open the deployment to more novice computer users so that we can better evaluate a typical user's impression of the system.

CONCLUSION

ISR leverages virtual machine technology and network accessible storage to improve the management of enterprise clients by (1) cleanly separating the client software stack into a portion that is managed by IT and a portion that can be managed by the user, (2) providing centralized management of the entire user software environment, and (3) providing a mechanism for simple, rapid environment migration. Through this combination, ISR is able to simultaneously deliver the administrative

benefits of centralized management and the rich user experience of thick-client computing.

REFERENCES

[1] Kozuch, M., Satyanarayanan, M., Bressoud, T., Helfrich, C., Sinnamohideen, S., "Seamless Mobile Computing on Fixed Infrastructure," *IEEE Computer*, July 2004, pp. 65-72.

[2] Bressoud, T., Kozuch, M., Helfrich, C., and Satyanarayanan, M., "OpenCAS: A Flexible Architecture for Building and Accessing Content Addressable Storage," *2004 International Workshop on Scalable File Systems and Storage Technologies*, September 15, 2004.

[3] Douceur, J., Adya, A., Bolosky, W., Simon, D., and Theimer, M., "Reclaiming space from duplicate files in a serverless distributed file system," *Proceedings of the International Conference on Distributed Computing Systems (ICDCS 2002)*, Vienna, Austria, July 2002.

[4] Goldberg, R., "Survey of Virtual Machine Research," *IEEE Computer*, June 1974, pages 34-45.

[5] Tolia, N., Harkes, J., Kozuch, M., and Satyanarayanan, M., "Integrating Portable and Distributed Storage," *Proceedings of the 3rd USENIX Conference on File and Storage Technologies*, 2004.

AUTHORS' BIOGRAPHIES

Michael Kozuch is a senior researcher for Intel Corporation. Mike received a B.S. degree from Penn State University in 1992 and a Ph.D. degree from Princeton University in 1997, both in electrical engineering. Mike has worked for Intel research labs since 1997, four years in Oregon and three years in Pittsburgh, Pennsylvania. His research focuses on novel uses of virtual machine technology. His e-mail is makozuch at ichips.intel.com.

Casey Helfrich is a research engineer at the Intel Research Lab in Pittsburgh. He received a Bachelor's degree in Physics from Carnegie Mellon University in 2001 and an additional B.S. degree in Computer Science from Carnegie Mellon University in 2002. He joined the Pittsburgh lab at its inception and helped design and build the IT infrastructure for Intel Research. His e-mail is casey.j.helfrich at intel.com.

David O'Hallaron has been a faculty member of the Carnegie Mellon University School of Computer Science since 1989. His interests include high-performance distributed systems, Internet services, distributed search, and scientific computing. David has co-authored three books including *Computer Systems: A Programmer's Perspective* and won the 2003 Gordon Bell award for special achievement. His e-mail is droh at cs.cmu.edu.

Mahadev Satyanarayanan is the Carnegie Group Professor of Computer Science at Carnegie Mellon University. His research interests include mobile computing, pervasive computing, and distributed systems (especially distributed file systems). From 2001 to 2004 he was the founding director of Intel Research Pittsburgh, where the Internet Suspend/Resume project was initiated. He is a Fellow of the ACM and the IEEE, and the founding Editor-in-Chief of IEEE Pervasive Computing. His e-mail is satya at cs.cmu.edu.

THIS PAGE INTENTIONALLY LEFT BLANK

Advancements and Applications of Statistical Learning/Data Mining in Semiconductor Manufacturing

Randall Goodwin, Technology and Manufacturing Group, Intel Corporation
Russell Miller, Technology and Manufacturing Group, Intel Corporation
Eugene Tuv, Technology and Manufacturing Group, Intel Corporation
Alexander Borisov, Sales and Marketing Group, Intel Corporation
Mani Janakiram, Technology and Manufacturing Group, Intel Corporation
Sigal Louchheim, Information Services and Technology Group, Intel Corporation

Index words: statistics, machine learning, data mining

ABSTRACT

Knowledge Discovery in Databases (KDD) is the non-trivial process of identifying valid, novel, potentially useful and ultimately understandable patterns in data [12]. Sometimes referred to as Data Mining, Machine Learning, or Statistical Learning (which we will use in this paper), it has many diverse applications in semiconductor manufacturing.

Some of the challenging characteristics of semiconductor data include high dimensionality (millions of observations and tens of thousands of variables), mixtures of categorical and numeric data, non-randomly missing data, non-Gaussian and multimodal distributions, highly non-linear complex relationships, noise and outliers in both x and y dimensions, temporal dependencies, etc. These challenges are becoming particularly acute as the quantity of available data is growing dramatically. To address these challenges, statistical-learning techniques are applied.

We begin the paper with a description of the problem, followed by an overview of the statistical-learning techniques we use in our case studies. We then describe how the challenges presented by semiconductor data were addressed with original extensions to tree-based and kernel-based methods. Next, we review four case studies: house sales price predictions, throughput time prediction, signal identification/separation and unit-level speed prediction. Finally, we discuss how enterprise-wide statistical models form a foundation for intelligent, automated decision systems, and we describe applications currently under development within Intel Corporation.

INTRODUCTION

In 1965, Gordon Moore, one of the co-founders of Intel Corporation, observed an exponential growth in the number of transistors per integrated circuit and predicted the trend would continue. Moore's Law, as the relationship was named, has been maintained to the present time and is expected to continue through the rest of the decade. A plot of the relationship is shown in Figure 1 for Intel microprocessors.

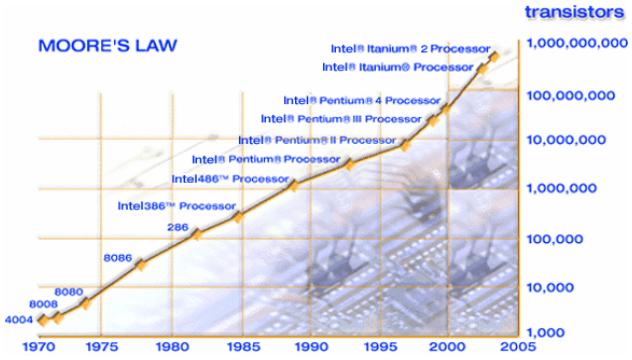

Figure 1: Moore's Law

To keep pace with Moore's Law, the semiconductor industry has relied upon many technical innovations and has grown significantly in complexity. The technological advances have been accompanied by an exponential growth in the quantity of data collected and stored during the manufacturing process. Examples of semiconductor manufacturing data include lot transactions and timestamps, process and equipment data, in line metrology data, electrical in-line test (e-test), wafer sort, and final electrical test/performance binning. Data may be at the lot level, wafer lever, or even the unit level. As Intel

Corporation ships several million microprocessors per quarter from its worldwide factories, the quantity of data stored in databases is now measured on the order of terabytes.

Traditional statistical approaches have served the industry well for many years and will continue to have an important place in semiconductor manufacturing. Often, a simple data plot, control chart, linear regression, or analysis of variance will tell an important story and enable the needed process discovery and controls. Design of Experiment (DOE) methods will also continue to play a key role in technology development and process health investigation/optimization. However, for multivariate/mixed data-type modeling of non-linear relationships, such as maximum functional frequency, lot throughput time and unit-level final bin classification, statistical-learning methods are required.

Our first case study is an illustrative one from outside the semiconductor domain on predicting the future sales prices of homes. When buying a home, consumers are guided by many factors such as location, square feet of the home, size of the lot, pool, etc. Real-estate agents and appraisers will readily provide "comps," short for comparisons, to existing homes that have recently sold to help the buyer (and seller) determine the value of a home. Traditional multiple regression techniques can be used to model home prices; however, using an advanced statistical-learning method, lower errors are achieved, modeling times are reduced, and requirements for domain-specific knowledge and modeling techniques are eliminated.

Our second case study focuses on a key measure of Fab performance–cycle time (CT), sometimes referred to as throughput time (TPT). In semiconductor manufacturing groups of wafers are processed together and move through the various processing steps in a Fab lot. Given the implications of multiple lot reentrant steps, a high mix of different products, lots with different priority flags, equipment preventive maintenance schedules, set-up times, etc., the ability to understand and predict lot TPT would immensely benefit factory planning and scheduling. Traditional approaches for predicting the remaining TPT of a lot include static linear modeling using Little's Law [13] and simulation techniques. However, static models do not handle stochastic variables, and simulations require large amounts of computing resources in order to predict the remaining cycle time. In this example, data-mining/statistical-learning methods use historical data to predict individual lot cycle-time by comparing key characteristics of a lot in progress to lots that have completed the target prediction operation. Results show that application of pre-clustering and a single decision tree results in better TPT predictions than with previous approaches.

The third case study focuses on utilizing statistical-learning technologies for signal identification/separation and for Advanced Process Control Systems (APCS). In this case study we are interested in identifying the key sources of variation of Sort Fmax–a measure of the maximum functional speed (frequency) of an individual microprocessor while still in wafer form. We are not only interested in identifying, ranking, and separating the sources of non-random variation, but also determining how the effects change over time. In our example, we use simulated data typical of those found in high-volume CPU manufacturing. Small Fmax "signals" were embedded in several simulated operations on a baseline normal variation (noise). By using advanced tree-based statistical-learning techniques and our newly developed extensions, we are able to separate out the subtle signals affecting Fmax not possible with traditional statistical-learning approaches or commercial data-mining software packages.

The final case study focuses on predicting the speed of individual microprocessors at the final test step. Natural process variation and both non-random and random defects result in microprocessors that are functional at different speeds. Predicting the final test outcome and speed bin at the unit level early in the flow presents unique challenges to traditional approaches; yet, such a prediction would yield clear benefits in planning, downstream test flow optimization, signal identification, yield improvement, advanced process control, and final assembly optimization strategies. A key enabler to this application has been the ability to trace individual units through key operations in the entire Fab and assembly test manufacturing flows. Unit-level numeric and categorical data (variables) form a basis for both training the models (when the final class test outcome is known) and predicting the final class test result of upstream units. Since the data are at the unit level, and hundreds of variables (or even thousands) are used for generating predictions, the dimensionality of the data sets is very large. Furthermore, the data often contain missing values from dynamic sampling schemes, outliers, temporal relationships, non-linear relationships and even dynamic sets of important variables. The novel extensions to tree- and kernel-based statistical-learning methods enable robust, accurate unit-level bin classification nearly impossible with traditional approaches or commercial data-mining software packages.

OVERVIEW OF STATISTICAL-LEARNING METHODS

In statistical- or machine-learning methods we are usually given an object with a set of variables/attributes, often

called "inputs" or "predictors" and a corresponding target, often called "response" or "output" values. The goal is to build a good model or predictive function capable of predicting the unknown, future target value, given input values.

When the response is numeric, the learning problem is called "regression." When the response takes on a discrete set of *k* non-orderable categorical values, the learning problem is called "classification." In predictive learning one uses data to build a good predictive model. A representative "training" data base with all response and predictor variables that have been jointly measured is assumed to exist. A "learning" procedure is applied to the training dataset in order to extract a good predicting function. There are many commonly used learning procedures including linear/logistic regression, neural networks, kernel methods, decision trees, etc. A technical overview of modern learning techniques is provided in [1].

Recently there has been a revolution in the field of statistical learning inspired by the introduction of three new approaches: the extension of kernel methods to support vector machines [4]; the development of reproducing kernel Hilbert space methods [5]; and the extensions of decision trees by application of boosting [6,7], bagging, and Random Forest (RF) techniques [2,3].

The complexity of the underlying data adds significant challenges when developing models for industrial applications. This includes mixed-type variables with blocks of non-randomly missing data, categorical predictors with a very large number of levels (hundreds or thousands). Very often datasets are extremely saturated: there are a small number of observations and a huge number of variables (tens of thousands). Both predictors and responses normally contain noise and mislabeled classes. Both regression and multi-level classification models are of interest. Thus, a universal, scalable and robust learner is needed.

Decision trees are one of the most popular universal methods in machine learning/data mining, and they are commonly used for data exploration and hypothesis generation. Classification and Regression Trees (CART), a commonly used decision-tree algorithm [8], use recursive partitioning to divide the domain of X variables into sets of rectangular regions. These regions are as homogeneous as possible with respect to the response variable and fit a simple model in each region, either majority vote for classification, or a constant value for regression [8]. The resulting model is a highly interpretable decision tree. Some of the principal limitations of CART are low accuracy, because piecewise, constant approximations are used, and high variance/instability.

One problem faced at Intel Corporation is multilevel classification problems (see the BinSplit case study below) with categorical predictors of high cardinality (*m* unordered values). Tree algorithms would need to evaluate $2^{(m-1)}$ possible partitions of the m values of the predictor into two groups. When *m* is large it becomes computationally intractable to evaluate all unique partitions. This problem is addressed in [9], and the algorithm is implemented in an internally developed set of statistical-learning algorithms. To reduce the high computational requirements, a hybrid clustering scheme (k-means and agglomerative) with a novel generalized distance metric was developed. This dynamic preprocessing method resulted in an efficient, computationally fast way to discover a small number of natural partitions of levels for such variables that have similar statistical properties in terms of categorical response.

Recent advances in tree-based methods such as Multiple Additive Regression Trees (MART) [7] and RF [13] have proven to be effective universal learning machines. Both are resistant to outliers in X-space, both have efficient mechanisms to handle missing data, both are competitive in accuracy with the best-known learning algorithms in regression and classification settings, and both handle mixed data types naturally. However, MART uses an exhaustive search on all input variables for every split and every tree in the ensemble, and it becomes extremely expensive computationally to handle very large numbers of predictors. At the same time, RF shows significant degradation in accuracy in the presence of many noise variables.

To address the computational limitations of MART and the susceptibility of RF to noise variables, a fast hybrid method using dynamic feature selection was proposed [10]. This method features a stage-wise stochastic boosting of shallow random trees built on a small intelligently sampled subset of variables. The method can be applied to noisy, massive regression and classification problems and has excellent predictive power. It is comparable to the best-known learning engines. Based on our experience, this combination of speed, accuracy, and applicability makes this hybrid method one of the best universal learners available. The hybrid method produced very competitive results at the 2003 Neural Information Processing Systems (NIPS) conference competition on feature selection in data sets with thousands of variables. There were over 1600 entries from some of the most prominent researchers in machine learning. Another method under development, based on an ensemble of kernel ridge classifiers [11], ranked as the second best entry in the NIPS 2003 competition, ahead of submissions from the world's best research universities.

Sometimes, it is necessary to identify patterns in a given set of predictors and/or reduce the dimensionality of the predictor variables. In this case, there are no response variables and the technique often referred to as unsupervised learning or clustering is used to find and group data into similar sets of observations. A hierarchical method is used to form and identify appropriate numbers of clusters based on training data. The nearest neighbor method is used (k-means) for assigning test data to clusters formed. Our second case study evaluated clustering techniques in addition to CART for TPT prediction analysis.

Many of the statistical algorithms discussed throughout this paper and used in the case studies have been integrated into an internally developed Windows* -based statistical-learning platform optimized for Intel Architecture, called Interactive Data Exploration and Learning (IDEAL). The application is updated as new algorithms are developed, and it is validated using a variety of internal test data sets.

PREDICTING THE SALES PRICE OF HOMES

Introduction

In this case study we use the buying and selling of a home to illustrate statistical-learning technologies. Some typical questions asked of a real-estate agent when buying or selling a home are as follows. What is my home currently worth? What is the fair price of that new home I am considering purchasing? Are the sellers asking too much or is it a good buy? Instinctively, we know many factors (variables) influence the value of a home: the size of the home and lot, its location, the home features and upgrades, its proximity to schools, the age of the home, landscaping, pool, number of rooms, garages, etc. Even factors such as the real estate agent of the buyer and seller likely influence the final selling price of a home.

Approach

In this case study we used a small set of both numeric and categorical variables to "learn" and then predict the future sales price of homes in Tempe, Arizona. This is an example of regression, where the target variable (sometimes called the response or Y variable) is numeric. The case study was used to generate interest in machine-learning technologies and is currently used in an internally developed statistical-learning class.

The first, and often the most difficult step in developing any statistical model is extracting and preparing the data. In this case, data on recent home sales in the area of interest were available from public sources such as the local newspaper and the county assessor's office. Data on approximately 1,300 homes were collected, and the data included such variables as the square footage of the home and lot, the size of the pool (set to zero if the house had no pool), the year the house was built, the sales date, and the subdivision name. A sample of the data is shown in Figure 2 along with a simple bivarate plot of one of the variables vs. the sales price (Figure 3).

Table 1: Home sales data

Subdivision Name	Land Size	Livable sq. ft.	Pool	Built	Sold Date	Sale Price ($)
TEMPE ROYAL PALM	6,042	1,391	0	1984	10/7/2003	$177,000
TEMPE ROYAL PALM	8,246	2,406	450	1988	5/28/2003	$259,900
COVENTRY TEMPE	10,812	3,456	600	1998	3/27/2003	$470,000
WARNER RANCH VIL	3,742	1,435	0	1985	10/1/2002	$147,500
ESTATE LA COLINA (	16,069	3,239	750	1980	8/1/2002	$334,500
TUSCANY	12,785	3,342	500	1996	5/1/2002	$487,000
PECAN GROVE ESTA	7,427	2,570	0	1990	1/1/2002	$238,000
RAINTREE UNIT 2 LO	16,575	2,638	400	1982	4/1/2001	$380,000
CIRCLE G RANCHES	36,647	4,057	500	1980	4/1/2001	$585,000
ALTA MIRA 1 LOT 1-	9,509	2,475	480	1983	1/1/2001	$204,000
ESTATE LA COLINA (	12,637	2,683	425	1981	11/1/2000	$279,000

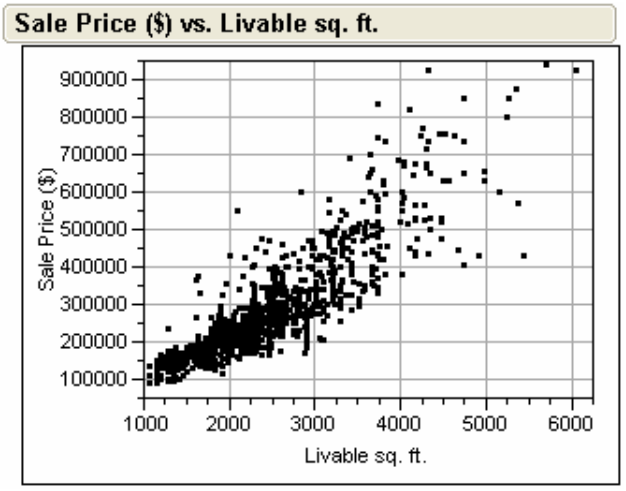

Figure 2: Sales price vs. livable square feet

To make the comparison to other statistical approaches fair and represent how the statistical model could be used in real life, the 1,300 home data set was manually divided into two parts based on the home sales date. The "training" data set included all but the 50 most recent home sales. The last 50 home sales were used as the test set to validate the model and quantify the error.

To compare with existing approaches, the training and test datasets (with the sales price withheld) were given to several statisticians within Intel Corporation to develop predictive models. Some statisticians used the data-mining

* Other brands and names are the property of their respective owners.

capabilities available in commercial software, while others applied traditional regression techniques. The traditional models took time to develop and required intermediate calculations such as average appreciation rates, assessment/screening of outliers, even driving through certain neighborhoods to view the homes in an attempt to build a more accurate model.

IDEAL was also used to create a gradient boosted tree (GBT) model to predict the sales price of homes in the validation data set. The mean error was compared between all modeling methods.

Results

A plot of the results achieved on the validation data set is shown in Figure 4. Using GBTs in IDEAL, the mean error was $14,473, while the error for other models ranged from $17,096 to $43,275. The time taken to model and produce the predictions using IDEAL was measured on the order of minutes, while most of the other statistical models took significant non-computational time to create and in some cases the statisticians had to manually intervene to identify and remove outlier observation(s) in the training data set.

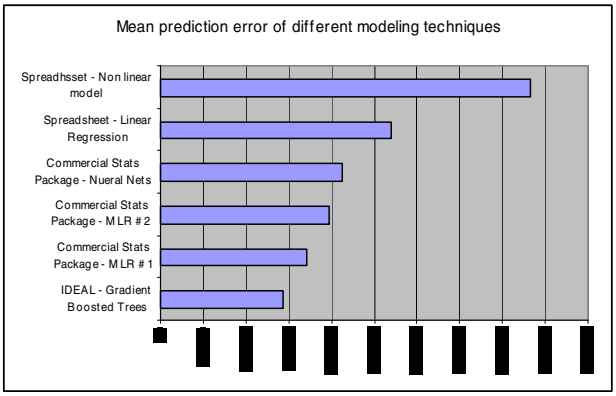

Figure 3: Prediction "contest" results

In addition to modeling speed and improved accuracy, IDEAL produced other outputs such as an explorable/drillable single decision tree (for single tree models), a variable importance pareto, and variable dependency plots. Examples of a variable importance pareto and dependency plot are shown in Figures 4 and 5. These tools aided in better understanding the relative variable importance and variable interactions/relationships present in the data.

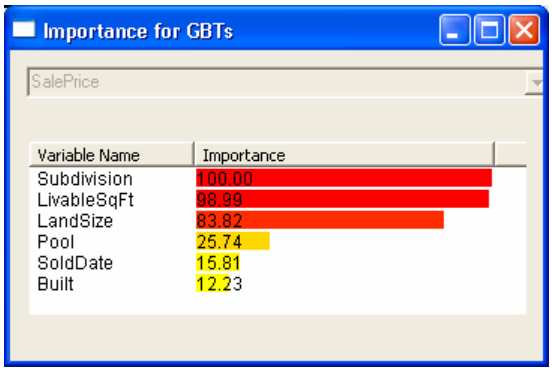

Figure 4: Variable Importance Pareto

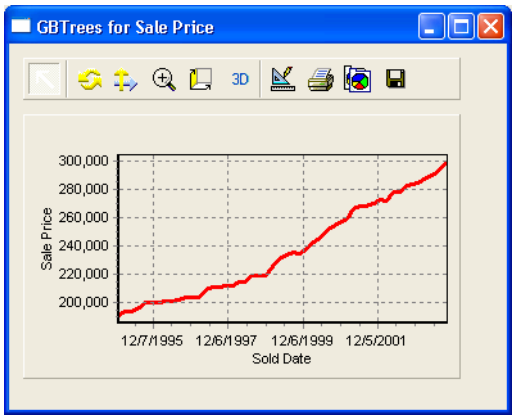

Figure 5: Dependency plot

Conclusion

Even in this simple example, where there are only six input variables, we can see the power of advanced machine-learning technologies. Because the data sets in semiconductor manufacturing are much more challenging, and the variable relationships often times non-linear, the use of statistical-learning techniques for accurate regression/classification and data exploration is a requirement.

THROUGHPUT TIME PREDICTION

Approach

A cross-functional team comprising groups from Intel and Arizona State University developed a proof of concept TPT data-mining technique using generic clustering and decision-tree techniques. Most of these techniques use historical data to provide prediction for current lots. This data-mining tool learns from past patterns in the factory and categorizes the current flow of products using its stochastic characteristics and existing data in order to predict lot TPT. To do this, the cumulative time of historical and existing lots by critical operation and route is used in addition to data on work in progress (WIP), and lot priority. The WIP values correspond to the number of

lots waiting to be processed at the critical operations at a given time. A typical data table in a manufacturing execution system database contains rows for each operation and each lot in production. Each row has information identifying the lot, the current operation, the production route, the time moved in to the current operation, the time moved out from the previous operation, and the time out from the current operation. There are also variables describing the type of production lot, the product type, and lot priority. There can be more than 400 rows for each lot with more than 15 columns of data, thus yielding more than 6000 variables to describe how an individual lot moves through the factory. A lot's cycle time for an operation can be described as queue time plus the tool processing time. Queue time is calculated as operation start minus previous operation out. Processing time is calculated as current operation out minus operation start. Queue time and process time were summed across all intermediate operations to get an aggregate cycle time between the critical operations.

$$Aggregate\ CT = \sum_{stepi}^{bottleneck} CT_i \qquad (1)$$

A lot velocity vector, which is a vector of aggregate cycle times for each critical step, was created, and different WIP measures were recalculated relative to move-in and move-out of the critical operations. Finally, the rows of data describing the individual steps for each lot were combined into a single row. Lots that did not have complete information were eliminated along with test or engineering lots. However, all filtered lots were included in the WIP counts since their presence affects the cycle time of production lots. Figure 6 pictorially represents the data-mining techniques and approaches adopted for lot TPT prediction.

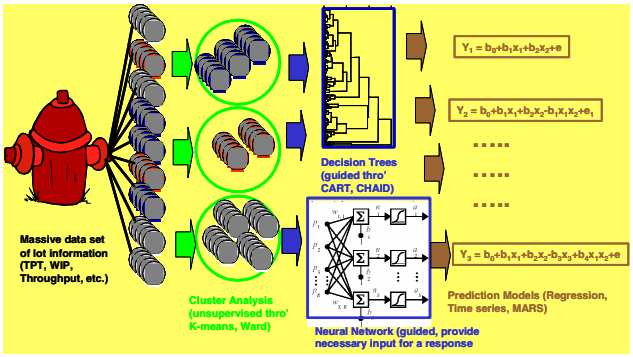

Figure 6: Data-mining techniques applied to lot TPT prediction

Little's Law states that on average CT = WIP/TP, where CT is the Cycle Time and TP is the throughput (processing rate) of an operation. Often the precision and reliability of the machines enable one to assume throughput to be nearly constant for a given machine. With this simplifying assumption, the cycle times of lots processed by the same tool are proportional to the ratio of the WIP as shown in equation 2.

$$CTb = (WIP_b / WIPa) * CTa \qquad (2)$$

In a linear production process with a first-in, first-out (FIFO) scheduling rule, WIP remains constant for each lot at each tool, and the cycle time of a lot for the series of tools is proportional to the ratio of WIP. The accuracy of this prediction depends on the variability in machine throughput. However, many production processes allow tools to be used in multiple operations, in which case WIP for a lot is not constant at every tool. In reentrant cases a lot may leave a tool at one step and return to the same tool several steps later. The WIP for the second pass through the tool includes new lots plus the lots making their first pass through the tool.

Although equation 2 cannot be used directly for cycle-time prediction, the measures of cycle time and WIP should be interchangeable as variables due to their proportional relationship. Also, lots that encounter similar values of WIP should have similar cycle times, and these can be predicted by comparing a new lot's cycle time at critical steps in the process to the cycle times of lots that have completed the process. The average cycle times for similar lots would predict the cycle time for the target lot. From the transactional data it is easier to derive an estimate of WIP state. To that end, averages of cycle times for adjacent steps, within a window of time, provide information about throughput and WIP. For the steps just ahead of the target lot, the average cycle time actually provides information about both the WIP ahead of the target lot and a measure of the throughput and WIP for prior lots. If the window is constructed to include lots that finish the current operation before the target lot and do not leave the operation before the target lot enters, then n counts all such lots, and n will be equal to the WIP for the target lot. Hence, we get the following equation:

$$Avg\ CT = \frac{\sum_{i=1}^{n} CT_i}{WIP} \qquad (3)$$

Given a target lot at step i, we can calculate the average cycle time at step i for neighboring lots. Each intermediate cycle time has an associated variance. The variance then for predicting the cycle time from the i^{th} tool to process completion is as follows:

$$Var(\sum_{i=j}^{m} CT_i) = \sum_{i=j}^{m} Var(CT_i) + \sum_{i=j}^{m}\sum_{h=j}^{m} Cov(CT_i, CT_h)\ i \neq h. \qquad (4)$$

In the data considered here, the covariance between intermediate cycle times was positive. Hence, the variance of the sum of intermediate cycle times increases as the number of intermediate steps increase. As expected, predictions of TPT made nearer to the end of the process have lower error rates since the amount of variability has decreased.

Results

In general, we found that data could be partitioned in a way that provides good predictions for future observations. However, the accuracy of the prediction varied from method to method. If cluster assignment is based on Euclidean distance from cluster center, prediction of multiple steps can be made from any current step using only one set of clusters. K-Nearest-Neighbors Prediction of multiple steps can be made from any current step by using same nearest neighbors. Regression trees perform automated handling of optimal tree size, and the tree must be constructed for each step to be predicted. These data-mining techniques were compared based on the mean absolute error and median absolute error (median is robust to outliers). CART and CART in Cluster performed better and provided prediction variability within two days as shown in Table 2.

Table 2: Comparison of model results

Method	Data Split	Median Absolute Error	Mean Absolute Error
KNN (K=5)	Random	2.10 days	4.88 days
KNN (K=10)	Random	2.35 days	4.91 days
CART	**CV**	**1.25 days**	**2.66 days**
Cluster (N=5)	Random	2.93 days	3.66 days
Cluster (N=10)	Random	2.60 days	3.38 days
CART in Cluster	**CV**	**1.09 days**	**1.65 days**
Neural Network	Random	2.55 days	3.61 days

The model validation also showed favorable results as shown in Figure 7 in which actual TPT is compared to predicted TPT. A demo product has been developed that combines these steps into a seamless process using a simple user interface.

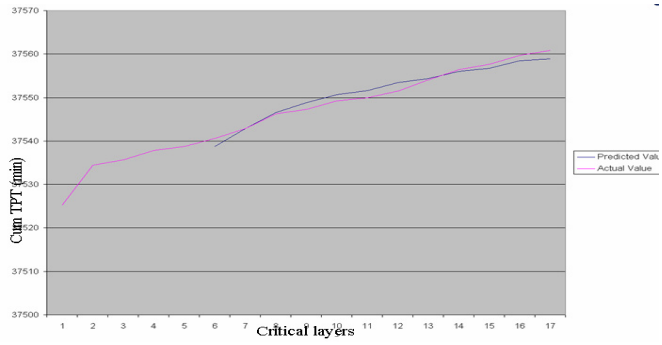

Figure 7: Comparison of actual TPT with predicted TPT

Conclusions

Analysis of data mining for lot TPT prediction on Fab data showed lot TPT prediction within two days and compared favorably with other static empirical models. Good predictions for lots currently in production can be obtained from similar lots that have already completed production. The team is currently evaluating a detailed pilot at one Fab. The powerful capability of this data-mining technique (Cluster and CART) is attracting many other internal customers. Some possible future projects include product health indicator, analysis/forecasting, demand planning, and process control.

SIGNAL IDENTIFICATION/SEPARATION

Introduction

Semiconductor fabrication is becoming increasingly complex, with routes stretching to several hundred process steps. Even with highly advanced process control systems in place, there is inevitable variation in yield and performance between and within manufacturing lots. A common practice is to associate this variation with equipment differences by performing analysis of variance at every process step where multiple pieces of equipment are used. Looking for operations where there are significant differences between process tools can reveal the sources of process variation.

Challenges

The one-at-a-time approach to equipment commonality studies has many shortcomings:

1. Most target variables of interest are affected by multiple process steps. For example, yield can be reduced by high particle counts at nearly any manufacturing step. The maximum frequency (Fmax) at which a part can run can be affected by a variety of lithography, etch, implant, and diffusion operations.
2. Lots are not distributed randomly across process tools. Often, material from one tool at one operation is run preferentially (or even exclusively) on a particular tool at another operation. In addition, some operations (notably lithography) are prone to running large blocks of the same product in a short time period, then not running any more for several weeks. Lots may also cycle through the same tool several times at different process operations.
3. The difference between tools can rarely be described by a constant offset. For example, lots from one tool are rarely a consistent 100 MHz faster than those from another tool over a significant period of time. Usually, the difference between process tools varies over time, sometimes dramatically. There is usually a

mixture of short-term and long-term trends in the data, often associated with the cycles for preventative maintenance.

Results

By using machine-learning techniques, we have overcome the above challenges, and can now look at all processing operations simultaneously, while also accounting for temporal trends in the data. The methodology we used is simple. At each process operation, we create two columns of variables: a categorical column indicating which process tool the lot was processed on, and a numeric column giving the time and date at which the lot was processed. These columns, which can number in the thousands, are fed into the learning engine along with the target variables of interest. We have found that stochastic, GBT models can give astonishingly accurate reconstructions of the time trends, even when there are significant differences at multiple operations.

These techniques are best demonstrated using simulated data, since the underlying trends in actual process data are unknowable. In this example, we simulated a hundred operation processes where five operations had significant differences between tools for Fmax. The programmed differences included steady-state offsets, saw-tooth patterns of short duration, step functions, and gradual drifts. Each lot had the same baseline frequency, to which the offsets from all of the affected tools were added. Additional random noise was also added.

One of the operations with an implanted signal is shown in the graph in Figure 8. Two of the five tools (entities) have square wave shape signals, while the remaining three tools have a consistent, zero offset. Other operations had different shapes of embedded signals.

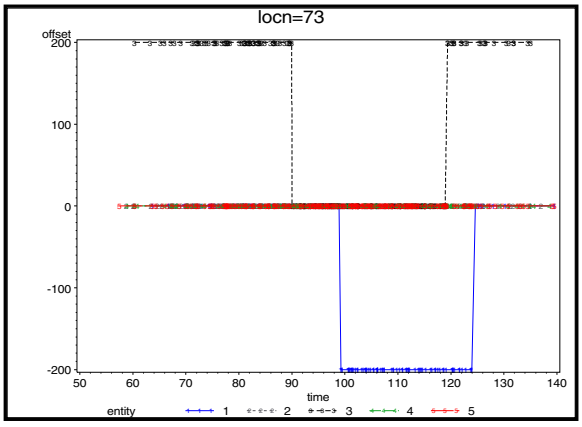

Figure 8: Implanted signals

Detecting which tools had non-random effects was the first challenge. Using stochastic, GBT models in IDEAL we were able to both identify which operations had embedded signals and accurately recover the patterns. To identify the tools and times with non-random patterns, we used the Variable Importance Pareto, which showed the relative variance reduction for all variables in the model. Variables with high variance reduction appeared highest on the pareto as in Figure 9.

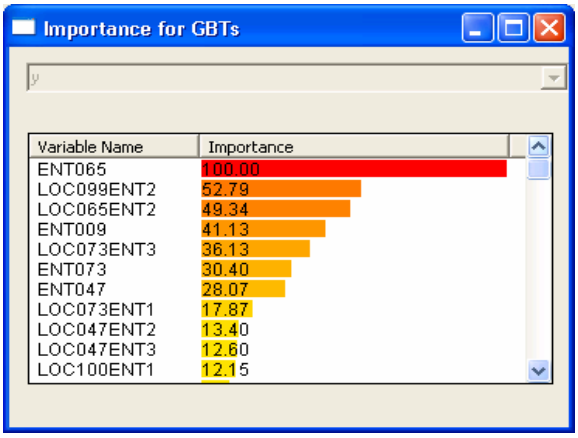

Figure 9: Variable importance

Signal separation is the next step. Figure 10 shows how the signal programmed in the data for one operation was lost in the added noise and the other valid embedded signals at other operations. Plotting just the response variable as a function of the time through the operation of interest did not show the important shift. Even using traditional approaches to average out the noise yielded only marginally better results–and required the upfront knowledge of which operations had signals.

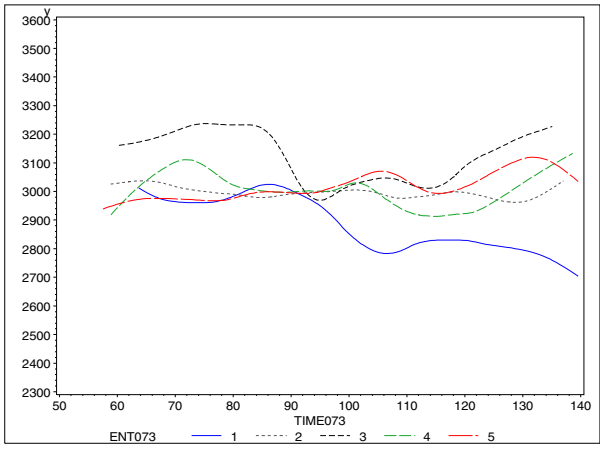

Figure 10: Hidden signal

Figure 11 depicts the recovered signal using GBT models in IDEAL. The recovery is not perfect due to the injection of random noise and the presence of valid signals from other operations. The recovery was far superior, however, to that given by a wide variety of classical statistical techniques, including multiple spline-based models.

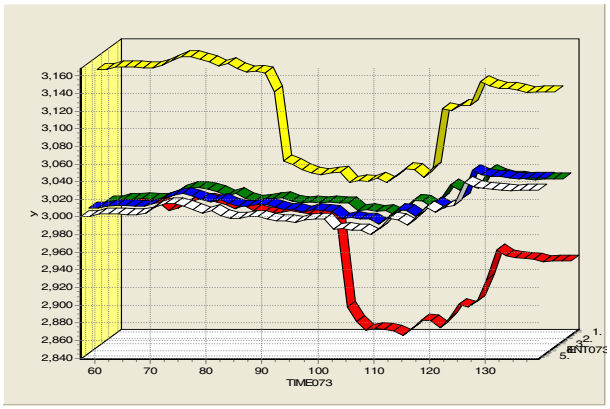

Figure 11: Recovered signals

The GBT-based machine-learning techniques are particularly attractive since they allow for both continuous and categorical target variables, and they allow for easy introduction of other predictor variables of mixed data types such as chemical vendors, queue times, or the results of inline metrology measurements.

UNIT-LEVEL BIN SPEED PREDICTION

Introduction

In the last case study we tackled a challenging application of statistical learning: accurately predicting the final test outcome (bin speed and yield) of individual microprocessors using upstream data from Fab and Sort. The results presented in this paper are for a stepping of the Intel® Pentium® 4 desktop processor on 130nm technology that is no longer in high-volume production. Analysis using the same techniques, with many additional variables and observations on current microprocessor generations has yielded similar, and in some cases better, results.

In this case, challenges emerge from a variety of factors such as the large number of observations and variables, non-randomly missing data (i.e., sampled), both categorical and numeric variables, presence of outliers, dynamic variable names, frequent process changes and improvements, non-linear variable relationships, etc. Even extracting the variables from databases and properly associating them with the appropriate unit level presented many challenges.

One traditional approach to classify the final speed of a CPU is to fit multiple linear logistic regressions by using a limited quantity of upstream numeric predictor variables such as Fmax and leakage current, measured at the wafer sort operation. However, these models assume prior knowledge of the underlying distributions, do not effectively address categorical variables, and require periodic, manual recalculations. Recent voltage and power optimization test strategies also invalidate many distribution assumptions.

Approach

To predict the final unit-level outcome, we first needed to extract and prepare the unit-level data set that contained both the predictors and responses (X's and Y's) for training the models. Several internally developed tools exist for extracting and joining unit-level data from databases across multiple factory data sources. These tools have improved in speed and capability as unit-level traceability has become the norm for semiconductor manufacturing. Even more specialized data marts have been created to lower the accessibility hurdles and improve speed and scalability.

We defined our response variable as "final speed if pass or fail" which resulted in six unique classes of units. The distribution of the response variable in the training dataset is shown in Figure 12.

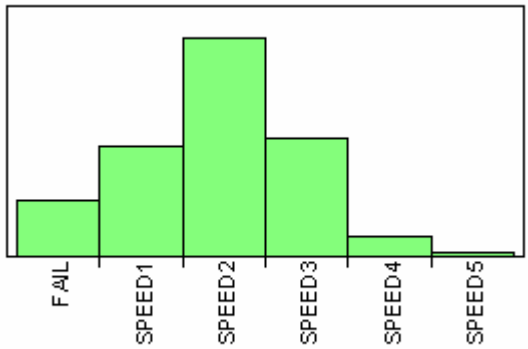

Figure 12: Distribution of response variable

For this multilevel classification problem we used GBT models in IDEAL to train the model. The model error was assessed using cross validation, i.e., using the model to classify observations that were withheld from the data set during the training process. Cross validation is a key component of IDEAL.

A bivariate plot of two of the continuous X variables (in arbitrary units) vs. the response variable (color/symbol of the points) is depicted in Figure 13. It is meant as a visual aid. If we had been using only two numeric variables to classify the units into the six possible outcomes, we would essentially be determining the set of two-dimensional

® Intel Pentium 4 is a registered trademark of Intel Corporation or its subsidiaries in the United States and other countries.

classification boundaries that minimize error on the test data set.

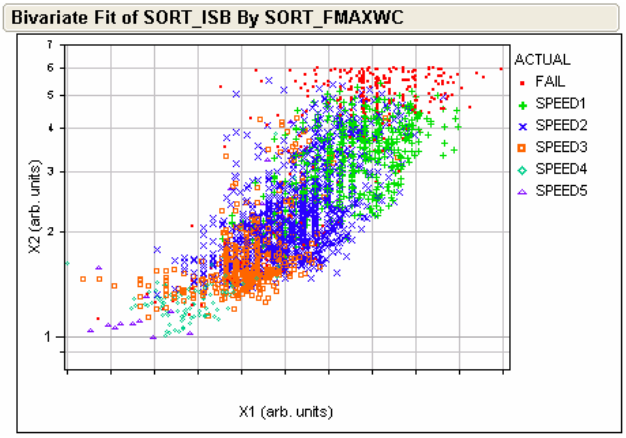

Figure 13: X2 vs. X1 (arb. units)

However, we are using many continuous variables as well as categorical variables as predictors, so the two dimensional classification boundaries now become an *n* dimensional classification hyper surface, where *n* is the number of variables in the model.

The overall error for a multilevel classification is determined by the misclassification percent across all classes of the target variable. In other words, it is the percent of total observations in the validation data set that were misclassified by the model. The goal of any classifier algorithm is to minimize misclassification. For multilevel and unbalanced classification problems, the rare classes may be "sacrificed" to minimize overall misclassification. IDEAL provides options in GBT models to change the relative weight, sometimes called misclassification penalty, of rare classes as the model is built. Overall misclassification may increase slightly; however, misclassification on rare classes is reduced. This is often the case when we are modeling rare classes such as failing units or a microprocessor speed bin that is faster than the one currently on the market. Other GBT options are available in IDEAL and were used for fine tuning the classification model. These included the learning rate, tree depth, number of iterations, and dynamic feature selection [10].

Results

The results from the case study presented in this paper show that low unit-level misclassification rates are achievable using numeric and categorical predictors from upstream (Fab and Sort). The overall cross-validated misclassification rate was 24% for the test data set. The misclassification matrix is shown in Table 3. The value in each cell is the percent of total observations in a given class that were predicted for each of the possible classes.

Table 3: Misclassification matrix

		PREDICT					
	Row %	FAIL	SPEED1	SPEED2	SPEED3	SPEED4	SPEED5
ACTUAL	FAIL	68.13	10.01	14.21	6.20	1.20	0.25
	SPEED1	2.26	82.44	15.30	0.01	0.00	0.00
	SPEED2	0.45	8.44	84.75	6.34	0.01	0.01
	SPEED3	0.09	0.73	20.07	77.46	1.62	0.03
	SPEED4	0.25	0.76	5.89	22.40	69.44	1.26
	SPEED5	0.88	7.51	27.69	11.05	13.25	39.62

For this data set, even a significant portion of true fails was predictable based on upstream Fab and Sort data. The decision to build die with higher probability of fail into units was taken because of an analysis of the relative die and final unit costs and other business considerations. Optimization of the decision process using financial variables and other business rules/constraints is currently an active area of research and development.

Although the primary goal of this case study was to determine how accurate a classification model can be created using statistical-learning techniques, several other benefits were also realized during the machine-learning process. A single tree model, although much less accurate than GBT models, was created in seconds using IDEAL. The single tree enabled visualization of the model and offered insights into non-obvious variable relationships. Also, IDEAL calculated normalized variance reduction, so we are able to see which variables are the most important in predicting the final speed. Lastly, the dependency plots are able to show the effect of an individual variable after averaging out the effects of all the other variables. Although the results are not presented here, these outputs from IDEAL were useful in identifying sources of equipment variation that impacted the final classification result–very similar to the previous case study on signal identification/separation.

Using more recent data on different microprocessor lines misclassification rates have been reduced from 20% down to 10%.

DISCUSSION

The results of the case studies demonstrate how data-mining/statistical-learning methodologies are being used at Intel Corporation to convert large amounts of semiconductor manufacturing data into real, actionable knowledge. These case studies only scratch the surface of the possible applications of these methodologies, many of which are active areas of research and development. Examples include simultaneous prediction of remaining cycle time and final test results, factory WIP prioritization and optimization, reformulation of yield and speed prediction models into unit-level predicted profit, autonomous learning prediction and optimization, and advanced multivariate process control systems.

All the applications mentioned above require robust and efficient statistical-learning technologies that can address the challenges of semiconductor data. IDEAL, developed internally with these requirements in mind, has demonstrated leading-edge capabilities in prediction accuracy, modeling speed, and model interpretability. Indeed, internal benchmarks have shown IDEAL to be faster and more capable when compared to commercial statistical, data-mining packages.

CONCLUSION

Utilization of data-mining and advanced statistical-learning methods in the semiconductor industry will continue to grow and will play an important role in maintaining Moore's Law. With each successive process generation, semiconductor manufacturing becomes more technologically complex. Concurrently, the quantity, complexity, and availability of data also march forward. These advanced methods provide us the means to understand and extract key lessons from our complex data.

ACKNOWLEDGMENTS

The authors gratefully acknowledge IT, R&D, and analysis help from Arizona State University, particularly Dr. George Runger and his students.

REFERENCES

[1] Hastie T., Tibshrirani R., and Friedman J., *The Elements of Statistical Learning*, Springer, New York, 2001.

[2] Breiman, L., "Bagging predictors," *Machine Learning 26*, 123-140, 1996.

[3] Breiman, L., "Random forests, random features," *Technical Report*, University of California, Berkeley, 2001.

[4] V. Vapnik, *The Nature of Statistical Learning Theory*, Springer, New York, 1995.

[5] Wahba, G., "Spline Models for Observational Data," *SIAM*, Philadelphia, 1990.

[6] Y. Freund and R. E. Schapire, "*Experiments with a New Boosting Algorithm*," in *Proceedings of International Conference on Machine Learning*, pp. 148-156, 1996.

[7] Friedman, J. H., 2001 "Greedy function approximation: a gradient boosting machine," *Annals of Statistics 29*, pp. 1189-1232, 1996.

[8] L. Breiman, J.H. Friedman, R. A. Olshen, and C.J. Stone, *Classification and Regression Trees*, Wadsworth Inc., Belmont, California, 1984.

[9] Tuv, E., Runger, G., "Pre-Processing of High-Dimensional Categorical Predictors in Classification Settings," *Applied Artificial Intelligence 17(5-6)*: 419-429, 2003.

[10] Borisov, A., Eruhimov, V., and Tuv, E., "Flexible Ensemble Learning with Dynamic Soft Feature Selection," forthcoming chapter in *Feature Extraction, Foundations and Applications*, editors: I. Guyon, S. Gunn, M. Nikravesh, and L. Zadeh, Springer, New York, 2004.

[11] Torkkola, K. and Tuv, E., 2004 "Ensembles of Regularized Least Squares Classifiers for High Dimensional Problems," forthcoming chapter in *Feature Extraction, Foundations and Applications*, Editors: I. Guyon, S. Gunn, M. Nikravesh, and L. Zadeh, Springer, New York, 2004.

[12] Fayyad, U., Piatetsky-Shapiro, G. and Padhraic Smyth, "From Data Mining to Knowledge Discovery: An Overview," Chapter 1 in *Advances in Knowledge Discovery and Data Mining*, pages 1–34, AAAI Press, 1996.

[13] Hopp, W. J and Spearman, M.L., "Factory Physics–Foundations of Manufacturing Management," The McGraw-Hill Companies, Inc., 1996.

AUTHORS' BIOGRAPHIES

Randall Goodwin graduated from Cornell University in 1992 with a B.S. degree in Applied and Engineering Physics. He joined Intel in 1992 and currently works in product and test technology development. One of his many interests is the application and optimal utilization of machine-learning technologies in semiconductor manufacturing. His e-mail is randall.s.goodwin at intel.com.

Russ Miller holds a B.A. degree in Physics from the University of Chicago, an M.S. degree in Statistics from Texas A&M, and an M.S.E.E. degree from Columbia University. He joined Intel in 1992 and has held a variety of positions in statistics, yield engineering, and strategic forecasting. His primary interest is improving the yield, reliability, and performance of high-volume microprocessors through the analysis of very large data sets. His e-mail is russell.miller at intel.com.

Eugene Tuv is a staff research scientist in the Enabling Technologies and Solutions Department at Intel. His research interests include supervised and unsupervised non-parametric learning with massive heterogeneous data. He holds postgraduate degrees in Mathematics and Applied Statistics. His e-mail is eugene.tuv at intel.com.

Mani Janakiram is a manager in the Enabling Technologies and Solutions Department at Intel. He has

18+ years of experience and has published 20+ papers in the area of statistical modeling, capacity modeling, data mining, factory operations research, and process control. He has a Ph.D. degree in Industrial Engineering from Arizona State University. His e-mail is mani.janakiram at intel.com.

Sigal Louchheim leads the Data Fusion research focus area in the Information Services and Technology Group. Her main research interests are Interestingness (what is interesting) in Knowledge Discovery and Data Mining. Sigal received her Ph.D. degree in Computer Science in 2003, her M.Sc. degree in Computer Science in 1996, and her B.Sc. degree in Mathematics and Computer Science in 1990. Her e-mail is sigal.louchheim at intel.com.

Alexander Evgenyevich Borisov was born in Nizhny Novgorod, Russia and received his Master's degree in mathematics (Lie algebras) at Lobachevsky Nizhny Novgorod State University where he is currently working on a Ph.D. degree in the area of context-free grammars. He currently works at Intel (Nizhny Novgorod) as a software engineer and researcher. His technical interests include artificial intelligence and data mining, especially tree-based classifiers. His e-mail is alexander.borisov at intel.com.

Metadata Management: the Foundation for Enterprise Information Integration

Thiru Thangarathinam, Information Services and Technology Group, Intel Corporation
Gregg Wyant, Information Services and Technology Group, Intel Corporation
Jacque Gibson, Information Services and Technology Group, Intel Corporation
John Simpson, Information Services and Technology Group, Intel Corporation

Index words: Metadata, Repository, XML, XMI, MOF, UML, CWM, Impact Analysis, Reuse, Data Quality, Interchange, Architecture

ABSTRACT

Metadata Management helps you understand what information you have, where it is located, and what value it provides to users. Users can view information in a context they understand, providing a more efficient and intuitive way to communicate. To achieve this kind of enterprise-wide information integration, companies need to describe and share, in a common way, the data in their disparate data sources. This should include the business description associated with the information asset, as well as location, connection details, data type details, and the information's relationship with other resources. Sharing this information leads to an increased visibility of information across an enterprise, shorter development times, and reduced operational costs as the organization can discover and eliminate redundant information sources.

In this paper, we explore how Metadata Management streamlines the application development process by reducing the development, deployment, and maintenance costs. This is made possible by the use of a single source of Metadata for logical, physical, and process aspects of the application environment, when tracking versions of the code and documenting all aspects of the application development life cycle. By providing a complete, integrated view of the development environment, Metadata helps identify redundant processes and applications, thereby reducing duplicated efforts. (For the purposes of this discussion, we are treating "Metadata" as a singular noun.) Developers can share and reuse existing objects such as data structures, programs, model definitions, and more. In addition, enterprise impact analysis greatly reduces the analysis and maintenance phase of the development life cycle.

INTRODUCTION

Have you ever wondered about the definition of a piece of data? Has something stopped working and you don't know why? Do you need to make a change in the environment and want to know ahead of time what will be impacted by your change? The answers to all of these questions can be found in Metadata. Metadata is information about data. It answers the who, what, when, where, why, and how of every piece of data being documented throughout the enterprise and is the enabler for reuse. Historically, Metadata has been defined as "data about data." Metadata can also be thought of as the "DNA" of your corporation. Through its systematic capture during the creation of the assets in your enterprise architecture, you can search, catalog, and ultimately reuse corporate assets, achieving enterprise-wide information integration.

To achieve this enterprise-wide information integration, companies need to describe and share, in a common way, the data in their different data sources. This should include the business description associated with the information asset, as well as its location, connection details, data type details, and the information's relationship with other resources. Sharing this information leads to an increased visibility across an enterprise, shorter development times, reduced operational costs as redundant information sources are identified and eliminated, and improved data quality as organizations begin to reuse approved information. The best way to manage and share this information is through a centralized Enterprise Repository that drives the connections between data, process, and applications, enforces standards, and is available to all employees.

As the number of applications in an organization increases and the time to design these applications decreases,

companies begin to recognize the need for a Metadata Repository. However, all too often they focus primarily on implementing a tool and neglect the Metadata Management aspect. This results in a Repository that is easily accessible, but not often used because the information lacks credibility. To be of value to the enterprise, Metadata must be managed using repeatable processes to identify and define the Metadata, standard deliverables to capture the Metadata, and approved governance processes to ensure ongoing Metadata credibility. This management allows users to understand what information they have, where it is located, and what value it provides. Plus, users can view information in a context they understand.

METADATA REPOSITORY ARCHITECTURE

As an enterprise infrastructure component, a Metadata Repository should provide a single, secure, and standard method for providing Metadata to end users. One of the key components of a Metadata Repository is its data collection architecture. There are three types of data collection architectures that can be employed. They are as follows:

- *Distributed Data Collection Architecture*. In this architecture, also known as the Active Repository, the Metadata Repository maintains pointers to external Metadata sources rather than creating and maintaining duplicate copies, thus eliminating all synchronization issues.
- *Centralized Data Collection Architecture*. In the centralized data collection architecture (known as the Static Repository), the Metadata is copied from various sources to the centralized repository. In a centralized Repository, accessing of the Metadata is independent of access to the original system, as the duplication of the data in the repository frees the system from any required access to the original Metadata.
- *Hybrid Data Collection Architecture*. This method utilizes both distributed and centralized data collection methods in a single repository. A hybrid approach consisting of both distributed and centralized approaches leverages the strengths and mitigates the weaknesses of both distributed and centralized architectures. The hybrid approach facilitates Metadata access in real time from third-party sources and provides the ability to capture additional Metadata attributes not existing in the source repositories. It also lets users create original data within the repository.

The Metadata Repository should also be designed in such a way that it can be fully scaled and customized, thereby enabling it to store a limitless amount of information for any particular object. This flexibility provides the Metadata Repository with the ability to conform to current and future standard initiatives, such as the Common Warehouse Metamodel (CWM) guaranteeing that users of the Metadata Repository will be able to conform to such initiatives as they receive wider adoption. To help enterprises increase information visibility and sharing, a Metadata Management solution should at least contain the following:

- A secure Web-based portal solution to search and analyze Metadata.
- A Metadata Management Engine that provides the standard processes for ensuring the credibility of the Metadata.
- Standards-based Metadata Repository and Metadata Representation. The common Metadata standards it should support are as follows:
  - Meta-Object Facility (MOF)
  - Common Warehouse Metamodel (CWM)
  - Unified Modeling Language (UML)
  - XML Metadata Interchange (XMI)

The following sections discuss each of these components in detail.

Secure Web Portal

The secure Web portal provides a graphically driven, browser-based interface for displaying the information stored in the Metadata Repository. Any authorized user with a secure Internet connection and a browser-enabled computer can have access to the Web portal. The Metadata Repository should also utilize a variety of security measures ensuring that databases and repositories remain secure while at the same time restricting access and usage privileges to only those users authorized by the Metadata Repository system administrator. By providing inherent security features, only those authorized to view appropriate information can view it. This enables business users to view information about the reports they run without confusing them with information about their other processes.

Metadata Management Engine

The Metadata Management Engine provides the core foundation for ensuring the credibility of the Metadata and it consists of the following components:

- *Archiving/Versioning Component*. This component tracks changes to the Metadata for all objects within the Metadata Repository. Through this function, users can monitor specific changes, restore Metadata records for specific objects to previous instances, purge changes from Metadata records, or create reports detailing the audit trail of all changes made for a Metadata record (such as date, time, user, and the nature of a change).
- *Repository Component*. This component is the database repository for storing all the Metadata objects, all of their relationships, and the properties associated with them. For the repository to be extensible and flexible, the Repository Component needs to be built on top of the Enterprise Metadata Model. The Enterprise Metadata Model not only defines all the data structures existing in the enterprise, but also how those data structures will interact within the application integration solution domain. The Repository Component also contains various internal functional subcomponents for governance, development lifecycle integration, configuration, security, and metrics.
- *Scanner Component*. This component is responsible for capturing Metadata from disparate Metadata sources and loading them into the Metadata Repository. Once the Metadata is loaded, it can then be accessed through the secure Web portal. By leveraging the industry-standard XMI specification for capturing Metadata, the Metadata Repository can exchange Metadata models with common modeling tools as well as import Metadata from databases into the Metadata Repository.
- *Administration Component*. This component contains the processes that perform operations such as monitoring quality, change management, classification, configuration, and access management. It also contains Metadata Management processes that are responsible for handling check-in, check-out, and for auditing scanned input to control changes and so on.

Support for Industry Standards

In this next section, we provide an overview of each of the standards and the support a Metadata Repository implementation should provide for each of these standards.

- *Meta-Object Facility (MOF).* The MOF standard defines an extensible framework for defining models for Metadata, and for providing tools with programmatic interfaces to store and access Metadata in a repository.
- *Common Warehouse Metamodel (CWM).* The CWM standard specifies interfaces that can be used to enable easy interchange of warehouse and business intelligence Metadata between warehouse tools, warehouse platforms, and warehouse Metadata Repositories in distributed heterogeneous environments.
- *Unified Modeling Language (UML).* The UML standard provides a graphical language for visualizing, specifying, constructing, and documenting the artifacts of distributed object systems. It also incorporates Action Semantics, which adds to UML the syntax and semantics of executable actions and procedures, including their run-time semantics.
- *XML Metadata Interchange (XMI).* The XMI standard specifies an open-interchange model intended to give developers working with object and distributed technology the ability to exchange data between tools, applications, repositories, business objects, and programs. This is a Stream-based Model Interchange Format (SMIF) that enables the exchange of modeling, repository, and programming data over the network and Internet in a standardized way. XMI is a very important specification that brings consistency and compatibility to applications created in collaborative environments.

Metadata Repository Usage Scenarios

In this section, we demonstrate the different usage scenarios of the Metadata Repository in terms of the roles of the users.

Types of Usage

Impact Analysis

Impact Analysis is a complex discovery of data dependencies in order to realize relationships and uncover exactly how changes to any data elements can cause a cascading effect throughout the relationships. Having a detailed understanding of data as they cut across many functional areas (finance, manufacturing, sales, and others) as well as usage methods (development, process, and workflow) has created new opportunities for leveraging the information we already have to allow for more detailed data-supported decision making.

Metadata Reuse

Metadata reuse is the ability to reuse objects (artifacts) with little or no interruption to the business process. This can be referencing an existing object and creating a relationship to it or using an object as a base to create a new (similar) version. Reuse is where we gain extreme savings in time and resources (people, systems, and data).

Reporting
Reporting is simply data in a summary or detailed view. Examples include understanding all the data elements within a specific table, or knowing how many databases on the physical layer map to a specific contextual entity.

Usage Scenarios

Since customers often approach the repository with problems needing solutions, we present usage scenarios of problems followed by solutions that can be found within the repository.

Data Analyst/Modeler
Impact Analysis. Customer identification numbers need to become longer. Since your data are pushed downstream into the data warehouse, you are unsure of all the systems that could be impacted by this change. These relationships would have been previously discovered in the Metadata Repository, which would help identify your data reuse and the impact of any changes.

Metadata Reuse. You have been given the task of creating the logical model for an upcoming business solution. Through discovery (search and browse in the Metadata Repository), you have found out that there are already the required entities in place to describe the data required. Utilizing the built-in check-out and reservation methods in the Metadata Repository, you can capture reuse associated with the existing entities. During subsequent modeling, you can also map the relationship between these entities and your logical model, thus increasing the usefulness of future data to other users.

Reporting. Data quality efforts are on the rise in your organization, and you are curious whether or not all of your data tables have the column descriptors per the new standard. Through reporting you can identify descriptions that are missing, items that are described incorrectly, as well as potential other problems you should address in the event of more strict requirements around quality. Summary analysis (on screen or in an extract) aid in speedy comparison and identification for future repair.

Applications Developer/Programmer
Impact Analysis. The current project requirement requires the removal of local data stores of user information and reuse of data stored within the Record of Origin (ROO) for customer first and last name. Quick analysis of your operational and physical data implementations show that your data do not leave your system. Without the tools at your disposal to uncover this basic truth, a lengthy discovery process must occur to ensure that all interfaced systems as well as downstream consumers for reporting are not impacted by your changes.

Metadata Reuse. In the example above, you have been given the task of removing a local data store of customer first and last name. In place of the local store, you are going to utilize the consumption of a Web service related to customer information, real-time, and display the subsequent results on screen. Through a search discovery process, you identify the associated Web service and methods necessary to obtain this source of data. Integration into your application is simplified through this process, and an increased level of reuse is realized in the enterprise. Additionally, through a registration process, you are now associated with this external system. This minimizes any future changes to the external system's data structures from adversely impacting your system performance and data quality.

Reporting. The only constant in business is change. This change can be with employees, systems, and data. After a recent organizational change, you find yourself the sole developer supporting a system that you did not write. Although the documentation was kept up to date regarding code and migration, the data interconnects on the enterprise were never developed. To ease you into your new job, you can run a report to identify all interrelated data connections as well as mappings within other systems (transformation and reuse). This helps you to further develop technical and business contacts and schedule your next release in a timely and effective manner. This same technique can be utilized when transferring data for outsourcing and/or offshore support.

Keep The Business Running (KTBR) Staff
Impact Analysis. As part of a new security measure, you are given the task of ensuring that all the Database Management Systems (DBMS) for systems you manage and interface with, have been updated to the newest versions of DBMS. Short of logging onto each system, there is no other way to perform this. However, with real-time components of a Metadata Repository, cataloging of application systems and environments are enabled and maintained accurate.

Reporting. You have discovered that a server has a faulty power supply causing the whole server to burn up during the night and go offline. The damage was so extensive that even the parallel storage system was destroyed beyond recovery. Adding to this problem, your cataloging system, which contains server versus application and data, was hosted on that server. Not knowing what customer contacts to utilize regarding the outage, recovery frustrations rise as application users start to flood the help desk and bog down engineers attempting to locate backups in hopes of restoring the systems. A simple roll-up of server-based data stores, application systems, and related transformation objects can be done per server. Through related Metadata objects, a simple discovery of

application owner and linked data stores (upstream and downstream) should help in notification and contingency planning for the recovery.

Security/Auditing Staff

Impact Analysis. Recent changes to governing policies have made the task of knowing the type of data you have and where they are stored a very high priority. Regulations now require you to secure employee phone numbers from all systems other than HR source systems. Historically, auditing the locations of this information requires manual interviews and data discovery processes. The only piece of information you have is the HR source system. Through the use of a Metadata Repository and intelligent impact analysis, you would be able to crawl the relationships and discover all related systems that this information was fed into and delivered from as well as ROO contact information in order to obtain data not readily available.

Reporting. A new federal policy has been passed by Congress requiring that people with access to personnel addresses should not be allowed access to sales customer databases. It has been shown that cross-referencing this can help to identify people who work for a company and also consume their products, which has led to actual acts of prejudice against those that do not consume their own company's products. In the past, a simple analysis like this would have taken weeks to perform. All interrelated systems had to be crawled, analyzed, documented, and validated. The Metadata Repository by itself does not provide this functionality. However, when the targeted data stores are identified, and coupled with a Role-Based Access Control (RBAC) system, you can easily flag conflicts and put in place rules to prevent this from happening in the future.

Acquisition Coordinator/Manager

Metadata Reuse. Because you have been acquired by a new company, several disparate applications have been marked for integration into the enterprise in order to more easily integrate new employees and their associated processes. Simple analysis discovers similar systems already inside the enterprise, and based upon data structures in both systems, you can easily engage with the source system to coordinate a data migration and integration effort.

Program/Project Manager

Impact Analysis. Prior to last year, you had been using a legacy application to track shipping information related to hazardous chemicals coming into and leaving the property. With the integration of this information into our enterprise shipping applications, it is now necessary to end the legacy system (and data) and concentrate your efforts on your new system. Your legacy system had been integrated into dozens of applications and reporting structures over the last seven years, and the analysts who helped perform these tasks have moved on with little or no documentation. A primary tenet of an Enterprise Metadata Repository is the integration of ROO data related to applications. This information is the first step towards accurate mappings-related data consumed by that system as well as mappings between data sources and reused objects.

Reporting. Executive staff have communicated the need for increased reuse with a subsequent decrease in single-use, non-reusable systems and environments. Your task is to identify systems that exist in a silo. Applications or data stores that have no data input and no data output are of zero reuse benefit to the company. After identification of these systems, you are given the task of identifying additional redundant systems that perform the same function. Your end state is to consolidate these systems in the hopes of increasing reuse and decreasing isolated data stores. The reporting aspect of the Metadata Repository can help to identify systems with no interconnects to data stores or applications. Categorization of these items can help to identify trending possibilities and through some subsequent analysis, overlaps in functionality will bubble up. This analysis which used to take months and was limited on penetration, can be done in minutes and can scope the entire enterprise. Isolation is no longer required when reuse is allowed.

BENEFITS OF METADATA MANAGEMENT AND METADATA REPOSITORY IMPLEMENTATION

There are several benefits associated with investment in a robust Metadata Management program and associated Metadata Repository. These include increased reuse of your corporate assets, improved impact analysis associated with these assets, increased quality for decision-making, reduced development and maintenance time, greater success in deployment of new enterprise capabilities, improved user access/usage, and better understanding of your corporation's assets. For corporations pursuing Enterprise Architecture, Metadata serves as the interstitial glue that links business processes to data to applications to the technical infrastructure. Some of the important benefits of the Metadata Management are discussed below.

Enabling Reuse of Corporate Assets

To increase reuse value, you must have a consistent set of Metadata attributes captured about all assets targeted for reuse. Examples of Metadata attributes might include author, definition, creation date, technology, and expiration date. This Metadata defines the key attributes of the corporate assets available for reuse and, through

this Metadata, corporations can realize significant reductions in Total Cost of Ownership (TCO), drive reduction of redundant assets, and enable agility through maximized return and increased utilization of their unique corporate assets. Reuse benefits through effective Metadata capture about corporate assets can also enable more consistent business process implementation and accelerated Enterprise Architecture adoption as the associated assets can be leveraged for subsequent instantiations. This is especially important when corporations want to maximize their resources on new capabilities versus reinventing existing assets due to lack of knowledge of their existence. Industry data has shown that upon discovery of reusable assets, the cost of reusing the asset is approximately 20% of the cost of creating the asset anew. Using a Metadata Repository also increases the impact of the reuse paradigm within corporations from one of ad hoc reuse (based on analyst, developer, or engineer interaction) to an enterprise-wide prescriptive reuse capability (where all assets are made available and mapped to the Enterprise Architecture for quick assessment and implementation).

Improved Impact Analysis

Enterprise data volumes are doubling every 12-18 months–this data explosion is fueled by data from inside corporations, data throughout the supply chain, and data from new sources. Likewise, the data growth is further exacerbated by the rapid adoption of XML as XML is roughly 10%-20% less efficient (in terms of size) as an interchange standard. Today's enterprise infrastructure is poorly suited for dealing with this continuing, rapid explosion in data and must rely on Metadata to maintain a manageable view of corporate assets. Through Metadata, a simplified infrastructure and a move toward a services-oriented architecture are essential to provide the needed impact analysis and associated business intelligence on this data growth that is both cost effective and timely.

Examples of Metadata captured for impact analysis, such as those discussed in usage scenarios, include the following:

- Business Metadata definitions and business rules.
- Data Metadata definitions.
- Data transformation rules, mappings, and processes.
- Data lineage (from models to database/tables to report usage).
- Application Metadata definitions.
- Technical Metadata definitions.
- Dependency across business processes, data, applications, and technology.

Without this impact analysis, Metadata analysts, developers, and engineers must spend an enormous amount of time in discovery mode searching out assets and then uncovering the relationships between assets; only a finite amount of time remains to put the asset(s) to use. To invert this 80% research and 20% deployment paradigm, and thus improve the user's interaction with corporate assets and raise productivity, companies must focus on the Metadata Management processes to collect this impact analysis Metadata (including the definition AND its context in the overall enterprise).

Increased Data Quality for Decision-Making

Corporations are quantifying the costs they are incurring due to inadequate, inaccurate, or incomplete data that has a direct correlation with the quality of decision-making. At the core of these problems lies a lack of rigor around the definition of data and their characteristics, or Metadata. By establishing Metadata processes to ensure needed Metadata definition and capturing this Metadata in a centralized Repository, Metadata Management ensures data consistency, thus enabling better business decisions. Corporations have identified data quality as key to their success, and have identified credible Metadata as the cornerstone of information quality.

Reduced Development/Maintenance and Increased Success in Enterprise Deployment

Metadata Management streamlines the application development process by reducing the development, deployment, and maintenance costs. This is made possible by the use of a single source of Metadata for logical, physical, and process aspects of the application environment, while tracking versions of the code and documenting all aspects of the application development life cycle. By providing a complete, integrated view of the development environment, Metadata helps identify redundant processes and applications, thereby reducing duplicated efforts. Developers can share and reuse existing objects such as data structures, programs, model definitions, and more. In addition, Enterprise Impact Analysis greatly reduces the analysis and maintenance phase of the development life cycle.

An example of measured results reported by one of the development teams within Intel's manufacturing group found that, based on Metadata reuse, for every $1 invested in using the Metadata Repository, $6 was saved in reduced development and sustaining costs for their Decision Support Systems (DSS) applications. Meta Group's recent "Data Warehouse Scorecard and Cost of Ownership" industry study quantified the enormous effect Metadata has had on the overall success of data warehouse initiatives. It reported that 75% of highly

successful data warehouse initiatives included formal Metadata Management facilities.

Metadata is also the foundation of effective enabling of future initiatives such as model-driven, event-driven, service-oriented architecture, Service-Oriented Development of Applications (SODA), Metadata-driven entitlement/security and role-based access control, and next-generation business intelligence architecture. "Through 2008, enterprises that use a Metadata Repository as part of a reuse program can develop SODA applications at least 20% more quickly and less expensively than those that do not (0.8 probability)." [1]

Understanding Corporate Assets through Metadata

Establishing an understanding of corporate assets via Metadata Management processes and a centralized Repository can benefit business users as well as the Metadata-producing organizations within a corporation. By providing a Metadata "card catalog" and implementing an enterprise data dictionary, business users can find processes, data, applications, technology, and reports more easily. In addition, the Metadata-producing organization can record the knowledge about assets for future use. As personnel change within an organization, institutional knowledge can leave that organization and undocumented assets can quickly lose their meaning and/or value. Subsequent employees may have little or no understanding of the corporate assets and may find they can't trust results generated from these assets if there is no context. This is especially important for documentation on legacy assets.

CONCLUSION

Corporations are beginning to see the value of Metadata to the business, and not just as a technology asset. Corporations that develop enterprise architecture frameworks that tie together business strategies, business processes, and data and instantiate the applications and supporting technologies based on those processes/data will benefit from Metadata Management. By focusing on Metadata Management processes, Intel expects to continue its approach to using standard deliverables defined with repeatable development processes to capture credible Metadata, and integrate that Metadata into the enterprise architecture framework. Over the past six years, since forming a centralized Metadata capability at Intel, Metadata usage and importance has increased in our environment. Since the latest iteration of our Enterprise Metadata Repository, we have seen a 400% increase in its utilization. We have also had a subsequent request for newer and better functionality in order to engage usage methods and customers that we had not initially realized would find value in the tool.

Metadata must be managed to be of value and connected to increase value. The value provided by a Metadata Management program will accelerate over time as the enterprise architecture framework continues to be built out and deliverables providing credible Metadata content are incorporated. Continued engagement in all project starts guarantees that Metadata gathering methods are employed up front, and our automated methods for refresh are put in place early. The best Metadata is accurate Metadata always. For further reading, see references [2-7].

ACKNOWLEDGMENTS

We would like to thank all of those involved in the creation of the Enterprise Metadata Repository and the Metadata Management program.

REFERENCES

[1].Michael Blechar, Research VP, Gartner.

[2].Brackett, M. H., *Data Resource Quality: Turning Bad Habits into Good Practices*, Addison Wesley, Boston, 2000.

[3].English, L. P., *Improving Data Warehouse and Business Information Quality*, John Wiley & Sons, Inc., New York, 1999.

[4].Marco, D., *Building and Managing the Meta Data Repository: A Full Lifecycle Guide*, John Wiley & Sons, Inc., New York, 2000.

[5].Redman, T., *Data Quality: The Field Guide*, Digital Press, Boston, 2001.

[6].Redman, T., *Data Quality for the Information Age*, Artech House, Boston, 1996.

[7].Tannenbaum, A., *Metadata Solutions: Using Metamodels, Repositories, XML and Enterprise Portals to Generate Information on Demand*, Addison Wesley, Boston, 2002.

AUTHORS' BIOGRAPHIES

Thiru Thangarathinam has been with Intel for two and a half years and has extensive experience in architecting, designing, and developing N-Tier applications using Service-oriented Architectures and related technologies. Thiru is currently the application architect of the Enterprise Metadata Repository application within Enterprise Data Architecture. He was part of the team that recently developed Intel's Metadata Repository and implemented the Enterprise Metadata Repository. Thiru has also coauthored a number of books on Service-oriented Architectures and has also been a frequent

contributor to leading online technology-related publications. His undergraduate degree is from The College of Engineering, Anna University, Madras. His e-mail is thiru.thangarathinam at intel.com.

Gregg Wyant is the chief data architect and director of Enterprise Data Architecture in the Information Services and Technology Group. With more than sixteen years at Intel, his career spans architecting Intel's early desktop and mobile solutions to driving software vendor optimizations for processor instruction set enhancements, to authoring the Ziff-Davis book *How Microprocessors Work* (translated into seven languages and used as the basic primer for Intel technology). He holds a B.S. degree in Computer Engineering from Iowa State and an MBA degree from California State University, Sacramento. His e-mail is gregg.wyant at intel.com.

Jacque Gibson has been with Intel for fourteen years and has experience in Production Planning and Information Technology. As manager of Product Data Management, she participated in Intel's initial ERP implementation. Jacque is currently the manager of Metadata Services within Enterprise Data Architecture. She and her team recently developed Intel's Metadata Program and implemented the Enterprise Metadata Repository. Her undergraduate degree is from Michigan State University. Her e-mail is jacquelyn.a.gibson at intel.com.

John Simpson is the technical programming lead and project manager within Enterprise Data Architecture for the Information Services and Technology Group. In his ten years with Intel he has moved from factory support services into Information Technology, launching one of the first database-driven applications on the company Intranet. His undergraduate degrees are from the University of the State of New York and the University of Phoenix, from which he also holds an M.S. degree in Computer Science. His e-mail is john.e.simpson at intel.com.

All of the authors are part of the Enterprise Data Architecture's Metadata Services team that recently won the 2004 DAMA/Wilshire Award for "Best Practices in Metadata Management" at the DAMA conference in May 2004. This is the highest award given in the data management industry for Metadata practices, and it is given to corporate data organizations to recognize business value, innovation, and excellence in design and/or implementation of Metadata as a critical component in data management success. Intel was selected as the winner based on the specified corporate cost reductions and productivity improvements as well as on its implementation of corporate data management practices.

The award press release and details can be found here: http://www.wilshireconferences.com/award/index.htm. Data Management Association (DAMA) International is an international not-for-profit association of data resource management professionals. DAMA International's primary purpose is to promote the understanding, development, and practice of managing information and data as a key enterprise asset. With presentations on topics ranging from Business Intelligence to Data Quality to Enterprise Architecture to Process Modeling, industry experts and corporate practitioners shared their strategies and success stories with the conference attendees.

Successful Application of Service-Oriented Architecture Across the Enterprise and Beyond

George Brown, Information Services and Technology Group, Intel Corporation
Robert Carpenter, Information Services and Technology Group, Intel Corporation

Index words: SOA, SOE, Business Process Automation, BPM, BAM, SOA, federated enterprise, business process model, value chain, reference model

ABSTRACT

This paper is divided into two major sections. In the first section we look at evolving service-oriented architecture to the service-oriented enterprise, and in the second, we examine integrated process and technology reference models.

The literature abounds with articles about Service-Oriented Architecture (SOA) and particularly Web service-based SOA. Yet, to be successful within a real enterprise, there needs to be more than simple agreement over protocol, data packaging, service invocation, and discovery. While the latter are essential to SOA by providing a foundation for the creation of services, they do not describe how those services function within the overall enterprise environment. They do not describe, for example, how to deliver reliable, scalable enterprise business processes built upon SOA–i.e., the Service-Oriented Enterprise (SOE). The literature on this topic is remarkably sparse.

In the first section of this paper, we explore SOA from the perspective of an SOE where the primary service building blocks execute within a managed SOA eco-system. This managed SOA ecosystem relies upon orchestration, business rules, and process correlation to provide monitoring and management, and business process analytics and autonomics. This framework weaves SOA into the enterprise fabric providing a rich service framework within which to implement enterprise business processes. We begin with an overview of SOE, and then discuss a technology development project to create a Business Rules service. Several usage models are discussed.

In a globally connected economy, enhanced collaboration increases visibility of critical supply and demand changes and streamlines the flow of information that governs critical design and supply chain decisions. The main barrier to digital supply chain collaboration is complexity and high costs associated with extensibility of business processes throughout the network of suppliers and customers. Business process extensibility is the ability of an enterprise to conduct digital collaboration with its partners, by extending processes throughout the supply chain to increase visibility and streamline the information flow.

Enterprises must be enabled to expose their business processes through SOA while retaining and extending the capabilities normally found in tightly integrated vertical applications. We must comprehend fully those collaboration models and the architectural elements necessary to support them.

In the second section of this paper, we explore methods for modeling and implementing business processes within the value chain as an integrated system and introduce a set of principles and guidelines for defining a reference architecture that consistently and accurately represents the value chain as an SOA. We explain how this can be accomplished with a reference model for modeling, simulation, and benchmarking business processes and with a reference architecture that supports deployment and management of collaborative processes within the value chain.

Our research with Collaborative Product Development Associates introduced us to a Federation Enterprise Reference Architecture (FERA*), a reference architecture that extends SOE to the federated enterprise.

* Other brands and names are the property of their respective owners.

GENERAL INTRODUCTION

As enterprises embark on the path to e-business, particularly supply chain management, they are challenged to abandon traditional modes of thinking about their business processes and computing infrastructure and to embrace an entirely new paradigm. Processes that were traditionally scoped entirely within the boundaries of an individual enterprise now participate as elements in a collaborative process within a value chain across a federation of enterprises.

At the same time as enterprises are challenged to cope with this new paradigm, they must also come to grips with fundamental changes in their own computing infrastructure. Most notable among these is the emergence of Service-Oriented Architecture (SOA) and particularly the Web service version of that architecture.

SOA creates problems for an enterprise on two distinct levels. First, the enterprise must come to grips with the complexities of collaborative processes across a federation of enterprises. To do this, it must develop a set of principles and guidelines for defining reference architecture that consistently and accurately represent the value chain as an SOA. This can be accomplished with a conceptual architecture for implementing technology to support modeling, simulation, deployment, and management of collaborative processes within the value chain. This architecture does not specify the detailed semantics of business transactions between the collaborative partners; rather, it distills from the structure of collaboration the key capabilities that must necessarily be present to support them. It becomes the framework for expressing those detailed semantics.

Second, enterprises must come to grips with the fundamental changes in their computing environments. SOA is creating a not-so-quiet revolution in the nature of business computing. Enterprise Resource Planning (ERP) vendors and many others are beginning to incorporate SOA into their products, and enterprises are beginning to incorporate SOA into their computing infrastructure. These directions are positive, but SOA, by itself, provides neither a framework for implementing business processes within an enterprise nor a framework to expose those processes to collaborative participation within a federation of enterprises. These goals can be achieved through a managed SOA ecosystem where process is implemented and managed across the "sea of services."

In the first section of this paper, we first develop a blueprint for the successful application of SOA across the enterprise and beyond. Our approach is decidedly and unabashedly process centric. We begin with a discussion of SOA exploring its limitations. We then propose an architecture to counteract those limitations. This architecture reconstructs the process from the various pieces that execute in discrete process controllers and services and provides a framework for the control of those processes, through an analytics framework and a business rules engine. The result is a flexible, agile, managed SOA ecosystem that supports process analytics and process autonomics across a heterogeneous process control environment within the enterprise. This architecture takes SOA and evolves it to a Service-Oriented Enterprise (SOE) where enterprise processes can be implemented in a consistent and robust way. Second, we discuss the results of two technology development projects that investigate this architecture and present an example that demonstrates the architecture's integrative capabilities. The investigations validate the architecture's value in evolving SOA to SOE.

Third, we return to the complex question of collaborative processes across a federation of enterprises. This inquiry is motivated by the fact that the main barrier to digital supply chain collaboration is the complexity and high costs associated with extensibility of business processes throughout the network of suppliers and customers. Given that the enterprise has evolved to SOE, it still faces a bewildering array of interaction models that have led to "one off" solutions that are neither consistent nor extensible. It is not that the SOE model has failed but rather that its scope needs to be expanded to comprehend collaboration models and business semantics across a federation of enterprises.

What is needed is a set of open source, standardized, integrated frameworks or reference models that would dramatically improve an organization's ability to design, develop, and maintain their business processes. The perceived value of a set of integrated models is far greater than the individual frameworks alone. To explore this proposition, the state of the art is reviewed and then the Federated Enterprise Reference Architecture (FERA) is described in some detail. FERA is proposed as a framework for providing those frameworks and reference models. To validate the potential of FERA, we examine how specific collaboration models are supported by components of the architecture. The results are both positive and promising.

The goal of this paper is to describe an integrated process and technology framework where business processes are enabled both within the enterprise and within the federation of enterprises. At the core of the exposition is the business process, which must be supported by correlation, process analytics; and process autonomics within the enterprise, and must be supported by standardized, integrated frameworks or reference

models within the federation of enterprises. Throughout, the SOA paradigm is maintained but more than simple services are exposed; the business processes themselves are exposed within and across enterprises. While the two sections of this paper may seem distinct, they are not. Our goal in these next two sections is to show how the evolution of SOA to SOE is nothing more than the architecture for the enablement of the process itself within the enterprise. The development of integrated frameworks and reference models across the federation of enterprises is nothing more than the architecture for the extension of process beyond the enterprise walls. Together they build an integrated process and technology framework to enable business processes for the e-corporation.

EVOLVING SERVICE-ORIENTED ARCHITECTURE TO THE SERVICE-ORIENTED ENTERPRISE

Our research explored SOA from two different perspectives.

1. In early 2004, we looked at the role of orchestration within SOA firstly from a practical, integrative perspective. The context for that investigation was the belief that orchestration engines, as process automation controllers, will function within a heterogeneous process control environment. This belief is driven by emerging capabilities in applications, ERP and others, and by the limited usability domains of individual product offerings.[1]
2. In the summer of 2004, we looked at the creation of generic services to support an SOE environment by using orchestration itself both as the process controller and as the methodology for exposing the service. A business rules engine was chosen as the object to expose, and several use cases were explored. The goal was to examine how a general-purpose service could both function within an SOE and support the SOE vision.

Interesting as these research activities were, they beg the question of what an SOE is. It is too early to give a definitive answer as the literature using this term has, by no means, reached consensus.[2]

For this reason, we focus on the more modest issue of the service taxonomy and architecture upon which an SOE will necessarily rely. For the purposes of this paper, a working definition of an SOE is as follows:

*A **Service-Oriented Enterprise** is an enterprise that implements and exposes its business processes through an SOA and that provides frameworks for managing its business processes across an SOA landscape.*

At the core of the SOE is the business process that should be elevated to an explicit object within the enterprise. It matters not whether the process in question is contained entirely within the enterprise or whether the process spans firewalls between "enterprises." As virtual enterprises grow in popularity, and as extended collaboration becomes a reality, internal and external processes will necessarily overlap and some issues around security and control will remain unresolved for now.

With process as the focus, it should come as no surprise that the traditional elements of SOA (discovery, protocol, data packaging, and service invocation), while assumed to exist within the computing infrastructure, are not the focus of SOE *per se*. These elements describe how to construct services and how to use services. They do not describe how sets of services support enterprise business processes or how atomic services function within an enterprise.

The central challenge facing the enterprise approaching SOA, and Web service-based SOA in particular, is how to implement business processes within an enterprise in such a way that the process is visible and manageable end-to-end. As the number of services available within the enterprise increase and as the services become increasingly intelligent, their execution pattern becomes increasingly difficult to define *a priori*[3].

The central contention of the SOE architecture is that more than mere inter-service infrastructure SOA is required to support SOE: a managed service ecosystem is required.

A Managed Service-Oriented Environment

Enterprises have come to expect a high degree of integrated capability from existing vertical applications. As a general rule, they provide a solution stack containing the following:

1. Infrastructure monitoring
2. Connectivity
3. Application monitoring and management
4. Application-level work tasks
5. Domain-specific logic
6. Program flow control
7. Presentation

Increasingly, enterprises have come to realize that their business processes are not strictly contained within the domain of a specific vertical business application but, rather, span multiple applications within and across enterprises. This, in turn, raises the question of how to

provide the same level of capability for processes when they are no longer executed within a single, integrated vertical application.

At a simplistic level, one could regard the applications as services and "orchestrate" a process as a sequence of steps each of which is contained within a separate application. The sequencing could be done formally by an orchestration engine, by an existing job scheduler, or by a real-time event model. With this approach, one has a functioning process but there are drawbacks. The fact that each application defines and manages its own stack independently creates a complex integration problem to answer the seemingly simple question: what is the state of the process now? More complex questions such as, what is the probability that a given process instance will complete within the prescribed SLA, remain elusive. What is missing in this approach is an overall context for the execution of the "services", i.e., a managed SOA ecosystem.

Solving these problems requires a fresh approach, one that views the process execution framework holistically. Taking this fresh look reveals a huge opportunity for refactoring and harmonization across the application landscape.

Refactoring the Application Landscape

The refactoring process has two phases: first, the elements of the stack need to be mapped to generally available enterprise services; second, additional services and frameworks need to be identified that complete the managed SOA ecosystem and provide the foundation for enterprise process management. At the very least, the first task is relatively straightforward, although the specific implementation within a given enterprise will map to its specific service offerings.

Table 1: Application stack to Service Mapping

Stack Element →	Service
Infrastructure monitoring	Std. Infrastructure monitoring
Connectivity	EAI
Application monitoring and management	Correlation, monitoring and management [BAM]
Application-level work tasks	Individual services
Domain specific logic	Business Rules Engine
Program flow control	Process orchestration and human workflow services
Presentation	Unified delivery, portal

At the heart of this exercise is the substitution of reusable, enterprise services to replace individually coded instantiations of capabilities within vertical applications. It represents the disaggregation of the vertical application into services. Most such services are obviously reusable. Initially, *Application-level work tasks → Individual services* may appear single use, but many will, in fact, be found to support reuse. In any event, the identification of such logic into discrete services enhances the potential for refactoring at a later point. The standardization of these services within the overall SOA framework achieves some immediate advantage in simplifying the management of the environment.

In addition to the basic mapping exercise, the following capabilities are essential ingredients of a managed SOA:

1. Process management across autonomous services.
2. Enterprise services to support processes.
3. Process monitoring and management.
4. Process analytics.
5. Process autonomics.
6. Integration of execution layer with high-level business modeling.

Together, this leads to a conceptual architecture:

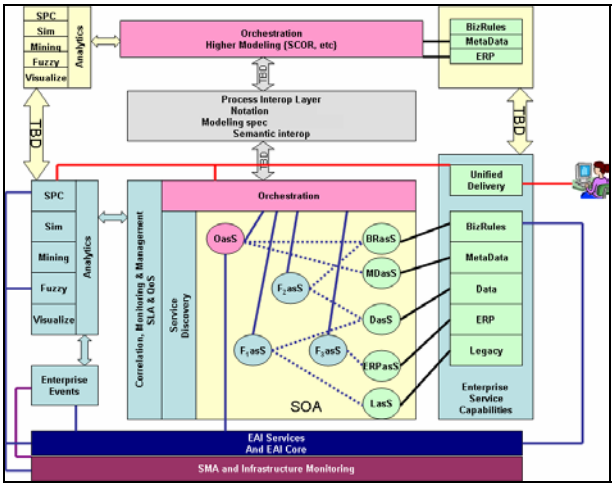

Figure 1: SOE conceptual architecture

The salient features are as follows:

1. Enterprise service capabilities represent (a) legacy applications and data that are wrapped and exposed as services within the basic SOA, [e.g., Data→DasS] and (b) new services, such as business rules. Authentication and authorization would be additional examples.
2. Within the managed SOA domain, services, orchestration, service discovery, correlation, and management provide the bare bones of a managed service environment.

3. The analytics piece relates correlation and management back to the execution environment through business rules to achieve autonomics.
4. Finally, principally through orchestration, business rules and data, higher-level process modeling is tied to the basic execution environment through a process interoperability layer[4].

Together this architecture provides a framework for creating and managing processes within a managed SOA ecosystem.

Correlation: Problem and Opportunity

It is a principle of managed SOA that processes are, where possible, under the control of an orchestration engine. The engine controls the flow of a process and invokes the individual services. For a given process, there may be one or many orchestration engines. These engines, $\{P_k\}$, may be arranged in several models:

1. Single instance invocation
2. Hierarchical
3. Peer-to-Peer. This is often managed by reducing the peer-to-peer model to a quasi-hierarchical hub and spoke model.

In practice, even relatively straightforward processes involve multiple process controllers. This is, in part, due to incorporation of process capabilities into Business-to-Business (B2B), ERP, and Extract, Transform, and Load (ET&L) products. An example will illustrate.

1. A company's manufacturing capacity is in constrained state according to its business rules.
2. A sales order is received over the standard B2Bi environment (B2Bi is P_1).
3. Because it is in constraint, the order is diverted to an analyst for review in a human workflow process (HWS is P_2).
4. The analyst requires a data refresh and cube rebuilds involving a host of data sources. This is managed by a process-based ET&L tool [ET&L is P_3].
5. The analyst concludes the HWP that notifies stakeholders, and executes against the ERP system, which may well have its own processes to execute before placing the order [ERP is P_4].

This is just the input side of the process. It has an external business rules engine in #1 and process controllers $\{P_1 , P_2 , P_3 , P_4\}$ in #2-5. When scenarios like this are extended across enterprises, the number of process controllers will, quite naturally, increase.

Enterprises demand answers to questions such as (1) What is the state of the process now? (2) Where is the order? (3) What are the processing statistics for orders of this type? (4) How do we analyze what has happened and design a better process? (5) How could we introduce SPC or autonomics into order handling? (6) What can we discover about this process?

The problem is that the entire process, $\mathcal{P}$, maps not to a process controller, which could in principle manage flows under its exclusive control, but to a set of them $\{P_k\}$ executing in an order determined not by an *a priori* model, but by logical outcomes based upon the data being processed. Recovering information about $\mathcal{P}$ is difficult, at best.

The solution to the problem has two parts. First, the parts of $\mathcal{P}$ executing in the different $\{P_k\}$ must be associated. This is the correlation problem. Second, information about $\mathcal{P}$ must be extracted from the correlated pieces in the $\{P_k\}$. This will feed the analytics and autonomics engines.

Consider the design pattern *Correlation Identifier*. It has been used to provide a method for associating a response with a request[5]. It is also being adopted and extended in the Web Services Business Process Execution Language[6] [WS-BPEL] where "correlation sets" are defined for similar purposes.[7] For our purposes, tthese standard-use cases, associating a response with a request, are too narrow in scope.

What is needed is an extension of the notion of a correlation identifier[8] or correlation set to meet the needs of process management within a managed SOA. Consider a correlation identifier that has the following properties:

1. It is an extensible xml object which is extended whenever
    a. receipt of process invocation, i.e., → P_k
    b. optionally during management by P_k
    c. at a process handoff P_k→ P_j
    d. processing ends at P_k
2. Its extensions [leaves] contain
    a. process information, sequence number within a path in $\mathcal{P}$, invoking system, invoked system [optionally tracing the call sequence to underlying services providing the "work" of the process]
    b. process context information, including but not limited to (i) a mapping to physical infrastructure, (ii) process partners, etc.,

and (iii) a key for mapping the activities of $\mathcal{P}$ in P_k against P_k's operational repository.[9]

c. process value adornments, e.g., value of sales order, etc., as name-value pairs.[10]

At each stage of $\mathcal{P}$, C_ID is extended with the state of the process. The next step is to collect the correlation identifiers into an analytics environment. Standard Enterprise Application Integration (EAI) such as Web Services (WS) can mediate the collection. Such a model leads to an architecture as shown in Figure 2.

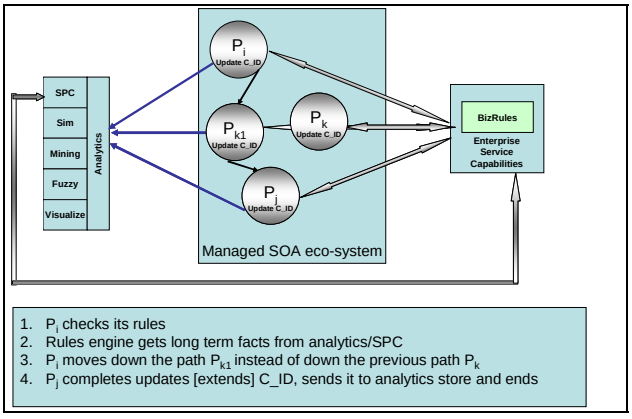

Figure 2: Correlation architecture

Individual process controllers implementing a business process $\mathcal{P}\{P_i , P_k , P_{k1} , P_j\}$ update the analytics environment with the correlation identifier, C_ID, as described above, with basic data and extended data as specified by the enterprise business rules. The first job of the analytics environment is to reconstruct a navigable process from the collection of C_ID. Fortunately, most will be simple linear chains or trees [directed, acyclic nets]. Since at → P_i , P_i receives C_ID and at P_i→{ P_j , P_k , P_l ..} [11] it passes it on in an extended version of C_ID, the overall reconciliation of the collection of C_ID over a path is not too difficult.

Once the process is reconstructed in the analytics environment, it can be subjected to standard process execution analysis[12], statistical process control, simulation with appropriate tools, data mining, fuzzy logic and heuristics, and process visualization.[13]

With this framework, correlation solves the initial problem of associating the segments of $\mathcal{P}$ executing in the different $\{P_k\}$ and creates the opportunity for a broad spectrum of process analytics opportunities.

Process Autonomics and Business Rules

With the correlation identifier strategy defined, we can now tackle the second part: how to link information about $\mathcal{P}$ to an autonomics framework.

Figure 2 illustrates that framework by providing integration of the managed SOA ecosystem to the analytics environment through a business rules engine.[14] By designing processes to be rules dependent, the control loop can be completed to enable autonomics.

Most rules engines evaluate rules against a set of facts and those facts come in two forms: instance facts are submitted to the rules engine when the rule is to be evaluated, and environmental facts are longer term and persistent within the rules environment. The integration occurs by taking the results of the analytics, (e.g., SPC computations, infrastructure alerts, etc.), and propagating them to the rules engine as events over the standard EAI bus. The rules engine stores them as environmental facts. Then, when a process executes, it submits its instance facts to the rules engine where, based upon the logic of the rules, they are combined with environmental facts to render a decision back to the process. Control can be exercised over the process through the analytics framework and through the rules engine. A simple example based upon infrastructure alerting will illustrate the principle.

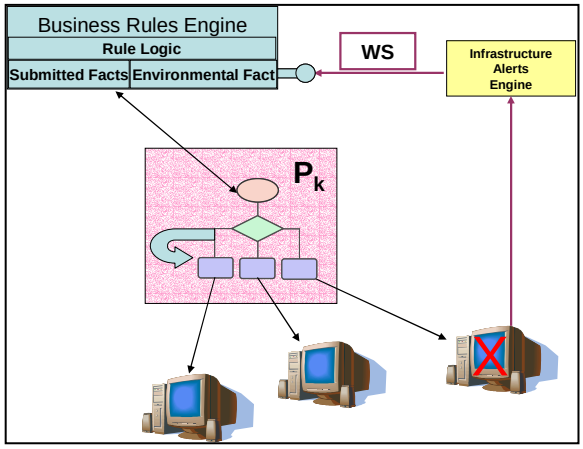

Figure 3: Infrastructure rules integration model

In this model, infrastructure events are sent to the infrastructure events engine (operations console) where an alert is sent via Web services to the rules engine. When a service goes down, the rules engine receives the event and stores it together with the expected time to availability as an environmental fact. When a process, $\mathcal{P}$, starts under control of P_k, P_k calls the rules engine to determine if the infrastructure is ready. The submitted process is checked against its dependencies and the non-availability stored in the environmental fact base. The rule computes the likelihood of completion within the SLA based upon the estimated availability time. P_k is instructed to proceed, resubmit in x minutes, or abort as appropriate. Additional analytics-based and

infrastructure-based process control scenarios are easily developed within this architecture.[15]

In the technology development activities ending in the summer of 2004, we evaluated a business rules engine in this architecture against the following scenarios:

1. Provide switches in other processes.
2. Provide thresholds in other processes.
3. Alternative source parameterization.
4. Add content to processes based upon business domain logic.
5. Demonstrate rules as an integration point.
6. Demonstrate rules-based reflective programming [variant of #4].
7. Demonstrate forward chaining for complex automation.

Each of these allowed us to provide a level of autonomic process control from a rules environment and allowed us to explore different sourcing options for the environmental facts. An additional benefit was that with rules-based processes, the process could be kept more generic driving the potential for reuse. When measured against the critical success indicators of the project,[16] the rules-based integration approach proved successful. The framework provided the richness necessary to move SOA toward SOE.

Challenges and Solutions

The challenge to move SOA toward SOE is met in large part by the architecture described here. The service taxonomy and architecture provide a framework that enables enterprises to expose their business processes through SOA while retaining and extending the capabilities normally found in tightly integrated vertical applications.

The principal challenge will be to evangelize the correlation identifier as used in this paper and to standardize its elements. While the architecture is not strictly dependent upon it, there will need to be efforts in the supplier ecosystem to drive acceptance of the model and to begin to develop an open standard for implementation. The advantages of doing so are enormous for the company seeking to move past SOA to SOE.

An additional challenge is to spur work on the process interoperability layer to define the standards and mechanisms needed to create transparency between models describing collaboration scenarios within a federation of enterprises, and models executing processes within an enterprise under this architecture.[17]

To enable enterprises to expose their business processes through SOA while retaining and extending the capabilities normally found in tightly integrated vertical applications, more than mere inter-service infrastructure SOA is required: a managed service ecosystem is required. This ecosystem leads to a natural division: enterprise services, managed SOA environment, analytics framework, and integration to higher-order modeling.

The challenge of the heterogeneity of the process control environment in turn drives the necessity of correlating the process segments of $\mathcal{P}$ executing in disparate controllers { P_j , P_k , P_l ..}. This is resolved through the correlation identifier, C_ID, extended by each P_k and managed in the analytics environment. C_ID provides basic process execution data, processes context and values for analysis, and provides the pointers back to the operational repositories of the $P_{k,}$ which allow for detailed tracing and trouble shooting.

The analytics environment together with the infrastructure management environment then provide input to the rules engine that serves as the integration point to the execution environment. This completes the autonomics control loop. A basic framework for exposing processes is achieved and the first steps from SOA to SOE are realized.

The significance of this research and technology development is that it solves the key challenge posed by emerging product capabilities. It provides a product independent framework for the correlation of processes across a heterogeneous process environment. The correlation identifier together with business rules then provides the foundation of process analytics and process autonomics moving the enterprise from the "sea of services" of an SOA architecture to a managed SOA ecosystem. The enterprise moves from SOA to SOE.

Equally importantly, this architecture creates a managed process execution framework for the enterprise that enables it to participate within a federation of enterprises precisely because the execution of externally triggered processes are managed and visible end-to-end within the enterprise. It provides the control necessary to guarantee reliable participation in collaboration scenarios across a federation of enterprises. The next step is to comprehend fully those interaction models and the architectural elements necessary to support them. The next section does just that.

INTEGRATED PROCESS AND TECHNOLOGY FRAMEWORK

In a globally connected economy, enhanced collaboration increases visibility of critical supply and demand changes and streamlines the flow of information that governs critical design and supply chain decisions. As a result, companies can better coordinate their internal product development plans and their design and engineering effort, and they can bring more new products to market faster while improving on-time delivery and reducing inventories and costs. However, information technology solutions currently used to support supply chain collaboration have at best only scratched the surface of the total opportunity.

The main barrier to digital supply chain collaboration is complexity and high costs associated with extensibility of business processes throughout the network of suppliers and customers. Business process extensibility is the ability of an enterprise to conduct digital collaboration with its partners, by extending processes throughout the supply chain to increase visibility and streamline the information flow.

Global economic pressures mandate that manufacturing companies find ways of faster and more intelligent collaboration throughout the supply chain. And while there are processes that can be integrated using relatively simple forms of collaboration, the vast majority of the most critical value-added collaborative processes are non-linear, non-deterministic, and multi-disciplinary in nature.

Integrated business process management and information technology architecture is critically important in the increasingly complex global operations. The standardized business process is a critically important factor for sustainable differentiation in global business as it provides baseline metrics for continuous improvement in understandable terms, and it contributes to the formulation of consensus across the participating companies in the value chain. An integrated approach is to utilize process reference models along with reference architectures representing a variety of technology deployment components serving loosely coupled collaborative processes that can be reconciled through direct mapping from the process models, defining business semantics in plain, non-technical language.

Such a reference architecture, described here, is based on the federated approach, where loosely coupled federated systems connect to each other through gateways based on common business semantics and open standards. This approach utilizes Web services for system integration and workflow orchestration and open standard business semantics for describing the federation and its working principles. Without predetermining and imposing any specific workflow and without entailing any additional investment on the part of the participants, this approach can penetrate as many tiers of the supply chain as required.

Service-oriented architectures generally presume a collection of services that enable the business much as pre-SOA infrastructure has. A case in point is the Service Reference Model for FEA, the Federal Enterprise Architecture as shown in Figure 4.

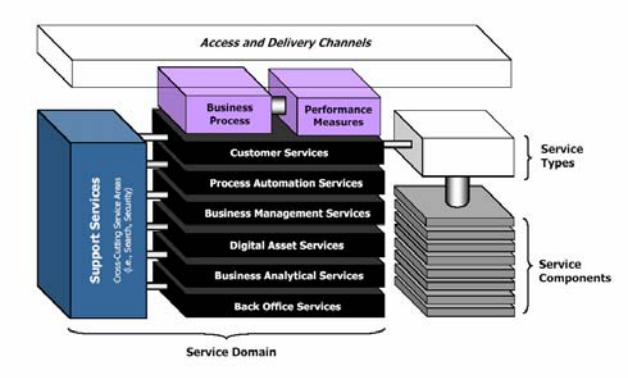

Figure 4: Service Reference Model for FEA

What is missing from these generic architectural patterns is any recognition of how business semantics determines the services that are defined. Moreover, there are no patterns to define how those service-based processes are instantiated. Lacking such patterns diminishes the distinct advantages of the well-formed architecture for the SOE as discussed earlier.

FERA[18] has potential to be the collaboration framework in a future state architecture that enables enterprises to expose their business processes across a federation of enterprises through SOA. To validate the potential of FERA, we examine how specific collaboration models are supported by components of the reference architecture. We look at how certain collaborations in process scenarios from different communities of businesses found in our value chain are enabled by FERA.

Defining of Business Semantics

It is a well-documented fact that business and IT managers rank business process improvement as their number one priority. Companies that are process-focused have well-defined processes that are both vertically and horizontally aligned, are well-governed, produce cost-

effective and reliable outcomes, and are able to adapt more quickly to market changes.

The corporate focus now shifts away from the speedy transformation of materials through the supply chains to understanding what drives product value chains. In turn, the fundamental organizational model may shift to promote better transparency and synchronization of the extended value chain. For example, centralized management may split into federated business units to reduce the scope in terms of the breadth of product coverage in each unit, and to tie product metrics more tightly to such criteria as time to market, acceptance, and profitability. For another, outsourcing of product design and manufacturing may not only enhance focus, but may also share the responsibility across a larger resource pool. Even the delivery model may shift from an emphasis on product manufacturing to service or solutions. In the process, corporations look for control and balance of the three major business payoffs that arise from total product value-chain management: time to market, market acceptance, and profitability.

Intel's IT research agenda in Business Process Systems has focused on the use of reference models to represent process scenarios from different communities of business with the intent of generating reusable templates that can accelerate the integration of trading partner business systems/processes into the supply network. This work started with application of the Supply Chain Operations Reference (SCOR) model and expanded to successful application of enterprise-wide reference models. In order to consolidate these business process reference models into a usable value chain reference model, it became clear that the reference models must be coupled with a collaboration framework that facilitates implementation of process scenarios across the value chain.

A set of open source, standardized, integrated frameworks or reference models would dramatically improve an organization's ability to design, develop, and maintain their business processes, and the perceived value of a set of integrated models is far greater than the individual frameworks alone. A general consensus has developed among partners developing essential collaboration models for product design for supply chain that the long-term value proposition is to focus on a Value Chain Operations Reference model (VCOR).[19]

Defining business semantics in terms of the common vocabulary of VCOR aggregates business applications and business processes to a higher level of abstraction. In this way, value chain integration enables coordination across departmental, organizational, and enterprise boundaries from an overall business-level perspective. The benefit is that it facilitates service-composed processes and, thereby, brings service-oriented relevance to a complex IT landscape in which ongoing, flexible adaptation is necessary.

Figure 5: Value Chain Operations Reference model, VCOR

Mapping of Business Process to Core Collaboration Capabilities

To classify information exchange patterns that can take place in a federated enterprise framework, one has to consider two dominant characteristics: how is information shared and how is the business context of collaboration administered and preserved. Because the information content and context are decoupled, a combination of the two process characteristics determines classes of collaborative patterns in a federated enterprise.

Information sharing has three patterns: people to people, systems to systems, and people to systems (and vice versa). In people to people collaboration obviously information gets exchanged using a vocabulary and semantics that enable all parties to properly interpret the content and context of collaboration. In systems to systems, information exchange semantics are left to systems to interpret and process. Business context administration has two patterns: centralized and distributed. Where a centralized pattern exists, a common business logic supports the entire scope of collaboration; where a distributed pattern exists, a separate business logic gets involved, thus some shared business semantics

need to reconcile the context of collaboration between semantics specific to each participating domain.

There has been considerable work done in defining and implementing e-Workforce solutions to facilitate synchronous and asynchronous collaboration in distributed environments for Classes 1, 2, and 3 as shown in Table 2.

Table 2: Synchronous and asynchronous collaboration in distributed environments

	People-to-People	**People-to-Systems**	**Systems-to-Systems**
Centralized (single authority)	Meeting	FERA Class 1: Bulletin boards and web meetings	Tight integration
Distributed (multiple authorities)	FERA Class 2: Personal interaction supported by collaborative software	FERA Class 4: Collaborative business process management	FERA Class 3: Publish and subscribe system-to-system

High-level collaboration service architectures exist that provide support for generic collaboration semantics through core collaboration business process services built on the collaboration services stack. For the pattern of collaborative business process management, Class 4, there is a need to bridge business processes with the core collaboration capabilities.

A significant lesson from this work is that business can be represented by business processes defined in terms of value chain reference models (e.g., VCOR), which can be used for process modeling, gap analysis, simulation, benchmarking, and consensus building. Secondly, an architectural representation can be used that maps process models to components of the conceptual architecture and resources used for accurate, fast, and flexible implementations of the process models in a federation. The two independent but reconciled process representations facilitate the mapping of business processes to core collaboration capabilities.

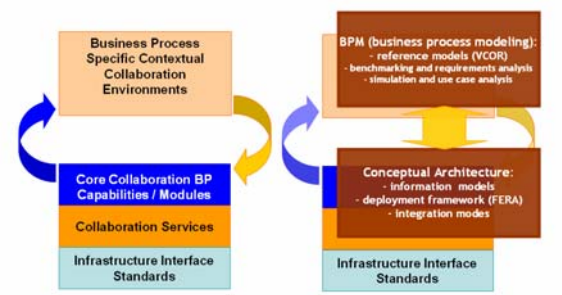

Figure 6: Integrated process and technology framework

Basic Components of FERA

FERA is a generic system architecture that describes all classes of deployment of the federated enterprise approach. FERA categorizes types of collaborative processes and patterns of technology deployment used to support those processes.

Being a common abstraction model representing a variety of technology deployment patterns of loosely coupled collaborative processes, FERA can be mapped into equivalent process models representing the same processes by using business semantics. Thus, in FERA an integrated process and technology framework can be established to speed up assessment of gaps, requirements, and deployment of the solutions used for supply chain collaboration.

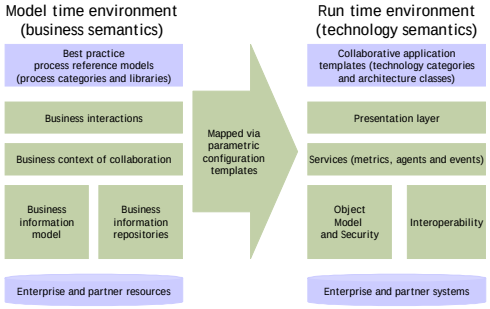

Figure 7: The usual representations of semantics

The two representations in Figure 7 are different views of the same process and need a mapping from the level of business semantics to that of technical infrastructure.

On the business semantics level, interactions between resources, their context of collaboration, information model, and repositories of information are all defined using business process modeling. This enables development of different models and their evaluation in a language that can speed up consensus building and provide for benchmarking and common representation of best practice reference models.

On the technical infrastructure level, the figure shows how technology components are used to deploy and support the processes defined in business process modeling semantics. A clear distinction must be made between the two sets of semantics. This distinction allows for an evaluation of the alternatives that can be used to support the same process, even if the deployment patterns need to be different in order to accommodate different levels of readiness of participants.

In general, FERA requires that participants can freely browse business models, evaluate them, and assess their own usage scenarios independently of each other, yet

understand the dependencies between the collaborating parties. After that, different participants can elect the deployment model that corresponds to the business process they decided to participate in, supporting it with their own capabilities and system-level readiness. [20]

There are seven basic components of FERA, and a specific configuration of these components can be used to support different classes of processes, where certain deployment patterns can support more processes and vice versa: one process can be supported by a specific combination of different deployment patterns.

Figure 8 presents the basic components of FERA and we explain each component in the list that follows.

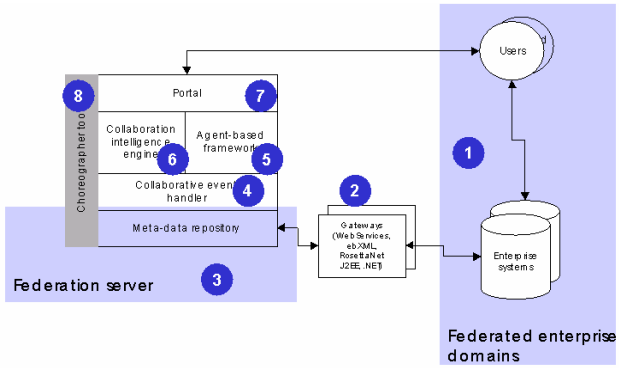

Figure 8: Basic FERA components

1. *Federated enterprise domains:* These are all systems that host the business logic, functions, and data for specialists and all users that need to collaborate.
2. *Gateways:* These components expose data contained in proprietary formats and local databases of the federated systems and users, and they establish open formats for data exchanges with federation servers. In FERA, gateways and federated systems have to conform to the set of standards commonly called SOA.
3. *Federation servers:* These open meta-schema servers publish and subscribe to and from gateways and portals. The servers support the persistence of collaborative sessions and preserve the context of the collaborative workflow. They also manage participants' profiles, and control access to data, security, and sign-on.
4. *Collaborative event handlers:* These rule-based engines control the start and finish and maintain the continuity of collaborative sessions, and they manage the status and change control of the collaborative workflow.
5. *Agent-based frameworks:* The component agents enable automation, pre-processing, and post-processing of data exchange steps in the collaborative workflow that is controlled by event handlers and maintained by federation servers.
6. *Collaboration intelligence engines*: These data collection, analysis, and reporting components provide insights into the effectiveness of collaborative processes to all participants.
7. *Portals*: These components enable Web-based user interfaces for participants to link directly to sessions on the federated servers.
8. *Choreographer tools*: These tools, used by federated instance administrators, are administrative tools that manage profiles, security, access, and sign-on protocols, and they link to federated system gateways and configurations of the meta-schema for the federated server. They are also used for choreography of business scenarios administered through contexts of collaboration and their version control.

Because of the autonomy of the enterprise domains, each needs an interoperability gateway. This gateway in FERA needs to be based on an open standard that will enable a reliable and scalable communications interface with the Federation Server in both directions. It should specify common functions that hide standard-specific terminology, constructs, and entities that can confuse end-users who do not need to know low-level standard specific details. Collaborative contexts are defined, preserved, and made persistent on one or more interconnected federation servers.

If the Federation Server comes with a registry-repository (reg/rep) component, the participants must register themselves in the Registry and Repository section, which governs security, access rights, and some other basic profile maintenance-related requirements.

The FERA built-in collaborative services–Collaborative Event Handler, Collaborative Metrics, Intelligence Engine, and an Agent Framework–enable collaborative process coordination and configuration. Agents, events, and metrics lookups are managed as built-in services. These three components are the heart of collaborative process choreography and enable complex collaborative patterns to take place in a non-deterministic business process design. Internal communication and collaboration between the Federation Server and the built-in services are conducted by Web services. This is a typical SOA.

Agents, metrics and events enable complex collaboration processes. Each FERA participant has its own created and configured agent. Agents interpret business process collaboration semantics that should be modeled and specified using standard business process modeling specification(s). The specification candidates for business collaborations (private and public) are ebXML Business Process Specification Schema (BPSS), Business Process Execution Language for Web Services (BPEL4WS), Business Process Modeling Language (BPML), etc. Although none of these has reached the level of sophistication required by FERA Class 4 process (discussed later), alternative approaches based on UML semantics have yielded sufficient results in many applications.

The Collaborative Intelligence Engine logs the progress of collaborations and collects the information used for detailed collaboration analysis (collaboration paths, overall system responses, system errors, etc. [21]).

The Portal enables user access to the Federation Server, contains user-defined views, filters and macros, enables invoking of services from a Web browser, connects other applications that are not announced in the federation into a coherent environment for data viewing and exchange (e-mail, Web meetings, etc.).

FERA Process Patterns

FERA describes a variety of collaborative solutions used in practice to support complex processes spanning organizational domains, business entities, and information technology systems that are based on three common principles: *autonomy, partiality and independence; open content and context semantics; and continuous operation.*

- *Autonomy, partiality, and independence.* Each participant in the collaborative process retains full autonomy of the internal workflow and ownership of her data. FERA does not require any change in local business logic or processes. FERA does not prescribe the collaborative process; the process is self-discoverable by all participants who can elect to join a subset of capabilities that fits their business needs and/or system readiness. Any participant can enter or exit the federation at any time; the processes supported by FERA will be able to continue.
- *Open content and context semantics*. To connect all participants and to maintain persistence of the collaborative workflow state, FERA requires that the federation and its working principles be defined using an open standard semantics that is equally accessible to all participants. A two-part description is required by FERA: context and content, and they both need to utilize an open standard semantics (e.g., for context BPEL, UMM, for content ISO 10303, PDX, etc.)
- *Continuous operation*. FERA does not need to move the data between participating systems in order to support the collaboration process. It only needs to enable information sharing through common definitions and/or references to the data changes supported by context and content definitions at the Federation Server, while preserving the context within which the information is being shared. FERA needs to support a variety of configurations of the same process to be continuously changed and executed within the same context thus following preferences, internal constraints, and capabilities of each participant.

These principles when translated to technology requirements mean that four basic requirements have to be met by all seven FERA components:

- *Service-based integration*. APIs between FERA components internally as well as communications with the external computing resources need to be supported by Web services. This enables a consistent method of searching, finding, locating, configuring, and executing all FERA APIs.
- *Dynamic configuration without coding*. FERA patterns of collaboration are generic, described using open standard semantics and therefore reusable and configurable to support a wide variety of processes on any FERA-compliant platform. FERA components are described using configurable templates that do not require coding, so a rule-based mapping between business process models (e.g., business modeling semantics like UML) to FERA deployment semantics can be achieved. FERA patterns can assist in classifying and declaring collaborative capabilities of different participants, but cannot prescribe the process in advance.
- *Reusability of components and patterns.* The context of collaboration needs to be defined using open standard semantics, while the ontology of contents needs to be consistent across all shared meta-data definitions. This enables the processes to be configured as they fit the business needs and reconfigured when the needs change. The process choreography itself evolves as the process is being executed by utilizing a combination of events, metrics, and agent services.
- *Meta-schema decoupled from business logic*. In FERA processes, context and content are modeled

separately and cannot constrain each other through direct relationship dependencies. This effectively means that relationships between entities defining the content of collaboration, as well as relationships between entities defining the context of collaboration, need to be declared in their respective meta-schemas as objects too.

SUMMARY

Closing the gap between process and technology resolves one major hurdle: the overwhelming complexity of the processes themselves in optimizing collaboration across the design and supply chain. Processes present far more intricacies than most companies recognize. All too often, IT organizations apply simplifying assumptions regarding requirements that inevitably lead to the wrong technical implementation, because ultimately, far broader ranges of interactions need to be supported. Either the rigidity only serves processes too trivial to matter, or else the lack of flexibility drives out the intelligence of effective decision making. Extreme care must be applied in considering the full spectrum of process interactions to understand their character completely before implementing technology, since that will determine the deployment framework.

This paper explored requirements of an SOE that executes within a managed SOA ecosystem, relying upon orchestration, business rules, and process correlation to provide monitoring and management, and business process analytics and autonomics. A framework was presented that weaves SOA into the enterprise fabric providing a rich service framework within which to implement enterprise business processes.

The SOA has great potential to enable more rapid deployment of service-composed processes but a gap may arise because technology does not speak a process language. Service descriptions do not communicate clearly to people regarding their roles, responsibilities, priorities, and milestones. Organizations tend to normalize processes and apply narrow definitions to activities involving broader requirements and deeper complexity. Process variations across the value chain break the systems that are based on those normalization efforts.

Value chain processes must be mapped to a well-defined service framework, based on common standards, which enable the sharing of content-rich information in a timely manner while maintaining an accurate business context. In order to establish faster and more intelligent collaboration throughout the value chain, a process framework must be explicitly defined in order to be integrated into the extended enterprise based on common objectives, on clearly establishing expectations and roles for all individuals, and on setting a reference baseline for continuous improvement.

By fully comprehending those collaborative processes within the value chain and the mapping to a managed SOA ecosystem comprising SOE, the enterprise is enabled to participate in a federation of enterprises, through the managed process execution framework, with a correlation of processes, process analytics, and process autonomics based on meaningful business semantics across a federation.

ACKNOWLEDGMENTS

We thank Vasco Drecun, partner and PLM research director at Collaborative Product Development Associates, LLC. Vasco is a strategic partner in the research on FERA and a contributor to that section of the paper.

ENDNOTES AND REFERENCES

[1] A secondary component of the investigation was to analyze the adequacy of BPEL as a language, where its well-understood limitations with regard to advanced synchronization scenarios were documented. BPEL constructs were developed to fill the gaps, but questions remained about both the language itself and the direction of the governing body, at least to the extent that BPEL is intended to support extended collaboration scenarios. On the vendor side, the inchoate state of thinking was confirmed through a series of meetings leading to the conclusion that further investigations remain necessary. Refer to Bartosz Kiepuszewski, "Expressiveness and Suitability of Languages for Control Flow Modelling in Workflows," *Ph.D. Dissertation* 2002, pp. 62-91. See also Petia Wohed, Wil M.P. van der Aalst, Marlon Dumas, Arthur H.M. ter Hofstede, "Pattern Based Analysis of BPEL4WS," *Technical Report* FIT-TR-2002-04, QUT. See also Petia Wohed, Wil M.P. van der Aalst, Marlon Dumas, Arthur H.M. ter Hofstede, "Pattern Based Analysis of BPML (and WSCI)," *FIT Technical Report*, FIT-TR-2002-05.

[2] See for example, "Envisioning the Service-Oriented Enterprise," *Zapthink*, 2003. See also, Sleeper, "Towards the Service-Oriented Enterprise," "service-oriented enterprise reflects a change in the human, business process, and organizational governance factors that shape how IT interacts with the business" at

http://www.webpronews.com/enterprise/enterpriseonline/wpn-13-20030903TowardstheServiceOrientedEnterprise.html*

An additional aspect of SOE is Service-Oriented Infrastructure [SOI], typically seen in terms of GRID

and the virtualization of the infrastructure layer. The relation between SOI and SOE is a topic of current research.

[3] With n services in the enterprise, there are n*(n-1) potential two-way interactions. Tracking the interactions through realistic business processes may produce graphs that are neither linear nor acyclic. Services without an execution framework quickly become unmanageable. This problem is made even more complex for dynamic processes where the process execution path is data dependent and not static.

[4] The process interoperability layer provides the bridge between the execution environment described in this section and the important business semantics and interaction models that are the focus of the second section of this paper. The two parts work hand in hand to provide a top-to-bottom blueprint for the SOE applicable to an enterprise and to a federation of enterprises.

[5] See for example, Gregor Hohpe and Bobby Woolf, *Enterprise Integration Patterns*, Addison-Wesley, Boston, MA, p. 163 (2004).

6 See Working Draft 01, 08 September 2004 at http://www.oasis-open.org/committees/download.php/9094/wsbpel-specification-draft-Sept-08-2004.html#s.Extensions.sect4*

[7] op cit §10, 10.1 and 10.2. See also §14.4 [Extensions–Correlation].

[8] The correlation identifier will be denoted C_ID.

[9] This provides a method to explore the run-time repositories of the individual P_k relative to a specific process instance $\mathcal{P}$.

[10] Suitably defined, the "value adornments" become the backbone of business activity monitoring within the framework.

[11] Even within a tree, a process controller P_i may call more than one successor.

[12] For example, if time stamp data are collected along the way, process QoS and SLA measurements can be taken, trended, and comprehended.

[13] In the second section of this paper, we explore a Federated Enterprise Reference Architecture in the context of an Integrated Process and Technology Framework. From this section's point of view the federation server, while possessing vital capabilities for managing business semantics and collaboration context, embodies another process controller, P_k. It can participate in correlation, based upon a unique process ID, at the moment of service invocation, resulting in a rich integrated analytics capability across the entire stack.

[14] The general benefits of a business rules engine *vis-a vis* traditional application-level coded logic are that business rules (1) change quickly; (2) promote agility; (3) enable the business people to control/change business processes; (4) reduce tight coupling between business behavior and code base; (5) leave the released code base stable yet enable dynamic environment; and (6) as proved by the research here, can form the basis of SPC and reflective programming control over the execution environment.

[15] An interesting example is created by use of automated data collection, such as RFID, feeding an operational or analytics environment that, in turn, sets environmental facts in the rules base via events. Transactional or aggregated information can then drive logistics processes in real or near real time based upon a full set of business rules and their associated inferential patterns.

[16] The critical success indicators of the project related to the business rules engine itself were as follows: modify business behavior without modifying the underlying code base; move business domain logic out of code to minimize impact to the release to production process; drive agility by giving business owners control over business processes; support globalization by explicitly partitioning business rules from code; drive code reuse by reducing dependence on domain specific logic in code; make business rules visible and consistently documented; and provide a framework for versioning and control of business rules.

[17] In the TD effort, higher-level process models describing collaboration scenarios were exported into BPML and then translated by hand into BPEL for execution. The process worked but an automated solution assuring semantic integrity is needed.

[18] FERA was introduced by Collaborative Product Development Associates, LLC, through sponsored research on product lifecycle management and assessing collaboration processes for product design for supply chains developed with enterprise reference models. FERA represents a benchmark for establishing the level of maturity in supporting collaboration within a federation.

[19] The Value Chain Operations Reference (VCOR) model is being developed by the Value Chain Group, an independent consortium dedicated to the consistent

integration of constituent frameworks to form a common language for value chain optimization.

[20] The first part of this article, "Evolving SOA to SOE," describes the instantiation of a rich service framework to support FERA within an enterprise domain.

[21] Supporting unique process IDs at the level of FERA through the execution environment, such as that described in the first section of this paper, would enable correlation between enterprise systems and provide extended analytics across the entire federation.

AUTHORS' BIOGRAPHIES

George W. Brown joined Intel in 1994 and is currently a senior program manager within the ISTG Research group. He focues specifically on methods and tools to ensure Intel reaches its goals in supply chain management by identifying opportunities to apply information technology in innovative ways to solve business problems and improve Intel business processes. He is also the past chairman of the Supply Chain Council and has represented Intel in external research and benchmarking activities as chair of the SCC Research Stategy Committee. His e-mail is george.w.brown at intel.com.

Robert E. Carpenter joined Intel in 1997 after a 20-year career as an attorney, retiring as district attorney of Greene County, NY in 1996. For many years, he taught mathematics at Bard College and later law at Albany Law School. After joining Intel, he focused on the development of architectural frameworks and most recently has been responsible for the definition of process automation architecture across the business landscape. His current interests include process algebras and the application of formal representation theory to virtualized process and infrastructure. His e-mail is robert.e.carpenter at intel.com.

THIS PAGE INTENTIONALLY LEFT BLANK

Bayes Network "Smart" Diagnostics

John Mark Agosta, Corporate Technology Group, Intel Corporation
Thomas Gardos, Information Services and Technology Group, Intel Corporation

Index words: diagnostics, troubleshooting, fault analysis, Bayes networks, knowledge engineering

ABSTRACT

Formal diagnostic methods are emerging from the machine-learning research community and beginning to find application in Intel. In this paper we give an overview of these methods and the potential they show for improving diagnostic procedures in operational environments. We present an historical overview of Bayes networks and discuss how they can be applied to diagnosis. We then give an illustration of how they can model the faults in a vacuum subsystem of a manufacturing tool.

INTRODUCTION

Substantial portions of Information Technology (IT) and manufacturing operations budgets are dedicated to diagnostics. In most enterprises today diagnostic methods are generally ad-hoc and rely on undocumented knowledge of a few experts. These practices commonly result in excessive down time of critical infrastructures and in wasted expense associated with unnecessary tests and occasionally unnecessary repairs. This situation will continue as manufacturing and IT operations become more automated, thereby reducing the need for human operators, while at the same time increasing the amount of equipment critical to keeping the business running. For example, in the next generation of semiconductor fabrication plants, most of the wafer handling will be automated, leaving maintenance as the predominant operations cost. The problem is complicated by the fact that the most highly skilled diagnostic staff are usually the most sought after for new deployment projects, taxing the availability of diagnostic expertise for on-going operations.

Bayes Networks

As it turns out, rigorous methods are emerging from the machine-learning research community that begin to address these problems. One of the most promising is Bayes networks, a formalism based on probability and graph theory. (Bayes networks are also known as "belief networks," "Bayesian networks," or "causal networks.") Formally, a Bayes network models the probability distribution of a set of random variables as nodes in a graph, and it models the probabilistic dependencies among variables as arcs in the graph. Causal relationships are represented by conditional probabilities, encoded in the distributions associated with each node, to capture the strength of the relationship. Such directed graphs visualize the probabilistic relationships among variables. Most important is that Bayes networks model independence among variables; the network contains only a small fraction of the possible dependencies and this makes working with them manageable. Just as the network graph makes the model design easy to grasp, it also makes possible efficient computation of probability updates.

Applied to diagnostic modeling, the network's random variables represent events, divided roughly into two classes: evidence (e.g., tests or symptoms) that can be directly observed, and root causes (e.g., component failures) that cannot be observed and must be inferred from evidence. We treat diagnostic models as causal Bayes networks. Cause-effect relationships beginning with root problems and the tests or symptoms that result are both elicited from experts and learned statistically with support of operational data.

Once the model is developed, diagnosis proceeds by inverting the conditional dependence relationships applying a generalization of Bayes rule, and thus calculating the probability of root causes conditioned on evidence provided by the user. The result is a ranked list of most likely root causes based on the evidence provided so far and a second ranked list indicating the most discriminating diagnostic tests to perform next. This methodology shows some very interesting and valuable properties:

- The ability to diagnose multiple simultaneous root causes, the combination of which may have never been anticipated by the experts.
- The dynamic creation of the diagnostic sequence– effectively, the flow chart.

- The ability to submit and retract evidence and recompute its effect in any order during the diagnostic process.

Bayes Networks as a Knowledge Engineering Tool

Another valuable result of this process is that the expert knowledge is captured and thus the diagnostic model forms the basis of a knowledge engineering tool. Moreover, the expert knowledge is encoded in a manner such that diagnostic inference can then be computed–a capability that most knowledge management approaches cannot deliver.

The information that comprises the Bayes network exists in an informal and unstructured way in equipment maintenance logs, but most of it never gets applied to subsequent occurrences. The Bayes network is a concise way to encode diagnostic knowledge from a combination of engineering knowledge and statistical data in such a way that all possible test and observation diagnostic sequences can be generated from it. By their nature, breakdown events are rare, and a troubleshooting model cannot be built solely by a data-driven approach.

Bayes network diagnostics are used in several commercial implementations, most notably by Microsoft, in General Electric's Condition Forecaster* tool and for troubleshooting websites (see, for example, "Parts America.com" [19]).

Intel has sponsored an annual Bayesian Application Workshop in cooperation with the Uncertainty in Artificial Intelligence conference where commercial implementation projects are presented [4].

PROBABILISTIC LOGIC AND BAYES NETWORKS

Why is probability the right way to reason about diagnosis? The reasons may appear obvious to the reader, but the topic has been the source of much discussion, both among students of logic and of statistics. This section offers a short answer to the question. A reader uninterested in the question may want to skip directly to the example in the next section.

There is a rich history of theory that Bayes network diagnostics draws upon, as the name "Bayes" suggests. Thomas Bayes was an 18th-century cleric whose name is associated with a comprehensive concept of probability, i.e., that any uncertainty can be represented as such. The Bayesian approach weaves together statistics, economics, and machine learning. It is the starting point from which the theory of rational decision-making has been developed, and the one on which this approach to diagnostics is based. A rational solution incorporates all the decision maker knows, to the point where the decision can stand in for the judgment of the decision maker. This is a tall order, but it offers a first principles approach to solving a problem. A Bayesian solution presents the full rationale for the decision, which among other things, gives a justification for automating the decision.

Bayes network diagnostics applies a decision-theoretic concept of probability as the basis to measure information. The cost of achieving a successful diagnostic outcome is a function of the time and cost spent collecting information, and this is a significant part of modeling diagnosis. The "value of information" (VOI) computation discussed in the section "The Value of Diagnostic Observations" minimizes the diagnostic steps to isolate faults. This is a good approximation to minimizing the cost to achieve the correct diagnosis, and this can easily be improved upon by assigning different costs to performing tests. The full decision optimization problem can be addressed by embedding the diagnostic Bayes network in an optimization framework, but this is not typically justified [10]. Interestingly, see Bayer-Zubek [5] for a novel Bayesian approach that does not take advantage of a Bayes network, but does compute an optimal diagnostic policy.

Cox's Theorem as the Foundation for Probabilistic Logic

Probability is a means to express and reason with degrees of uncertainty precisely. This section outlines a derivation for probability, and the rules for updating probabilities, that is more widely applicable than conventional statistics-based derivations, but maintains consistency with them. The conclusions presented here are based on an argument first presented by R.T.Cox [7, 13, 18] He begins with three premises:

I. **The uncertainty of an event is described by a real number**. It is possible to find an equivalence between any real, uncertain value and a *certain equivalent*. For instance, you may offer a price for a used car before having the opportunity to inspect it. Your offered price weighs the possible different conditions of the car that are uncertain; but your offered price is a certain quantity, called the *expected value* of the uncertain quantity, the value of the car. A *probability* is just the expected value of an event outcome, where the certain occurrence of

* Other brands and names are the property of their respective owners.

the event has value *1* and its negation, the absence of the event, has value *0.* In most expositions the concept of expectation is derived from probability. In this case, similarly to Whittle [24] the reverse is more natural.

II. **Probabilities combine in common sense ways.** When events are certain, probabilistic logic is reduced to familiar Aristotelian logic. When probabilities can be defined by frequencies of event occurrences, probabilistic logic is consistent with counting frequencies.

If events are not certain, then to combine them and obtain their joint probability, we need to keep track of the *state of information* of each event, say *A* and *B*, in order to combine them. This is represented by another event or proposition, *C*, called the *conditioning* event. The probability of the *conditioned* event (e.g., *A, B*) is supported by *C.* Conditioning is shown by a vertical bar "|", hence "*A* given *C*" is written *(A | C).* Unlike the Boolean algebra of certain propositions, the probabilities of uncertain propositions must have consistent conditioning to be combined, and the outcome of their combination will depend strongly on their conditioning. An essential difference when working with uncertainty is that uncertain inferences are "non-monotonic" in the inclusion of changing conditioning by additional evidence; additional evidence may weaken an inference, whereas once a logical proposition is proven, no additional consistent fact will change the conclusion.

III. **Combining the same knowledge in different orders gives equivalent results.** The largest part of reasoning with probabilities is reasoning about how new information–a new fact or assumption–changes our belief in a proposition, as represented by a probability. This is called *belief update.* Solve this, and the method for reasoning with probability is completed.

The specific problem of belief update can be stated as deriving the combining function, *f(),* for the probabilities *A* and *B,* where one is conditioned on the other, and they share a common state of information *C*:

$$pr(A \text{ and } B \mid C) = f(pr(A \mid B \text{ and } C), pr(B \mid C))$$

To make a long story short, the derived combining rule, *f(),* not surprisingly turns out to be multiplication. Perhaps what is surprising is that Cox's theorem derives the conventional probability rules without recourse to making assumptions about sample spaces, event frequencies, set theory, additivity of the probability of mutually exclusive events, and the definition of conditional probability measures. **Thus we have a general result about reasoning under uncertainty that is consistent with, and subsumes conventional notions of statistical probability.** From this "information update" combining rule it is an algebraic derivation to obtain the update rule known as the Bayes rule:

$$pr(A \mid B \text{ and } C) \propto pr (B \mid A \text{ and } C) \, pr(A \mid C)$$

Reasoning About Causes

The notion of cause is closely entwined with probability. A cause may function unreliably, but the cause itself–as a general principle–is not at question. For example, we don't question the tendency of plumbing to wear out and leak with time, but that is not saying that any specific valves or pumps will do so. The inverse question, of reasoning from specific examples to establishing the general cause is the problem of inference. Statistical inference is applied to prove the existence of a cause in principle. In comparison, in working on diagnostic problems our models will select from a set of general causes which specific cause or causes can explain a breakdown.

Why not use an entirely qualitative (e.g., symbolic) calculus? Because there aren't any that don't reduce to computing with probability numbers [14], and because there is a bonus in using probability-based logic–it can be extended to statistical techniques. This has a direct benefit when data are available from which to learn Bayes network probabilities. Bayesian statistics consider the proper way to combine expert judgment and statistical knowledge. For a comprehensive review see especially the article by Heckerman in [16].

The models we develop work fine with the rank-ordered probability numbers estimated by the technicians. The arguments why Bayes net methods work well with order of magnitude probabilities and other methods don't are discussed extensively in the literature [21].

History of Bayes Network Diagnostics

Early work in perceptrons, precursors to neural networks, was sympathetic to Bayesian approaches. Progress in Bayesian approaches to automated reasoning along with other quantitative reasoning methods stagnated when purely symbolic approaches became popular in the early 1980s. The current blossoming of Bayesian methods coincides with Pearl's introduction of "Belief networks" [20] and Heckerman and others work on the Pathfinder project [9], as an improvement on the Mycin medical diagnostic rule-based expert system.

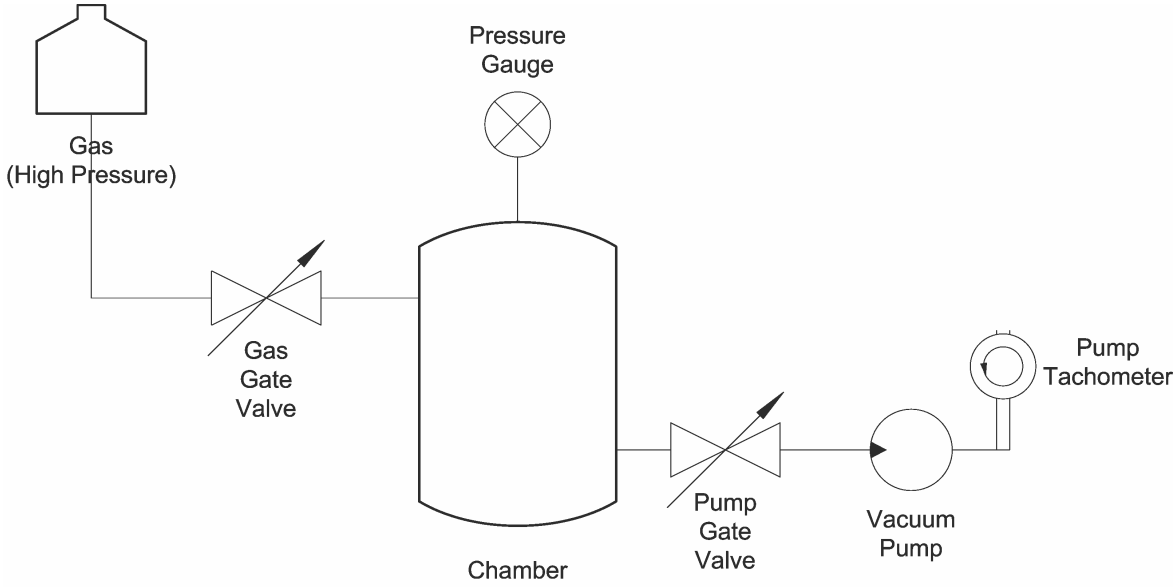

Figure 1: Chamber, valve, and pump assembly for diagnostic example

Several members from this group were early members of Microsoft Research, and they implemented a wide variety of applications for computer and software configuration diagnostics. About the same time, the field was spurred on by general-purpose Bayes networks solution methods by "join-tree" propagation, published by Lauritzen and Spiegelhalter [18]. There is also a significant Bayes network component in machine-learning research that has contributed to the growth of the field.

AN EXAMPLE DIAGNOSTIC PROBLEM

Consider the following system shown in Figure 1 that is loosely based on a chemical-vapor deposition process in silicon wafer manufacturing [1][2]. In this simplified system the chamber is pumped empty of all the reacted gas so that it is clean and ready for the next process to take place. The gas gate valve is closed to maintain pressure. The pump gate valve opens and the vacuum pump removes the reacted gas and pumps down the chamber to a low vacuum pressure of about $1/100^{th}$ of an atmosphere[1]. A gauge indicates the pressure inside the chamber and a tachometer indicates the rotational speed of the vacuum pump.

For this example, there are five components whose failure we may have to infer in a diagnostic process. The five components along with their operational states are as follows:

- Gas gate valve (good, leaking)
- Pump gate valve (good, leaking)
- Vacuum pump (good, insufficient vacuum)
- Pump tachometer (good, out of calibration, no reading)
- Pressure gauge (good, out of calibration, no reading)

Each component will be represented by a node in the Bayes net diagnostic model.

Assume there is a chamber pressure alarm that will sound if the chamber pressure strays outside of specified limits in the clean cycle.

Next, we characterize how a component failure may manifest itself. This is knowledge that an expert with the tool would possess.

- A failure of the gas gate valve would prohibit the chamber pressure from reaching the cleaning pressure level and would cause the chamber pressure to rise quickly after being pumped down.
- A failure of the pump gate valve would cause the chamber pressure to rise quickly after being pumped down, although the chamber would initially reach the target clean cycle pressure.
- If the pressure gauge failed or went out of calibration, it would fail to indicate that the pressure in the chamber reached its target and would not indicate accurate pressure levels for any diagnostic

[1] An atmosphere is a unit of pressure equal to the pressure of the air at sea level.

tests. This could confound the chamber clean pressure alarm and the rate-of-rise test.

- If the vacuum pump didn't operate correctly, the chamber pressure would not reach its target pressure level for the clean cycle.
- If the pump tachometer failed or went out of calibration, it would not give an accurate reading of the pump speed which would confound the vacuum pump speed reading.

The first four component failures would manifest themselves as a chamber clean pressure alarm. The last component failure could confound a diagnostic procedure but would not trigger a pressure alarm.

The following are the diagnostic procedures available to the technician. The implications of each procedure correspond to an arc from a failure node to a diagnostic test:

- *Rate-of-rise test*. This test measures whether the chamber can pump down to target pressure and measures how quickly the pressure rises with both gate valves closed. There are three outcomes: normal; fail-to-pump-down, and failed-rise. Fail-to-pump-down implicates a fault in the vacuum pump. Failed-rise implicates a fault in either of the gate valves.
- *Gas gate valve physical leak check*. This checks to see if there is a physical leak by using a handheld leak sensor. Possible outcomes are pass or fail. This test implicates a fault in the gas gate valve.
- *Pressure gauge check*. The pressure gauge is removed and placed on a test card and tested for operation and calibration. Possible outcomes are normal, out-of-calibration, and bad. This test implicates a fault in the pressure gauge.
- *Vacuum pump speed check*. The vacuum pump is set to a predetermined speed that is checked via the pump tachometer. Possible outcomes are pass and fail. This test implicates a fault in the pump or the tachometer.
- *Pump assembly leak test*. The pump and pump gate valves are checked for leaks. Possible outcomes are leak and no-leak. This test implicates a fault in the pump gate valve or pump gaskets.

Diagnostic Problem

Now suppose that an alarm sounds indicating that the chamber pressure did not reach the target vacuum pressure. The chamber pressure alarm is the primary indication that something is wrong, and a diagnostic procedure must commence. The technician must then decide the sequence of the diagnostic and repair steps, in order to get the system operational as soon as possible.

There are 15 combinations of failure for the first four components for each of which the pump tachometer may or may not fail, resulting in thirty total component failure combinations.

Once a breakdown has been identified we need to sequence the steps to isolate equipment faults; this needs to be done in a principled way to eliminate the typical uninformed "replace and test" maintenance procedures.

BAYES NETWORKS APPLIED TO DIAGNOSIS

The Diagnostic Process

As shown by the example, diagnosis starts with a general indication that something has failed–the *primary indication*–and narrows down the cause of the indication by a series of observation and test steps. The diagnostic process ends by finding one or more root causes that explain the failure.

Diagnosis as a Bayes Network

In diagnostic models the nodes fall into two sets: root causes or faults, and observations or tests. In the course of diagnosis a (hopefully small) subset of the observation variables will become certain. The subset of likely faults will be inferred from these. Faults won't be certain, except perhaps when a faulty component is replaced and its condition is observed directly.

Fault states predict observations. From observation states we can infer faults: we can see this with a two-node Bayes net (Figure 2). The arrow from fault node to test node respects the causal direction from fault to observation.

A Two Node Diagnostic Model

The simplest example of a diagnostic model has a fault variable that conditions a test variable (Figure 2).

Figure 2 Simple two-node Bayes net[2]

[2] All Bayes net visualizations in this document were created using the GeNIe software developed at the Decisions Systems Laboratory at the University of Pittsburgh and is freely available at http://www.sis.pitt.edu/~genie

Table 1: Example prior and conditional probability tables for a fault node and test node, respectively

fault	
broken	0.1
working	0.9

test	broken	working
pr(abnormal | fault) =	0.99	0.1
pr(normal | fault) =	0.01	0.9

Table 1 shows the probability assumptions for the two nodes. The fault node failure probability is 0.1 prior to the test. The probability that the test indicates abnormal given the fault node is broken is 0.99 and is referred to as the test sensitivity. The probability that the test indicates normal given the fault node is working is 0.9 and is called the specificity.

With the data in Table 1 we can calculate the marginal probability that the test result is abnormal by adding the two weighted conditional states:

$$pr(abnormal)$$

$$= pr(abnormal \mid fault)pr(fault)$$

$$+ pr(abnormal \mid working)pr(working)$$

$$= (.99)(.1) + (.1)(.9) = 0.189$$

It is largely because of the poor specificity that the test is predicted to be abnormal with probability 0.189. Figure 3 shows the nodes with probabilities indicated.

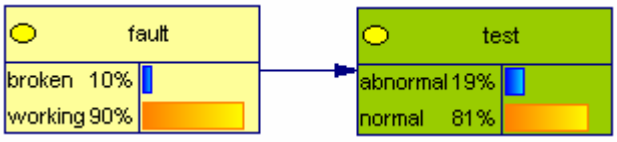

Figure 3: A two-node Bayes net

When evidence for the test is applied, the test value becomes certain, and the updated probabilities of other nodes (just one in this case) are conditioned on the evidence. Applying Bayes rule to this obtains the probability of the fault given the test, in this case for an abnormal test result (Figure 4).

Given this evidence, the probability that the fault node is broken (0.52) is now slightly greater than the probability it is working (0.48).

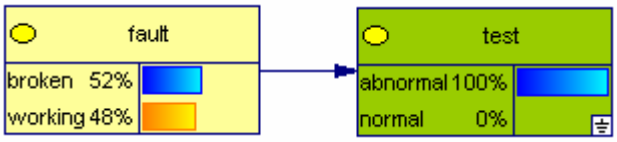

Figure 4: Inference in two-node Bayes net

Vacuum Chamber Diagnostic Model

We can now model the faults and the tests of our pressure chamber example as in Figure 5. We have color coded the nodes for easier interpretation. The five yellow nodes forming a column on the left side represent the components whose failures can be inferred. The blue node by itself in the center represents an internal unobservable state, in this case the chamber pressure. The green nodes on the right of the graph represent the diagnostic observations and tests that an operator could perform.

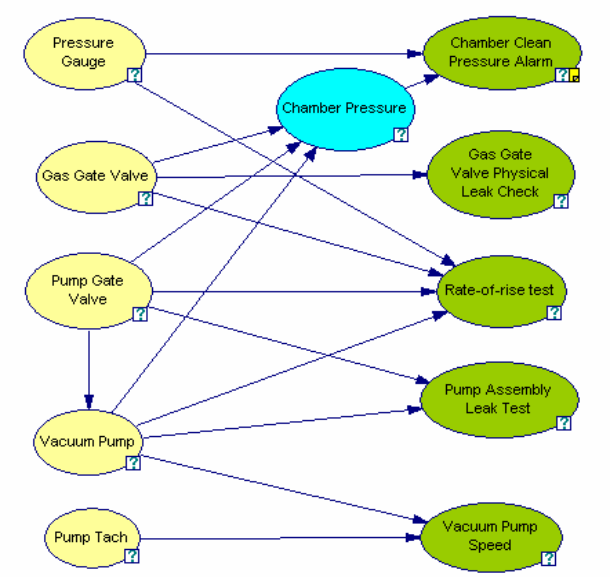

Figure 5: Bayes net diagnostic model of pressure chamber example

Figure 6 shows the same graph but with bar charts representing the marginal probabilities of the model, computed with no evidence set.

All but one of the fault nodes along the left side of the graph have no arrows entering from so-called parent nodes. The probability assignment for these nodes represents the prior probability of failure given nothing is known about the results of tests. Their values reflect the components' reliabilities. So the pressure gauge is the least likely to fail with a 1% chance, and the pump gate valve is the most likely to fail with a 6% chance.

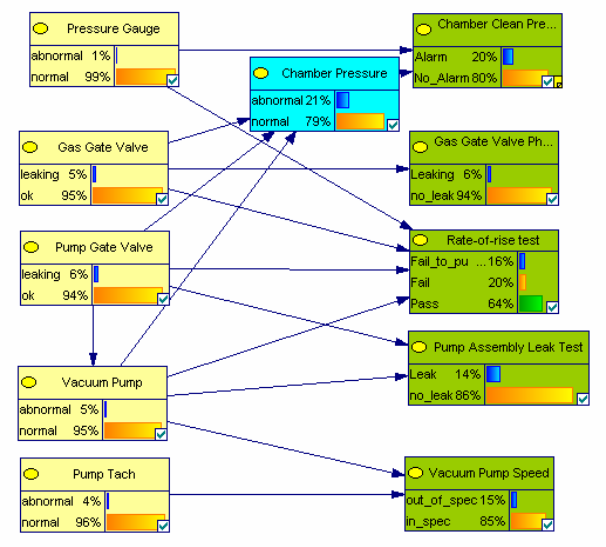

Figure 6: Example Bayes net showing marginal probabilities

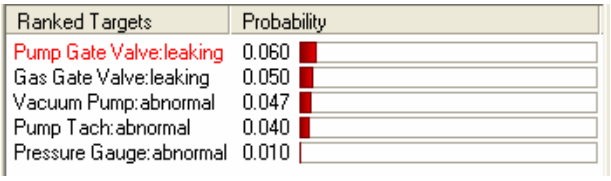

Figure 7: Ranking of possible component failures given no diagnostic test information is known

The vacuum pump node has the pump gate valve as a parent node reflecting the fact that if the pump gate valve is leaking it can make the pump more likely to fail.

Since the vacuum pump has a parent node, a conditional probability table is defined similar to the one used in Table 1. In this example, we use the values shown in Table 2.

Table 2: Conditional probability table for the vacuum pump node. The top row refers to the states of the pump gate valve which is a parent node. The abnormal and normal states refer to the vacuum pump.

Pump Gate Valve	leaking	ok
pr(abnormal | Pump Gate Valve) =	0.15	0.04
pr(normal | Pump Gate Valve) =	0.85	.96

With these data and the prior probabilities for the pump gate valve, we can calculate the marginal probability that the vacuum pump is normal by adding the two weighted conditional states as we did in the two node example to get

$$pr(normal) = 0.9534.$$

Since the abnormal and normal states are mutually exclusive and exhaustive, the probability of the vacuum pump being abnormal is simply the complement:

$$p(abnormal) = 1 - p(normal) = 0.0466.$$

We now have the probabilities of failure for all the components before observing any diagnostic tests. The ranked faults can then be displayed as a Pareto chart as shown in Figure 7. In the next section we look at how these probabilities change as we provide diagnostic evidence and update the model.

First we look at another challenge of modeling the local conditional probability for a node. Consider the chamber pressure node in Figure 5. Its possible states are normal and abnormal. Because each of its parents can take one of two states, there are eight combinations of parent states and thus eight columns in the conditional probability table that would need to be assigned (Figure 8). In general, the probability model scales exponentially with the number of variables. Most people find this level of specification unintuitive. Fortunately, if we make the reasonable assumption that the parent nodes are *causally independent* [10] given the current node, we can employ a local model, such as the Noisy MAX, that scales linearly with the number of parent variables. In short, causal independence assumes a combining rule so that it is sufficient to consider the effect separately of each parent on the child node [20].

Using a Noisy MAX model, the specification of conditional probabilities for our example is greatly simplified. Figure 9 shows the same node as the Noisy MAX model. The LEAK column on the right-hand side allows for the probability, usually small, that the node can take the abnormal state by some other means not represented in the model.

Hard Diagnostic Problems

Diagnosis can be hard for several reasons. The pressure chamber model has most of these characteristics:

- First, the primary indication may tell little or nothing about the source of the problem. An "idiot light" is an example. In the model, the chamber pressure alarm could be the consequence of four out of five faults.

Gas Gate Valve	leaking				ok			
Pump Gate Valve	leaking		ok		leaking		ok	
Vacuum Pump	abnormal	normal	abnormal	normal	abnormal	normal	abnormal	normal
abnormal	0.99685	0.991	0.9685	0.91	0.9685	0.91	0.685	0.1
normal	0.00315	0.009	0.0315	0.09	0.0315	0.09	0.315	0.9

Figure 8: Conditional probability table (CPT) for a node with three parents

Parent	Gas Gate Valve		Pump Gate Valve		Vacuum Pump		LEAK
State	leaking	ok	leaking	ok	abnormal	normal	
abnormal	0.9	0	0.9	0	0.65	0	0.1
normal	0.1	1	0.1	1	0.35	1	0.9

Figure 9: Using the Noisy MAX local probability model the conditional probabilities can be simplified for this case of a node with three parents

- The things observed are not the things that go wrong. We use the terms *observations*, *symptoms,* and *tests* for the things that can be observed, and *causes* or *faults (*or *diseases* in clinical diagnosis) for the things the diagnosis attempts to identify. Causes are *inferred* from observations. Occasionally faults can be observed directly. In this case the problem is reduced to searching efficiently and inference is not necessary. The model in the chamber pressure example is an instance where none of the failures can be observed directly.
- One root cause may cause many observations, and one observation may be non-specific. If there is a specific, accurate test for each failure and no opportunity for confusion among them (i.e., silver bullet tests), then again, diagnosis is reduced to search. In this model the "gas gate valve physical leak check" is specific to the "gas gate valve" leaking. Otherwise all faults must be isolated by a combination of observations and tests.
- Causes may cascade, so the problem is compounded by having to find the first cause that started the chain. Note how the arc from "pump gate valve leak" to "vacuum pump" models how over time the former can cause the latter to degrade.
- Tests and observations may either interfere with, or improve another test's diagnostic value. A destructive test is an extreme example of a test that interferes with others. Tests that are only valuable in combination improve each other's value. Interactions among test results can be modeled by arcs between tests.
- There may be numerous, possibly expensive tests that relate to numerous faults, so tests must be chosen efficiently to make progress with the diagnosis. While the size of the model is linear in the number of tests, the number of test sequences is exponential in the number of tests.
- The diagnostic problem may not be completely resolvable with the available tests, time, and money. In this situation narrowing the list of suspect faults is the best that can be done. Probability updates have the desirable property of providing useful guidance even if a complete resolution of the diagnosis is not possible.

These last three characteristics are not illustrated in this small model. However, as we see, each of these characteristics can be managed to advantage with the available modeling techniques.

In the next sections we talk more about computing inference with this diagnostic model.

Efficiently Isolating Faults by Ranking of Root Causes

As we discussed, the value of a diagnostic tool depends on how rapidly it isolates the correct fault, since even simple approaches, let alone other automated methods for diagnosis, will eventually come to a resolution of the problem. In the worst case a machinery problem can be approached by brute force replacement of all components. Clearly this is excessive in time and cost. The problem is to minimize the collection of evidence. In the best circumstances, observation and test selection will be parsimonious, and most observations will not be made.

A useful diagnostic model comprises a large number, typically hundreds, of possible observations, alarms, measurements, or tests that can be performed. These may entail costs, take time, or require materials, equipment, or skilled labor. The common thread is that it is

uneconomical to seek all inputs. Therefore, there must be a tradeoff between the "diagnostic" value gained as progress toward isolating system faults and the cost expended in finding the faults We use the term "cost" to include any resource expended to generate an input to the model. In the diagnostic procedures that we consider, time will be "of the essence" and cost will be measured as a function of time to find the fault.

During diagnosis the Bayes network updates the probability of each fault based on the values of observations made to that point. Typically this is displayed as a ranking of the fault variable probabilities as shown in Figure 6. Isolating the one or more faults at the root of the problem consists of driving the probability of the faults to extremes–the probability of true faults to one and the rest to zero. An appropriate way to measure this is to compute the entropy over the set of faults [3]. Thus the most efficient set of observations is the set that maximizes the decrease in entropy of the fault set. Since the Bayes network can be used to compute the full joint probability distribution of the faults and observations under any state of information, it can be used to compute these entropy changes. The computation of these values to determine the best set of observations to make is loosely referred to as "value of information." Strictly, value of information refers to the increase in expected value of the outcomes under the optimal decision were the information value available when the decision was made [12].

In our example, if we set diagnostic evidence so that the "rate-of-rise" test is set to "fail to pump down" we get the recalculated ranked probabilities shown in Figure 10. Note that the vacuum pump probability went from 0.047 with no evidence set to 0.278 with this diagnostic test.

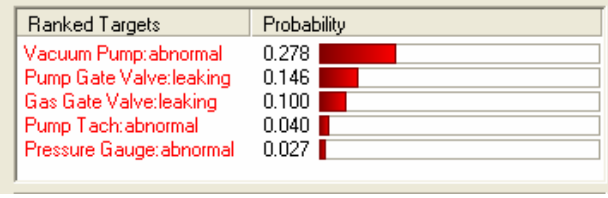

Figure 10: Ranked probabilities of component failure with rate-of-rise test set to "fail to pump down" state

If we further set a second diagnostic test, vacuum pump speed, to its "out of spec" state, we see that the vacuum pump is implicated further, increasing the probability to 0.779 (Figure 11). Note that the pump tachometer probability of failure went from 0.04 to 0.112 with the evidence of pump abnormality, representing the fact that the pump diagnosis could be confounded by a bad tachometer. The probability of the tachometer being abnormal could have been even higher except for the fact that the chamber pressure failed to pump down to vacuum during the rate-of-rise test making a pump abnormality more feasible. We would see this more clearly if we were to clear the rate-of-rise test result and only have evidence of the vacuum pump being out of spec. In that case the pump tachometer probability would jump to 0.265.

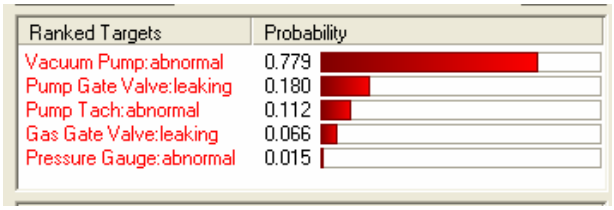

Figure 11: Ranked probabilities of component failure with both rate-of-rise test set to "fail to pump down" and vacuum pump speed set to "out of spec"

Value of Diagnostic Observations: Finding the Best Test Sequence

The challenge in diagnostic reasoning is computing which is the best observation to select next, to make progress towards finding the faults [23]. This is called *differential diagnosis*, at the stage in the diagnosis when the current information suggests several candidate faults with a high probability, to be confirmed or refuted to find what the actual faults are. As explained, a decrease in the entropy of this set of faults is the heuristic to measure progress toward confirming the faults.

The difference between the fault set entropy conditional on a set of observations and the entropy lacking the observations is called the "mutual information" between the fault and observation sets. Mutual information captures roughly the dependence between the fault and observation sets; the more positive the mutual information, the stronger the dependence. If the probability distributions of faults and observations are independent, then their mutual information is zero. Obviously such observations would be useless.

Finding the best set of observations to distinguish among the current fault candidates by mutual information becomes a hard computational problem. Each set requires computing over the Bayes network, which typically is expensive. Furthermore, mutual information over a set of observations may not be well approximated by the mutual information of the individual observations; for example, an observation that interferes with another, or is needed in combination with another, will affect the other's mutual information. In the interest of computational efficiency, in practice, these dependencies may be disregarded [3].

Finally, mutual information of any set of observations will change as other observations are made. This may be

due to dependencies among observations. Also the current fault probabilities play a role: faults with extreme values are less likely to change and hence they affect mutual information less. Thus it is insufficient to rely on mutual information computed at the start of the diagnosis. The observation selection task in diagnosis may be thought of as a dynamic feature selection problem, conditional on the current observation values.

Since diagnostic value as indicated by mutual information for most of these inputs depends heavily on the current state of the diagnosis, it is worthwhile to re-compute mutual information after each observation. The procedure is thus to re-compute each observation's diagnostic value individually after each observation is taken, and the state of belief is updated, then to select the next observation with the highest value. This sequence of steps is repeated until either the probability of one or more faults approaches one, or the value of all remaining observations approaches zero. This approximate guided search method is called *myopic* test selection. Simulations in typical diagnostic models have shown that myopic test selection closely approximates optimal [11]. The mutual information computation over all possible observations as required by myopic test selection may be onerous, and further approximations may be necessary [13].

Continuing with our example, Figure 12 shows the diagnostic value of each of the tests in the chamber pressure model. Note that the rate-of-rise test shows the highest diagnostic value. If we set evidence for that test to "fail to pump down" as we did earlier in our example, that test is removed from the list. The re-ranked list is shown in Figure 13. As one would expect, the failure of the chamber to pump down implicates the vacuum pump and so the vacuum pump speed test is now highest on the list.

Ranked Observations	Diagnostic Value
Rate-of-rise test	0.286
Pump Assembly Leak Test	0.260
Vacuum Pump Speed	0.229
Chamber Clean Pressure Alarm	0.228
Gas Gate Valve Physical Leak Check	0.181

Figure 12: Diagnostic value of information for chamber pressure example before any diagnostic evidence is set

Ranked Observations	Diagnostic Value
Vacuum Pump Speed	0.240
Pump Assembly Leak Test	0.228
Chamber Clean Pressure Alarm	0.144
Gas Gate Valve Physical Leak Check	0.127

Figure 13: Diagnostic value of information for chamber pressure example after rate of rise test is set to fail to pump down

ON DEPLOYING A BAYES NETWORK-BASED DIAGNOSTIC MODEL

In our work we are modeling a subsystem of a chemical vapor deposition tool. We have been working with the domain experts, in this case the tool engineer and lead technician, to elicit the Bayes net model for that subsystem. Our network consists of 65 nodes and 81 arcs. Of those, 23 represent components likely to fail, 12 represent internal unobservable states, and 25 represent diagnostic tests. We also have three nodes that we call conditioning nodes. They are not probabilistic but rather represent some known state of the system such as the number of hours of operation of the tool greater or less than 10,000 hours.

Our experience has shown that the cause-effect relationships modeled by the Bayes net graph were quickly picked up by the domain experts. In fact, after an introductory session, the equipment engineer quickly sketched out a thirty node graph. Refining the graph proceeded more slowly, however. We calculate that we averaged about three nodes per hour during subsequent sessions. During these sessions, we not only extended the topology but also defined the conditional probabilities. Relating the probabilities to the graph was not quite as intuitive an exercise, but once we started evaluating the model with test cases, the domain experts became more comfortable with that aspect as well.

Our next steps are to develop a simple user interface to the model inference engine and then pilot the tool on the factory floor where we can measure the diagnostic performance against current methods.

FUTURE CHALLENGES

The combination of current diagnostic modeling theory and practice is mature and offers a valuable set of next-generation tools to improve manufacturing and operations practices. Looking to the future, research in the field promises several ways that current Bayes networks diagnostics can be extended to offer more value.

Intel Research has sponsored a Strategic Research Project (SRP) in Statistical Computing on Bayes Networks and Graphical Models. Many of the emerging techniques discussed in this section (in addition to the algorithms needed for diagnostic inference) have been implemented in Intel's release of the open source *Probabilistic Network Library* [22].

Diagnostic modeling can be applied to process faults in the same way the example in this paper applied it to machinery faults. In semiconductor manufacturing most problems first come to light either because a problem (a "primary indication") is detected in the process or in the product itself. For example, particle counts on the wafer or electrical tests of the dice indicate that something is wrong in the process. Conceptually, a Bayes network can model the set of causes leading up to the problem and infer from measurements which cause is at fault. It can only do this, of course, if the possible causes are understood well enough to model them, and therefore the model would be limited to well-understood parts of the process. Furthermore, process diagnostic models could initiate the use of a machinery diagnostic, since the first indication of a machinery breakdown is often detected by the process.

There are extensions to Bayes networks for temporal modeling (Dynamic Bayes Networks) that could be used to consider how problems evolve over time. "Unscheduled maintenance" by its nature does not exploit the temporal aspect. It takes place entirely when the breakdown occurs. A diagnostic model that tracked the evolution of the machine or process state over time could predict the machine state, for use in predictive maintenance.

The dynamics of the manufacturing process itself may be amenable to Bayes network modeling. The difficulties with modeling poorly understood processes apply here also. A current trend in research in the field is to build a preliminary model from the process diagnostic model and use this as a prior distribution for learning parts of the model from data where data are available. The advantage of such an approach is the insight the diagnostic model offers into the workings of the larger model. Techniques for learning model structure are an area of active research.

There are also Bayes networks extensions to decision making, called Partially Observable Markov Decision Processes (POMDPs). Control problems that arise in Automated Process Control (APC) can be formulated and solved as POMDPs. POMDPs have the advantage of transparency, something they share with dynamic Bayes networks. POMDPs are computationally challenging and also an area of active research.

Bayes networks and their extensions have applications generally to modeling where uncertainty and optimization apply, such as automating network and server cluster management, product test sequence optimization, and supply chain management.

SUMMARY

Innovation occurs as a combination of theoretical and practical advances. Bayes networks offer a comprehensive and principled theory for diagnostic modeling. Their adoption in industry has been slow, due in part to the learning curve of the theory to apply them. Although such models prove challenging to build, they are capable of automating hard diagnostic tasks.

ACKNOWLEDGEMENTS

We acknowledge the assistance and contributions to this work by the DSL lab, the University of Pittsburgh, and Steve Zambroski, Jolyon Clarke, and Quoc Truong. All examples and network illustrations were created with DSL's Genie* software.

REFERENCES

[1] "Integrated Circuit," *Encyclopædia Britannica Premium Service*, Oct. 1, 2004, http://www.britannica.com/eb/article?tocId=34340*.

[2] "Semiconductor device fabrication," Wikipedia. 2004, http://en.wikipedia.org/wiki/Semiconductor_device_fabrication*.

[3] Agosta, J. M. and J. Weiss, "Active Fusion for Diagnosis, Guided by Mutual Information Measures," *Proceedings of the 2nd International Conference on Information Fusion*, Sunnyvale, CA, pp. 337-344, July 1999.

[4] Agosta, J. M. and O. Kipersztok, editors, "Proceedings of the 2nd Bayesian Modeling Applications Workshop" of *Uncertainty in Artificial Intelligence*, 20, Banff, CA, 7 July 2004.

[5] Bayer-Zubek, Valentina, "Learning Diagnostic Policies from Examples by Systematic Search," *Uncertainty and Artificial Intelligence, 20,* pp. 27-34, 2004.

[6] Cowell, R. G., Dawid, A. P., Lauritzen, S. L., Spiegelhalter, D. J., *Probabilistic Networks and Expert Systems*, Springer-Verlag, NY, NY, 1999.

* Other brands and names are the property of their respective owners.

[7] Cox, R.T., *The Algebra of Probable Inference,* John's Hopkins Press, Baltimore, MD, 1961.

[8] Druzdzel, M.J. and van der Gaag, L.C., "Building probabilistic networks: Where do the numbers come from?", *IEEE Transactions on Knowledge and Data Engineering 12*, pp. 481-486.

[9] Heckerman, D., B. Nathwani., "An evaluation of the diagnostic accuracy of Pathfinder," *Computer and Biomedical Research, 25*, pp. 56-74, 1992.

[10] Heckerman, D., J. S. Breese, "Decision-Theoretic Troubleshooting: A Framework for Repair and Experiment " *Microsoft Technical Report* MSR-TR-96-06, Redmond, WA: March 1996.

[11] Heckerman, D., J. S. Breese, and Koos Rommelse, "Decision Theoretic Troubleshooting," *Communications of the ACM,* Vol. 38, No. 3, pp. 49-56, March 1995.

[12] Howard, R. A., "Information Value Theory*" IEEE Transactions on Systems Science and Cybernetics,* Vol., SSC-2, No. 1, pp. 22-26, August 1966, in *Readings in Decision Analysis,* Stanford Research Institute, Menlo Park, CA, 1977.

[13] Jagt, R. M., "Support for Multiple Cause Diagnosis with Bayesian Networks," *M.S. Thesis, Information Sciences Department*, U. of Pittsburgh, 2002.

[14] Jaynes, E.T., *Probability Theory: the Logic of Science,* University Press, Cambridge, UK, 2003.

[15] Jensen, F. V., *Bayesian Networks and Decision Graphs,* Springer-Verlag, NY, NY, 2001.

[16] Jordan, M. Ed., *Learning in Graphical Models,* MIT Press, Cambridge, MA, 1999.

[17] Kalagnanam, J. and M. Henrion, "A Comparison of Decision Analysis and Expert Rules for Sequential Diagnosis," in R. D. Shachter, T. S. Levitt, L. N. Kanal and J. F. Lemmer, Eds. *Uncertainty and Artificial Intelligence, 4,* pp. 253-270, New York, NY: North Holland, 1990.

[18] Lauritzen, S. and D. Spiegelhalter, "Local computations with probabilities on graphical structures and their applications to expert systems," *J. Royal Statistical Society B*, 50, pp. 157-224, 1988.

[19] *Parts America; The Fastlane to Auto Parts and Accessories* http://www.partsamerica.com*.

[20] Pearl, J., *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan-Kaufman, San Mateo, CA, 1988.

[21] Pradhan, M., M. Henrion, G. Provan, B. Del Favero, and K. Huang, "The sensitivity of belief networks to imprecise probabilities: An experimental investigation," *Artificial Intelligence Journal, 84:1-2,* pp. 357, July 1996.

[22] "Probabilistic Network Library" on SourceForge https://sourceforge.net/projects/openpnl/*.

[23] van der Gaag, L.C. and M. Wessels, "Selective Evidence Gathering for Diagnostic Belief Networks," *AISB Quarterly, 86,* pp. 23-34, 1993.

[24] Whittle, P., *Probability via Expectation*, Springer-Verlag, NY, NY, 1992.

AUTHORS' BIOGRAPHIES

John Mark Agosta is part of the Statistical Computing SRP in Intel Research. Prior to joining Intel, he built Bayes networks for medical, avionics, and utility and automobile clients at Knowledge Industries and at SRI International. John received his Ph.D. degree in the Engineering-Economic Systems Department (now Management Science and Engineering) of Stanford University in 1991. At Stanford he participated in the early development of Bayes networks. His e-mail is john.m.agosta at intel.com.

Thomas Gardos has been with Intel since 1993 and is currently a senior researcher in Intel's IT Research team. His main interests are multimodal interfaces for mobile computing applications, Bayesian network-based diagnostic systems and digital video coding. Tom has thirteen patents in the areas of digital video coding and signal processing. He received his Ph.D. degree in Electrical Engineering from the Georgia Institute of Technology in 1993. His e-mail is thomas.r.gardos at intel.com.

An Architecture and Business Process Framework for Global Team Collaboration

Cynthia Pickering, Information Services and Technology Group, Intel Corporation
Eleanor Wynn, Information Services and Technology Group, Intel Corporation

Index words: collaboration, globalization, team, architecture, virtual

ABSTRACT

Tools for remote team collaboration within businesses have been available since the mid-1980s. Two opposing trends cause complete collaboration solutions to remain elusive. On the one hand, core tool capabilities are developed as point solutions, and then extra functions are added. These added functions may not integrate well with or be as fully developed as the core functionality. On the other hand, enterprises are rapidly globalizing and becoming more dependent on comprehensive collaboration applications to coordinate distributed teams. This means that overall productivity is affected by how well tools, processes, and capabilities are integrated; the tools should not be just a collection of separate features/functions.

An audit of collaboration tools used at Intel showed both overlaps and gaps between remote tools and day-to-day activities of workers. When an employee has so many tools to choose from and furthermore, works on multiple teams, the choices become overwhelming and confusing. The underlying architecture of a realistic solution to these overlaps and gaps must provide integration interfaces within the team collaboration environment, and to other business applications, information technology services, and infrastructure. In this paper, we describe a multi-level approach to integration, and we discuss unique findings about Intel's remote teams that justify our model. An essential element of progress towards the goal of an integrated solution will be the deployment of enabling platforms and the likelihood that these practical, indeed necessary, innovations in collaboration will also provide market pull for Intel's core products. By identifying and addressing our own needs, we can also provide solutions for a significant percentage of the Fortune 500 market that engages global workforces for knowledge work.

INTRODUCTION

Global expansion, outsourcing, competitive pressure to do complex projects more efficiently, and increased focus on work-life balance drive the need for employees to collaborate more effectively.

Globally expanding companies face challenges in melding multiple cultures with diverse values, histories, and perspectives. This can make it hard for individual employees to understand their colleagues even when they share a corporate culture–especially when a company is functionally distributed.

Today, we use information systems to improve the productivity of individuals or to automate tasks. However, these mainstream information systems do little to improve the ability of groups of people to work together on collective tasks such as collaborative problem analysis, idea synthesis, decision making, design, conflict resolution, and planning. Team productivity and performance has the potential to yield exponential results due to synergistic factors, knowledge creation and construction, diverse perspectives, and coping with complexity.

Almost every business process or project involves some form of collaboration and coordination between participants. Globally dispersed teams incorporate talents from different locations, and key team members can be chosen for their proximity to important customers and other stakeholders [1]. With the right collaboration tools, companies can become more agile and reduce product time to market [2].

Teams that effectively collaborate avoid or significantly reduce the following cost factors:

- *Time to market*: Cost of not meeting market window, loss of competitive advantage.
- *Time to information*: Project delays due to lack of information or incorrect information.

- *Cost of duplicate projects*: Unintended duplication of effort.
- *Cost of poor coordination*: Increased risk of severe product flaws and recalls.
- *Travel and relocation*: Remote coordination instead of face-to-face meetings and co-location.
- *Opportunity cost of intellectual capital*: Knowledge worker hours can produce more variation in value than manual worker hours; teamwork hours can produce exponentially more value than individual worker hours.

Remote collaboration software and related research have been under development since the mid-80s [3]. In most cases, the market has offered clusters of point solutions that represent their own core capabilities. Not only do these point solutions fail to encompass complete sets of collaboration needs, they also tend to embody outdated models of how corporations work that no longer apply in the global enterprise.

Moreover, team collaboration products on the market do not support enterprise-scale multiteaming; lack interoperability of needed applications; do not support business process and application integration; nor do they promote fluid switching between asynchronous and synchronous collaboration modes within a single environment. End users must learn different tools for different collaborative activities and move information across multiple environments. At the scale of an 80,000 person enterprise with two-thirds of the employees engaged in multiple teams, these constant shifts in applications and environments are counter-productive. Most collaboration tools lack rich, expressive, user-friendly interaction models, the kind that are needed for groups that rely on distance collaboration full-time and that include members who may never have met face-to-face. For this reason, they may fail to attract pervasive recurring usage.

To address these challenges, Intel's IT Virtual Collaboration Research Team [VCRT] decided to do the following:

- Survey Intel's "virtuality" on five dimensions: time, space, business unit, media, and culture.
- Create a baseline of related external research and current collaboration tool use at Intel.
- Identify the "desired user experience."
- Design user-oriented solution concepts for Intel and similar organizations.
- Define a service-oriented architecture for team collaboration.
- Specify enabling IT infrastructure dependencies.

The baseline and survey work validated the first-hand observations of team members and led to a definition of the desired user experience. By starting with the desired user experience as a goal, the team avoided the trap of thinking in terms of predefined capabilities. Instead we worked from the experience to identify supporting capabilities.

New capabilities offered in the collaboration environment map to the unmet needs identified in the virtuality survey and baseline work. For example, 64% of our employees belong to upwards of three or more teams; yet, multi-teaming is not addressed by existing commercial tools.

Other unique capabilities in our concept vision (see section entitled Visionary Concept) include individual and team workspaces for coordinating among multiple teams, asynchronous workspaces for tracking collaboration across time zones, and expressivity for social bonding when employees don't meet face-to-face. The design also addresses ease of use and navigation via an object-oriented 3D graphic desktop.

These new capabilities are as important as existing ones such as document storage and shared visual communication. However, there was a new focus on integrating all user needs into an interoperable collaboration environment.

The VCRT has made significant progress towards understanding and measuring virtuality, and in developing the overall concept design. Because of the interdependencies among the user interface, business process, applications, architecture, and enabling infrastructure layers, the team continues to explore these layers and their interconnections. Future research will evolve the asynchronous and mobility capabilities, prototype an interactive environment, validate user acceptance, and increase understanding of enabling technologies, architecture requirements, and feasibility.

INTEL VIRTUALITY DATA

A 2003 survey conducted at Intel Corporation [4] with 1260 respondents created an index to measure Intel's degree of team "virtuality." The purpose was to identify the potential payback of a radically new collaboration solution. Virtuality was defined by an initial set of five "discontinuities": time, space, organization, culture, and media [5]. Two new discontinuities added to the construct of virtuality were multiple teaming and differences in tools and practices. Three factors emerged in analysis: team distribution, variety of practices, and workplace mobility. Team distribution measures the degree to which people work with others distributed over different geographies and time zones. Variety of practices measures the degree to which employees experience cultural and

work process diversity on their teams. Workplace mobility is the degree to which employees work in environments other than regular offices, including different Intel sites, home, travel routes, and places outside the company.

On the one hand, we found that being geographically distributed in and of itself had no impact on team performance as measured by conformance to Intel values of discipline, quality, customer-orientation, risk-taking, results orientation, and great place to work. On the other hand, lack of shared work practices and structure, and workplace mobility negatively impacted performance. Cultural differences also posed a challenge to perceived performance. 71% of employees work on teams with people of a different culture or with people who speak different native languages or dialects of English. Thus, cultural differences are a key factor in any collaboration tool strategy.

In interviews with employees about team processes a consistent complaint was that the variety of processes, particularly the variety of tools used in a process, was a main source of frustration. (Variety of tools means employees need to use different tools and processes for different projects, or they have to resort to different tools to accomplish a single task.) The data also pointed out that two-thirds of Intel employees belong to three to ten teams simultaneously. This situation compounds the problem of different tools and processes, as employees must often switch tools for the same activity when going from one team to another. For instance, documents may be stored in a collaboration portal, on a Web site, on a shared drive, sent as e-mail attachments, or simply shared on the desktop in real time. Multiple teams times multiple methods compounds the negative impact of diverse practices.

The survey was conducted to answer questions that came up during the discussion of collaboration tool design to test whether our observations could be generalized across the full spectrum of Intel job types, ranks, business units, and geographies. We were surprised to find that virtuality experienced on the factory floor was similar to that of non-factory knowledge workers. The data provide strong support for our envisioned collaboration solution, which provides a platform for the integration of different processes and tools.

In 2004, we repeated the virtuality index survey and our results revealed that the five key metrics and associated indicators used for the virtuality index all increased. Some of these increases were marginal, while others were statistically significant.

Figure 1 shows indicators with significant per-employee trending from 2003-2004.

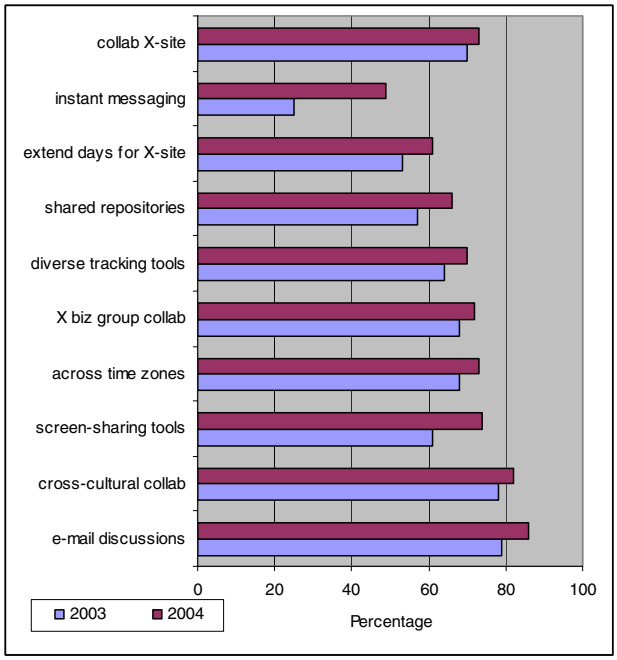

Figure 1: Virtuality trending: 2003-2004

These changes have differing impacts on productivity. Our earlier study found that distance and crossing organizational boundaries alone had no effect on perceived performance, whereas using diverse tracking tools had a negative effect on performance. The latter has increased by 6%, indicating that the need for a common collaboration platform with shared business process tools is increasing. The growth in various available tools also denotes more cross-site work and uptake of existing technologies. Instant messaging in particular had just been introduced at the time of the last study and is showing nonlinear growth with respect to other indicators. This reinforces the VCRT's identification of presence awareness and sociability tools as a key need of the global workforce.

Overall metrics of team performance to Intel values (as perceived by the individual) have either stayed steady or improved (e.g., "work fairly distributed" has increased 10%, from 44-54%). However, one disconcerting finding is that project timeliness has declined from a 52% rating to a 48% rating. There may be many explanations for this change, but a likely one is that coordination is suffering from the increases in the other virtuality factors. If so, this is a problem that needs an immediate solution.

These data confirm our initial hypothesis in tracking virtuality: it is increasing rather than static. An additional hypothesis is that there can be a critical mass threshold that would create nonlinear changes in the organization's performance. Supposing this to be the case, then the increasing dependence on remote teams makes the need

for effective collaboration support a core infrastructure requirement. This hypothesis could also explain why integrated tool sets have not reached the market already: the conditions that impact a company's bottom line have not been fully present up until now. There is no indication that virtuality will decrease.

Valuable research from the academic community is also being incorporated into our ongoing plans. In particular, the work of Carmel and Espinosa [6] has informed the team's sense of urgency and research about time zone differences and the growing need for asynchronous teaming tools to meet the demand for a better coordination capability. Majchrzak and Malhotra [7,8] describe a range of team needs for both cognitive and social integration, as well as task execution, tying these needs to how IT departments can respond. Hinds et al. [9] have compiled recent case studies on global collaboration, and the VCRT expects further engagement with these researchers. As well, the user interface needs of asynchronous team tools, and their ability to maintain user engagement across very large time separations will lead the research into new areas of engagement involving theories of "flow" [10] and experiments with multi-user video games that show what visualizations and activities keep users engaged and bring them back to the online encounter. Some of these findings can be applied to the problem of working environments where analogous activities may occur [11, 12].

VISIONARY CONCEPT

Intel Information Technology (IT)'s Collaboration Vision was developed by our cross-functional VCRT. The team had worked for two years on analyzing the current state, defining the new requirements and then combining them into an integrated vision. Whenever the team tried to communicate the vision, people immediately tried to map it to the tools they were most familiar with. This mapping created a false impression because it did not include key attributes such as integration, interoperability, multi-teaming, or expressive user interface. A well-constructed presentation provided the first step for people to begin to recognize differences between our vision and the existing tools they already used. The deciding step for gaining user buy-in occurred when IT Research funded a "Concept Car" [13] to build a dynamic, visual mock-up of our vision. This dramatically improved our ability to communicate key concepts to managers and employees.

Our vision was featured at the 2003 IT Strategic Long Range Planning and IT Product Line Business Plan internal planning forums where it inspired a multi-year program in IT to drive towards the vision.

Unique characteristics of our vision include the ability to see all "my" multi-team activities in one place, work without time and location boundaries, interact expressively with remote team members, and move effortlessly among business applications and team spaces.

Usage scenarios include meeting in real time across space and knowing who is participating, location of participants and their time of day; working asynchronously on a shared document or project tasks while tracking progress; coordinating responsibilities across projects and meetings, and managing diverse tasks.

Figure 2: Virtual collaboration vision concept

The envisioned solution in Figure 2 supports the VCRT's and individual's natural work flow by coordinating participation in multiple networked teams and providing an immersive interactive experience that engages the user.

Comparison to Other Tools

A comparison of twenty-one different collaboration tools in the market and how they scored against differentiating attributes showed that no one vendor met all of the attributes key to our concept vision [14]. The differentiating attributes that we examined included the following:

- Integration and interoperability.
- Aggregated views for multi-teaming.
- Singular, expressive Electronic Person.
- Combined project and activity timeline integrated with personal calendar.
- Intuitive search scoping and cross repository search.

Scoring was based on third-party reviews and vendor data. If a product met 40% of a differentiating attribute, then it received credit for that attribute. Fifteen of the products exhibited at least one attribute, and six of them at least two. No one product provided all of the above capabilities and none of the products offer cross workspace aggregation to support multi-teaming.

To summarize, gaps between the concept vision and current tools include the following:

- *Too Many Tools and Not Interoperable:* Multiple separate tools required to fulfill all needs; limited interoperability between these tools.
- *Siloed Workspaces:* Hard to share and view content between team workspaces in current tools; need a user-centric view that gives greater flexibility for managing cross-team activities.
- *Interface Not Engaging or Coordinated:* Too many windows to flip through; cannot see what is important in one view; cannot easily get to other applications or transfer data between applications. Lack of graphics and too much reliance on text and hierarchies.
- *Limited Expressivity:* Hard to send and perceive subtle, but important, cues and feedback during virtual meetings (e.g., audio, visual, body language, side-bar interactions). Reliance on media that only some members are strong in.
- *Hard to Find Related Content and People:* No automated way to find related content on Intranet/Internet; randomness to identify "right" people to participate (or consult) on virtual teams.

OVERALL SOLUTION STACK

Technology and globalization are dramatically impacting enterprise organizations. The strong interactive relationship between the changes of the organization and the underlying infrastructure needed to support it [15] forecasts an inflection point or strategic change in the way global teams coordinate.

The needed changes span the entire solution stack including people, process, interaction models, applications, and technology infrastructure/platforms. We will focus on the people and process aspects, then the underlying architecture and standards, followed by the infrastructure and technology platform implications.

Business Process Framework

Figure 3 shows the framework consisting of core business processes for team collaboration, their relationships, dynamic team views, and the aggregating substrate. The virtual team workspace serves as a container for conducting team activities and storing and sharing team information. Teams have meetings that can involve either problem-solving or shared work activities. Because people work on multiple teams, they need cross-team activity management, views that aggregate multiple teams, and the ability to switch easily to the current team context. Each one of these primary components is described next.

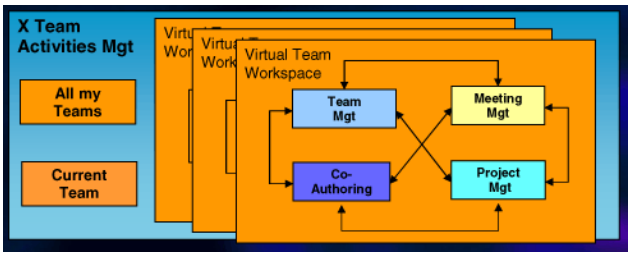

Figure 3: Business process framework

Team Management

Looking at team management from a process perspective leads to questions such as the following:

- What is the work that the team needs to do? Important activities under this question include mission and charter definition, scope, deliverables, roles and responsibilities, timing, and resource requirements.
- How do I get sanctioned by senior management to proceed?
- How do I form my team? Important activities include recruiting team members with the right availability and skills to get the job done, obtaining buy-in from team members' management for members' participation.
- How do I get team members on board with what we are trying to accomplish and help them get to know each other?
- Once the team is performing, how do I keep it performing at optimal efficiency?
- After our work is done, how do we disband, yet still preserve valuable artifacts that point back to the key results of our work?

Team management processes benefit a great deal from services that aggregate and powerfully manage team activities across all teams that a user is participating in. This includes the ability to easily search and share content across team spaces and to summarize and roll up content.

Underlying services for team management throughout the team lifecycle of forming, norming, performing, and disbanding support team activities such as threaded discussions, brainstorming, team self assessment, using team member personas for relationship building, increasing cultural awareness, and providing a launch pad to enhance social interaction and informal communication.

Meeting Management

The core meeting management process is based on effective meeting principles and practices. It includes pre-meeting, during, and post-meeting business processes. Pre-meeting processes involve setting up calendars of people and scheduling resources such as audio bridges,

data and video conferences, and conference rooms and equipment.

During meetings, processes and associated services include real-time usage of the aforementioned resources by the people involved in various roles for the meeting and the capturing of key information from the meeting. Post-meeting processes include preservation of meeting memory via integration with the team's persistent store, as well as ensuring that pertinent items such as required follow-up actions and topics are carried forward and included in future meeting agendas.

We have discovered that there are different interaction styles and activities that lend themselves to either structured, pre-planned meetings as characterized by the core meeting management process described above, or meetings that serendipitously unfold, much like a hallway or chance "water cooler" conversations (ad hoc meetings). Both have their merits. Ad hoc meetings do not have a formal process associated with them, but do need capabilities such as remote presence awareness and management, "click to meet," chat or instant messaging, voice over IP (VoIP), data and application sharing, and video conferencing, if desired.

The lack of overlapping work hours in which to schedule meetings is a critical issue for teams spanning multiple time zones, such as 24x7 software teams [6]. A cross-Pacific team may have only one or two hours of overlapping work time; the same is true of an Israel and US West Coast team. When you add to this different week patterns and national holidays, the amount of available overlapping time becomes even smaller. The compensating behavior is to stretch the workday at both ends. To solve this problem we have had to think seriously about a new meeting type: the asynchronous meeting. This would increase the time span of the meeting event to include a range of available working hours. The asynchronous meeting would be more bounded and structured than a serial hand-off of work, and would require special design features to maintain interactivity. Activities could be completed anywhere within the time window that the asynchronous meeting takes place.

Team members are assigned activities with pre-conditions and deadlines to start and complete them. A workflow scheduler suggests possible open timeslots to assignees that do not violate the start and deadline conditions Assignees can place the activities on suggested open timeslots in their calendar most convenient to them and within their normal work day. Activities that lend themselves to asynchronous meetings include preparing assigned sections of content for documents or presentations; review and annotation of presentations and documents by team members; gathering and summarizing information that the team needs to take into consideration; and completion of other tasks assigned in prior meetings.

A status-monitoring view enables the team members to see everyone's progress towards completion of the asynchronous activities. Personalized reminders go out automatically relative to an individual's time zone and calendar prior to their own settings for task start, at task start and due times, and for overdue tasks.

Asynchronous meetings do not preclude real-time interaction. For example, if two team members happen to notice that they are both online in the asynchronous team workspace they can initiate an ad hoc meeting for a quick real-time dialog. This is one reason that the ability to easily switch between both modes of collaboration is needed.

Often, asynchronous meetings precede real-time meetings where the team gathers to do the types of activities that lend themselves to real-time interaction such as reaching consensus and making decisions, discussion of complex topics that have some ambiguity, and in general to promote team norming and storming.

Project/Program Management

The project/program management core process supports project or mission teams throughout the project lifecycle. This capability leverages and augments the team collaboration workspace, asynchronous, and real-time collaboration capabilities. Additional project management processes and services include the project lifecycle framework, issue tracking, project schedule tracking, task assignments, resource management, critical path management, workflow management, and status reporting.

Co-Authoring

The co-authoring core process includes collaborative creation of unstructured content such as documents (our initial scope) and specialized content creation (future scope) such as Web authoring, software coding, and complex designs. Collaborative services such as in-line markups, annotations, comments, and review and approval are also included in the co-authoring capability. Co-authored content is stored and managed in the Team Collaboration Workspace while it is work in progress (WIP).

Team Collaboration Workspace

The team collaboration workspace is the primary substrate that enables teams to organize and manage the work that they must do together. This includes creating, sharing, and managing documents, e.g., unstructured content, asynchronous discussions, polling and team surveys, meetings, decisions, and action required (AR) tracking. The primary mode of interaction with the workspace is asynchronous or non-real-time, although team members

may visit the workspace simultaneously. Tight integration with communication and real-time collaboration tools, such as instant messaging and application sharing, allow team members to connect with each other while in the team workspace if desired. Capture and storage of real-time collaboration sessions to the team workspace for future reference is also supported. Meeting workspaces that support effective meeting practices enable capture of pre-, during, and post-meeting collateral.

Cross Team Activities Management

This capability collects information across multiple teams from the personal, topical, and business perspective. It provides services to manage team activities across all teams that a user belongs to, a subset or all teams in a business group, and teams working on related projects. Examples include the ability to easily find and share content across team spaces and to summarize and roll up content across teams, as well as to provide a common interaction model across multiple teams for the core collaboration processes and capability areas. This capability is essential to creating the integrated experience desired by users when using team collaboration tools.

Contextual Views

This capability provides contextual collaboration, or the ability to connect and collaborate with others from your current context, which is at the forefront of attention. A simple example of contextual collaboration would be integration of presence awareness and instant messaging with e-mail and the team collaboration workspace, or the ability to conduct in-line discussions within on-line documents. A more complex example involves embedding core team collaboration processes and capabilities with specific business processes and applications, such as product design, allowing collaboration to take place within the context of the business process.

Electronic Person

Electronic Person bundles information about what individuals choose to share with others. It includes both static and dynamic profile data. For example, Electronic Person may include picture(s), location information, presence and availability data, a bio, a resume, an emoticon that reflects mood or emotions, or time of day. Electronic Person also serves as a convenient launch pad to contact the associated person(s) via convenient communication channels such as instant messaging, email, soft or regular phone, and/or video. Electronic Person is depicted in Figure 2 as pictorial profile cards of people on the team. It is our mechanism for bringing more expression into the collaboration environment and for helping people to get to know each other better and to increase social interaction and cultural awareness among team members who are not located in the same place.

Architecture and Standards

The solution approach is based on a services-oriented architecture with well-defined, standard interfaces for services that are exposed to application developers, system engineers, and other shared application service provider sources. We add another level of value-added abstraction by collecting services into assemblies of services that define common core collaborative business processes–which we call business process service groups. Examples of business process service groups include meeting management (pre, during, post) and team management. Figure 4 shows the high-level conceptual architecture.

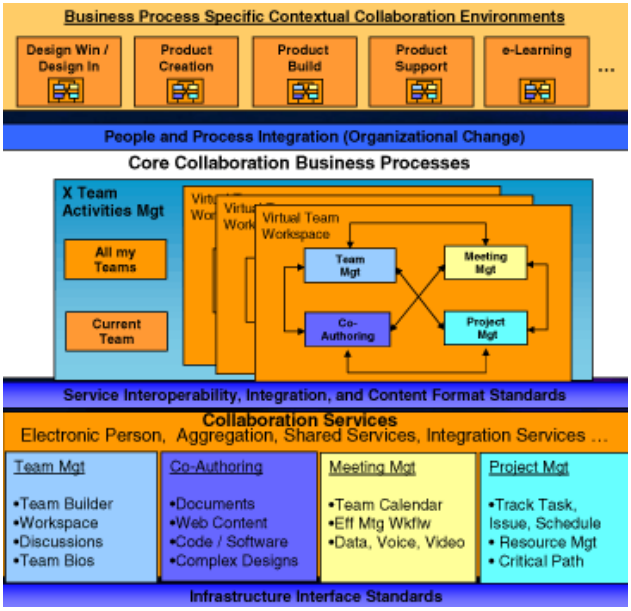

Figure 4: Conceptual architecture

The top layer is the business process-specific contextual collaboration view that is based on vertical business processes that make up a horizontal product lifecycle. The boxes in this layer represent business applications in the different vertical customer segments that have embedded core capabilities based on collaborative activities that most teams engage in. These core collaboration business processes are shown expanded in the middle layer and were discussed previously. The middle layer shows how the core collaboration processes are inter-related and used across multiple team workspaces. It also includes an aggregation function to provide cross team, personal, and contextual views. The bottom layer shows lower-level collaboration services that are collected into business process service groups that realize the higher-level business processes in the middle layer. The collaboration services can also be assembled differently to form processes more specific to the business need and/or called individually by applications. The thinner layers between the primary layers represent standards and integration

between services and the infrastructure, as well as organizational considerations for assimilating collaboration capabilities and processes to improve the way teams work together.

Industry standards key to enabling collaboration include the following:

- **VoIP:** voice, multimedia conferencing.
- **SIP:** session initiation protocol. SIP is a text-based protocol, similar to **HTTP** and **SMTP**, for initiating interactive communication sessions between users. Such sessions include voice, video, chat, interactive games, and virtual reality.
- **SIMPLE**: SIP extensions for Instant Messaging and Presence Awareness.
- **CPIM**: Common Profile for Instant Messaging/Chat, Presence Awareness.
- **ebXML**: eBusiness XML.
- **HumanML**: Human Markup Language. HumanML is designed to represent human characteristics through XML. The aim is to enhance the fidelity of human communication.
- **xCal:** XML DTD for iCalendar.
- **iTIP:** calendar free/busy time.
- **VideoML:** Video Markup Language.
- **MPEG4:** Video Compression.
- **SMIL:** Synchronized Multimedia Integration Language that enables simple authoring of interactive audiovisual presentations. SMIL is typically used for "rich media"/multimedia presentations that integrate streaming audio and video into images, text, or any other media type.

Technology Infrastructure Implications

The underlying technology infrastructure needed to support global team collaboration faces the following challenges:

- *Infrastructure Readiness:* Network, security, mobility, and information management infrastructures will require significant and costly upgrades to enable our vision for collaboration.
- *Desktop Real Estate and Design:* Designing for an integrated solution may require 3D interfaces or multiple screens. At the same time, the platform needs to work on small mobile devices and low-bandwidth networks. Design for flow [10, 20] experience is essential for asynchronous teaming.
- *Inter-company is hard:* Security, legal, interoperability issues are magnified for heterogeneous environments.

Several emerging platforms present an opportunity to explore alternative environments to anticipate and influence the global team coordination inflection point. These platforms include software developed by Intel Research and HP Labs, including PlanetLab [16, 17], Miramar [18], and Croquet [19]. We have begun engaging with developers and researchers to further explore these platforms and their role in the overall solution architecture.

CURRENT PROGRESS AND ROADMAP

In 2003, we conducted the virtuality index research, and developed the vision concept, architecture, and five+ year roadmap to get to the vision. A reuse program was also begun with the goal of broadly sharing components and saving money by not having redundant development efforts.

The 2004 implementation focus has been to lay the foundation for consolidating redundant tools into a common platform, providing some common workspace templates for teams, meetings, and simple projects, and to eliminate redundant products in the environment. We have augmented the reuse program with developer standards and guidelines specific to our collaboration platform and are validating component certification processes with real components targeted to land in the environment. In 2004, we also conducted research in emerging usage models, supporting capabilities, technologies for asynchronous and ad hoc meetings, social networking, adaptive physical workplaces (conference rooms and offices), converged communications, expressivity via Electronic Person, and multi-tasking. Detailed reference and solution architectures for capabilities to be implemented in 2005 have also been developed.

In 2005 implementation priorities include the integrated project management platform, SIP-based ad hoc meetings infrastructure with initial 1:1 and 1:2 usage models, secure external collaboration, simple document approval processes, digital conference rooms, and some targeted integration with enterprise and line of business portals. We also hope to lay the foundation for the Electronic Person. Research will continue in the areas of asynchronous meetings, expressivity, multi-teaming implications such as multitasking, and mobile collaboration needs. We will also develop interactive prototypes of the vision concept to validate feasibility, acceptance, and to evolve the interaction design and underlying infrastructure requirements. Cross-repository integration and search, and integrated workflow capabilities are some of the infrastructure components

needed to support multi-teaming. Again, the design of detailed architectures for 2006 implementations will be informed by research and technology development projects.

In 2006, we will begin to introduce effective meeting processes for asynchronous meetings and pre-scheduled real-time meetings, as well as for managing teams throughout their lifecycles. The processes will be embedded into the environment via workflow tools. Mobile collaboration will see some improvements such as roaming voice conferencing. In 2006, significant information and content management infrastructure upgrades such as the meta data registry will be available.

During 2007, we will upgrade multi-party ad hoc meetings, integrated multi-media conferencing, and integration of team workspace and office applications with business intelligence for collaborative decision making. We will also implement a seamless multi-teaming interface, enhance Electronic Person expressivity, and upgrade portfolio management and richer user interface shell constructs. By 2007, we expect to be reaping the benefits of our re-use program more broadly, which will also accelerate the move towards contextual collaboration.

CONCLUSION

Changes in the global business environment are apparent in many dimensions. They are regularly reported and analyzed in the business press. Our own study of Intel employees confirms the generalizations found in the marketplace and gives us an idea of how Intel is trending in terms of distributed teaming. We have described an approach for understanding the organizational situation, and have responded to it as information technology researchers and designers. We now have a strong concept of the desired user experience. We have identified social and technical factors involved in remote teaming, and defined core team collaboration processes

Since interactions that people previously conducted face to face happen increasingly over the network, it is critical that collaboration tools maintain context and interactivity to promote fluid and sustained communication. This requires interoperability of tools, architecture, and processes as well as new capabilities never imagined when collaboration applications were first invented.

Our architecture provides the foundation for global team collaboration processes and capabilities that integrate with business applications.

To maintain alignment with our vision and make progress against our roadmaps requires strong links between research, technology path finding, proof of concept prototyping, architecture, and the implementation program.

Establishing enterprise-scale core collaboration solutions that allow some customization into specific business process contexts increases the value of information technology to the business.

ACKNOWLEDGMENTS

VCRT members responsible for defining the visionary collaboration concept include Brad Anders, Tim Brooke, Mark Chuang, Keith Feher, Tammie Hertel, Chuck House, Mei Lu, Don Meyers, Phil Tierney, Amanda Ueno, and Nathan Zeldes.

REFERENCES

[1] Levenson, A. R., and Cohen, S. G., "Meeting the performance challenge, calculating return on investment for virtual teams," in *Virtual teams that work: Creating conditions for virtual team effectiveness* , C.B. Gibson & S. Cohen, Eds., San Francisco, CA, Jossey Bass, pp. 145-174, 2003.

[2] Maxfield, J., Fernando, T., and Dew, P., "A distributed virtual environment for collaborative engineering," in *Presence*, 7(3), 241–261, 1998.

[3] Greif, Irene, *Proceedings of the first Conference on Computer Supported Cooperative Work*, 1986, ACM Press, Austin, TX, USA.

[4] Lu, Mei, Wynn, Eleanor, Chudoba, Katherine, Watson-Manheim, Mary Beth, "Understanding Virtuality in a Global Organization: Toward a Virtuality Index," *ICIS 2003*, Seattle, Washington, USA.

[5] Watson-Manheim, M. B., Chudoba, K. M., and Crowston, K., "Discontinuities and continuities: a new way to understand virtual work," *Information Technology & People*, 15 (3), 191-209, 2002.

[6] Espinosa, J. Alberto, Carmel, Erran, "The Impact of Time Separation on Coordination of Global Software Teams," *Journal of Software Process: Improvement and Practice,* (In Press).

[7] Majchrzak, A., Malhotra, A., Lipnack, J., and Stamps, J., "Can absence make a team grow stronger?", *Harvard Business Review,* May 2004.

[8] Malhotra, A. and Majchrzak, A., "Enabling knowledge creation in far-flung teams: Best practices for IT support and knowledge sharing," *Journal of Knowledge Management* (expected: vol. 8, # 3, 2004).

[9] Hinds, Pamela and Sara Kiesler, *Distributed Work*, MIT Press, Cambridge, MA, 2002.

[10] Csikszentmihalyi, M., *Creativity: Flow and the Psychology of Discovery and Invention*, Harper Collins, New York, NY, 1996.

[11] Reeves, Byron, *The Media Equation: How Computers, Television and Interfaces are Social,* Cambridge University Press, Cambridge, UK, 1988.

[12] Reeves, Byron, *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*, Center for the Study of Language and Information, Stanford University, Palo Alto, CA, 2003.

[13] Pickering, Cynthia, "Using IT Concept Cars to drive innovation," in *IT Innovation for Adaptability and Competitiveness,* Fitzgerald, B and E Wynn, editors, *IFIP WG 8.6 Working Conference,* Kluwer Academic Publishers, Dordrecht, Holland, 2004.

[14] Meyers, Don, "Collaboration Tools Comparison," *Intel Information Technology Internal Report,* 2003.

[15] Ciborra, Claudio U., "From Control to Drift: the Dynamics of Corporate Information Infrastructures," Oxford University Press, Oxford, UK, 2000.

[16] Peterson, Larry, Anderson, Tom, Culler, David, and Roscoe, Timothy, "A Blueprint for Introducing Disruptive Technology into the Internet," in *Proceedings of ACM HotNets-I Workshop 2002*, Princeton, New Jersey, USA.

[17] Moore, Terry, Beck, Micah, and Plank, James S., "An End-to-End Approach to Globally Scalable Programmable Networking," in *Proceedings of the Workshop on Future Directions in Network Architecture (FDNA'03)*, Karlsruhe, Germany.

[18] Light, John, Miller, John David, Miramar, "A 3D Workplace," *IEEE IPPC*, 2002.

[19] Smith, D. A., A. Kay, A. Raab, and D.P. Reed, "Croquet: A Collaboration System Architecture," at http://www.opencroquet.org, 2003.

[20] Chen, Hsiang, Rolf Wigand and Michael Niland, "Exploring web users' optimal flow experiences," *Information Technology & People,* v. 13.4 MCB University Press, UK, 2000.

AUTHORS' BIOGRAPHIES

Cindy Pickering is an Information Technology principal engineer in Intel's IT Research. She focuses on global team collaboration, including people, process, and technology. Her research combines needs assessment, new concept development, emerging usage models definition, and enabling architecture and technologies exploration. Cindy received a B.S.E.E. degree from Penn State University in 1981. She joined Intel in 1991 and has 23 years of industry experience. Her e-mail is cynthia.k.pickering at intel.com.

Eleanor Wynn does research and innovation planning in information technology. She holds a Ph.D. from the University of California, Berkeley, specializing in social interaction. She has worked in technology requirements for 25 years. She is co-editor of *Information Technology & People*. At Intel, Eleanor analyzes the fit between the organization and new technologies. She represents Intel at Santa Fe Institute. Her e-mail is eleanor.wynn at intel.com.

For further information visit:

developer.intel.com/technology/itj/index.htm