
Publisher

Richard Bowles

Managing Editor

Stuart Douglas

Content Architect

Jackson He

Program Manager

Stuart Douglas

Technical Editor

David Clark

Technical Illustrators

MPS Limited

Technical and Strategic Reviewers

Jim R Blakley

Dusty Robbins

Nikhil Sharma

Mauri Whalen

Bo Huang

Steve Birkel

Jackson He

Billy Cox

Intel Technology Journal

Copyright © 2012 Intel Corporation. All rights reserved.
ISBN 978-1-934053-56-0, ISSN 1535-864X

Intel Technology Journal
Volume 16, Issue 4

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except as permitted under Sections 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 750-4744. Requests to the Publisher for permission should be addressed to the Publisher, Intel Press, Intel Corporation, 2111 NE 25th Avenue, JF3-330, Hillsboro, OR 97124-5961. E-Mail: intelpress@intel.com.

This publication is designed to provide accurate and authoritative information in regard to the subject matter covered. It is sold with the understanding that the publisher is not engaged in professional services. If professional advice or other expert assistance is required, the services of a competent professional person should be sought.

Intel Corporation may have patents or pending patent applications, trademarks, copyrights, or other intellectual property rights that relate to the presented subject matter. The furnishing of documents and other materials and information does not provide any license, express or implied, by estoppel or otherwise, to any such patents, trademarks, copyrights, or other intellectual property rights.

Intel may make changes to specifications, product descriptions, and plans at any time, without notice.

Fictitious names of companies, products, people, characters, and/or data mentioned herein are not intended to represent any real individual, company, product, or event.

Intel products are not intended for use in medical, life saving, life sustaining, critical control or safety systems, or in nuclear facility applications. Intel, the Intel logo, Intel Atom, Intel AVX, Intel Battery Life Analyzer, Intel Compiler, Intel Core i3, Intel Core i5, Intel Core i7, Intel DPST, Intel Energy Checker, Intel Mobile Platform SDK, Intel Intelligent Power Node Manager, Intel QuickPath Interconnect, Intel Rapid Memory Power Management (Intel RMPM), Intel VTune Amplifier, and Intel Xeon are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products.

For more complete information about performance and benchmark results, visit www.intel.com/benchmarks

[†]Other names and brands may be claimed as the property of others.

This book is printed on acid-free paper. (∞)

Publisher: Richard Bowles
Managing Editor: Stuart Douglas

Library of Congress Cataloging in Publication Data:

Printed in USA

10 9 8 7 6 5 4 3 2 1

December 2012

Notices and Disclaimers

ALL INFORMATION PROVIDED WITHIN OR OTHERWISE ASSOCIATED WITH THIS PUBLICATION INCLUDING, INTER ALIA, ALL SOFTWARE CODE, IS PROVIDED “AS IS”, AND FOR EDUCATIONAL PURPOSES ONLY. INTEL RETAINS ALL OWNERSHIP INTEREST IN ANY INTELLECTUAL PROPERTY RIGHTS ASSOCIATED WITH THIS INFORMATION AND NO LICENSE, EXPRESS OR IMPLIED, BY ESTOPPEL OR OTHERWISE, TO ANY INTELLECTUAL PROPERTY RIGHT IS GRANTED BY THIS PUBLICATION OR AS A RESULT OF YOUR PURCHASE THEREOF. INTEL ASSUMES NO LIABILITY WHATSOEVER AND INTEL DISCLAIMS ANY EXPRESS OR IMPLIED WARRANTY RELATING TO THIS INFORMATION INCLUDING, BY WAY OF EXAMPLE AND NOT LIMITATION, LIABILITY OR WARRANTIES RELATING TO FITNESS FOR A PARTICULAR PURPOSE, MERCHANTABILITY, OR THE INFRINGEMENT OF ANY INTELLECTUAL PROPERTY RIGHT ANYWHERE IN THE WORLD.

Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products.

For more information go to <http://www.intel.com/performance>

Intel’s compilers may or may not optimize to the same degree for non-Intel microprocessors for optimizations that are not unique to Intel microprocessors. These optimizations include SSE2, SSE3, and SSE4 instruction sets and other optimizations. Intel does not guarantee the availability, functionality, or effectiveness of any optimization on microprocessors not manufactured by Intel. Microprocessor-dependent optimizations in this product are intended for use with Intel microprocessors. Certain optimizations not specific to Intel microarchitecture are reserved for Intel microprocessors. Please refer to the applicable product User and Reference Guides for more information regarding the specific instruction sets covered by this notice.

Notice revision #20110804

INFORMATION IN THIS DOCUMENT IS PROVIDED IN CONNECTION WITH INTEL PRODUCTS. NO LICENSE, EXPRESS OR IMPLIED, BY ESTOPPEL OR OTHERWISE, TO ANY INTELLECTUAL PROPERTY RIGHTS IS GRANTED BY THIS DOCUMENT. EXCEPT AS PROVIDED IN INTEL’S TERMS AND CONDITIONS OF SALE FOR SUCH PRODUCTS, INTEL ASSUMES NO LIABILITY WHATSOEVER AND INTEL DISCLAIMS ANY EXPRESS OR IMPLIED WARRANTY, RELATING TO SALE AND/OR USE OF INTEL PRODUCTS INCLUDING LIABILITY OR WARRANTIES RELATING TO FITNESS FOR A PARTICULAR PURPOSE, MERCHANTABILITY, OR INFRINGEMENT OF ANY PATENT, COPYRIGHT OR OTHER INTELLECTUAL PROPERTY RIGHT.

A “Mission Critical Application” is any application in which failure of the Intel Product could result, directly or indirectly, in personal injury or death. SHOULD YOU PURCHASE OR USE INTEL’S PRODUCTS FOR ANY SUCH MISSION CRITICAL APPLICATION, YOU SHALL INDEMNIFY AND HOLD INTEL AND ITS SUBSIDIARIES, SUBCONTRACTORS AND AFFILIATES, AND THE DIRECTORS, OFFICERS, AND EMPLOYEES OF EACH, HARMLESS AGAINST ALL CLAIMS COSTS, DAMAGES, AND EXPENSES AND REASONABLE ATTORNEYS’ FEES ARISING OUT OF, DIRECTLY OR INDIRECTLY, ANY CLAIM OF PRODUCT LIABILITY, PERSONAL INJURY, OR DEATH ARISING IN ANY WAY OUT OF SUCH MISSION CRITICAL APPLICATION, WHETHER OR NOT INTEL OR ITS SUBCONTRACTOR WAS NEGLIGENT IN THE DESIGN, MANUFACTURE, OR WARNING OF THE INTEL PRODUCT OR ANY OF ITS PARTS.

Intel may make changes to specifications and product descriptions at any time, without notice. Designers must not rely on the absence or characteristics of any features or instructions marked “reserved” or “undefined”. Intel reserves these for future definition and shall have no responsibility whatsoever for conflicts or incompatibilities arising from future changes to them. The information here is subject to change without notice. Do not finalize a design with this information.

The products described in this document may contain design defects or errors known as errata, which may cause the product to deviate from published specifications. Current characterized errata are available on request.

Contact your local Intel sales office or your distributor to obtain the latest specifications and before placing your product order.

Copies of documents which have an order number and are referenced in this document, or other Intel literature, may be obtained by calling 1-800-548-4725, or go to: <http://www.intel.com/design/literature.htm>

INTEL® TECHNOLOGY JOURNAL

END TO END CLOUD COMPUTING

Articles

Foreword	7
Cloud Computing and Open Data Centers.....	8
The Past, the Present, and the Future of Cloud Computing	20
China: A Booming Market for Cloud Computing	26
Service Security and Compliance in the Cloud	34
On the Concept of Metadata Exchange in Cloud Services	58
OpenStack and Intel Technology Best Practice	78
The Hadoop Stack: New Paradigm for Big Data Storage and Processing	92
Some Key Cloud Security Considerations	112
Implementing Cloud Storage Metrics to Improve IT Efficiency and Capacity Management	126
Distributed Cloud with Edge Server	140
A Journey Toward a Compute Continuum: Client-Aware Cloud Services for Smart Clients	154
Graphics Acceleration for HTML5 and JavaScript Engine JIT Optimization for Mobile Devices	164
Mobile OS Architecture Trends	178

FOREWORD

Reuven Cohen (@ruv)

Senior Vice President Virtustream,

Contributing Author Forbes Magazine's "The Digital Provocateur"

The year 2012 marks an important milestone in the emergence of cloud computing, including significant industry collaboration. We've seen a remarkable transformation in how we interact with Internet technologies on an individual basis and collectively as an industry. From new industry alliances to open application stacks, never before have we witnessed such a rapid transformation in how we work and interact.

Often described as a revolution, cloud computing is an important transition, a paradigm shift in IT delivery. It refocuses how we view IT while creating new opportunities and challenges. Cloud computing has the potential to transform the design, development, and deployment of next-generation technologies.

In this issue of the *Intel Technology Journal* we will explore some of the technologies, trends, opportunities as well as the challenges facing this exciting transition in our industry. We've assembled an experienced team of authors and contributors at the forefront of cloud computing who will act as your guide through this new world we call "the cloud."

AUTHOR BIOGRAPHY, REUVEN COHEN

Reuven is an early innovator in the cloud computing space as the founder of Toronto based Enomaly in 2004 (Acquired by Virtustream in 2012). Enomaly was among the first to develop a self service infrastructure as a service (IaaS) platform (ECP) circa 2005. As well as SpotCloud (2011) the first commodity style cloud computing Spot Market.

Apart from his current role as Senior Vice President at Virtustream, Reuven writes "The Digital Provocateur" column for Forbes Magazine, is the co-founder of CloudCamp (100+ Cities around the Globe) CloudCamp is an unconference where early adopters of Cloud Computing technologies exchange ideas and is the largest of the 'barcamp' style of events. He is also the co-host of the DigitalNibbles Podcast sponsored by Intel.

You can learn more about Reuven at his website <http://ruv.net> or on Twitter at www.twitter.com/ruv

CLOUD COMPUTING AND OPEN DATA CENTERS

Contributors

Kai X. Miao

Cloud Infrastructure Group,
Intel Corporation

Jackson He

Software and Services Group,
Intel Corporation

Cloud computing, as an emerging paradigm in the IT industry, promises significant gains in efficiency and flexibility at a time when demands on data centers are growing exponentially. As the tools, building blocks, solutions, and best practices for cloud computing are evolving, challenges exist in the industry to develop sound cloud solutions to meet the requirements of different end users. The technology and industry leadership that Intel brings to this environment is broader and deeper than most realize. Intel's Open Cloud Vision is that cloud computing is federated, automated, and client-aware. Moving the industry toward that promise will require a focus on three industry-wide pillars of cloud computing—efficiency, simplification, and security—and on developing solutions that are open, multi-vendor, client-aware, and interoperable. The nature of diverse applications and services requires diverse cloud architecture and instances. This article also addresses the challenges of building towards a distributed cloud-based environment on Open Cloud Vision.

Introduction

Cloud computing is an evolution in which IT consumption and delivery are made available in a self-service fashion via the Internet or internal network, with a flexible pay-as-you-go business model, which requires a highly efficient and scalable architecture. In a cloud computing architecture, services and data reside in shared, dynamically scalable resource pools, often virtualized. Those services and data are accessible by any authenticated device over the Internet or an internal network. The key attributes that distinguish cloud computing from conventional computing are:

- Compute and storage functions are abstracted and offered as services
- Services are built on a massively scalable infrastructure
- Services are delivered on demand through dynamic, flexibly configurable resources
- Services are easily purchased and billed by consumption
- Resources are shared among multiple users (multi-tenancy)
- Services are accessible over the Internet or internal network by any device

“Rather than a revolution, cloud computing is an important transition, a paradigm shift in IT delivery.”

Rather than a revolution, cloud computing is an important transition, a paradigm shift in IT delivery—one that has broad impact and significant challenges to consider. Cloud computing offers the potential for a transformation in the design, development, and deployment of next-generation technologies—technologies that enable flexible, pay-as-you-go business models that will alter the future of computing from mobile platforms and devices to the data center.

The impetus behind cloud computing is the ever-increasing demands placed on data centers that are near capacity and resource-constrained. These demands include growing needs to manage business growth and increase IT flexibility. In response to these challenges, cloud computing is evolving in the forms of both public clouds (deployed by Internet companies, telecommunications companies, hosting service providers, and others) and private or enterprise clouds (deployed by enterprises behind a firewall for an organization's internal use).

Public clouds are being driven by explosive growth of Internet data and traffic as the Internet matures and Internet-based services proliferate. By 2015, over 3 billion people with more than 15 billion devices will access the Internet^[1]—over twice today's demand. The monumental requirements associated with the data center build-outs needed to satisfy this growing demand can only be met with the increased efficiency, performance, and flexibility of cloud architectures.

Private clouds are being driven by the expanding business demands on enterprise IT. More and more data centers find themselves facing real limits, whether based on lack of power, lack of room, lack of server capacity, or lack of network bandwidth. Expanding traditional infrastructures to meet these challenges quickly uncovers multiple inherent inflexibilities.

Cloud computing is far beyond data center virtualization. Initially, virtualization technologies allowed data centers to consolidate server infrastructure to save cost. Next, flexible resource management technologies added the ability to more dynamically allocate data center resources. This further reduced costs and also increased data center flexibility and performance, ushering in a new era of technology development and deployment. Software vendors have begun to design robust management features and technology optimizations for enterprise and public clouds based upon virtualization. Hardware vendors have extended their management tools and reliability features to include increased flexibility. The era of cloud computing can be seen as the next natural step, where significant automation and scalability become possible. Cloud computing offers a path to optimized use and rapid deployment of resources, improved operational efficiency, and potential for significant cost savings. When fully realized, cloud computing infrastructures can provide competitively significant IT agility, flexibility, and adaptability through systems that are efficient, simplified, and secure.

Today's workers are increasingly likely to utilize multiple devices, including smart phones, tablets, and PCs to access information. They embrace new applications and devices in their personal life, and expect those same capabilities to be available at work. Yet today, most of these devices function on a standalone basis, requiring users to juggle multiple independent devices. In addition, when it comes to the ability to access, display, manipulate, or secure data, clearly some devices are more capable than others. Yet today, most Internet services are “dumbed down,” perhaps being able to recognize screen size or display, but with limited ability to take advantage of enhanced security or performance on more capable devices.

“By 2015, over 3 billion people with more than 15 billion devices will access the Internet^[1]—over twice today's demand.”

“Cloud computing offers a path to optimized use and rapid deployment of resources, improved operational efficiency, and potential for significant cost savings.”

Though cloud computing can be viewed as an evolutionary step, it is a fundamental shift and there are challenges to consider:

- Maintaining the stability of mission-critical applications as you transition into cloud environments is paramount.
- Intellectual property protection, data security, and privacy all require additional attention and new tools if shared resources in a public cloud are to be used.
- The automation and flexibility of resource pools will be imperfect while cloud computing tools evolve.
- Solutions must be selected that provide for flexibility and interoperability.
- Cloud-based applications must enable (rather than negatively impact) user productivity, regardless of the device used or connectivity.

Intel's Open Cloud Vision

Cloud computing technology is evolving at a fast pace and many cloud services and vendors are entering the market to enable the development of private clouds for enterprise IT. Several public cloud providers are expanding their services to support enterprises and small and midsize businesses. In Intel's numerous conversations with vendors, analysts, and customers, we've identified key themes that emerge as critical to what customers want from cloud computing infrastructures and solutions. Intel's vision for cloud computing over the next five years centers on three themes that are essential to help overcome key challenges and realize the full potential and value of cloud computing, *Federated*, *Automated*, and *Client-aware*, as shown in Figure 1.^[2]

"Intel's vision for cloud computing over the next five years centers on three themes that are essential to help overcome key challenges and realize the full potential and value of cloud computing."



Figure 1: Intel's Open Cloud Vision
(Source: Intel Corporation, 2011^[2])

Federated means communications, data, and services can move easily within and across cloud computing infrastructures. To accomplish truly federated systems, smooth interoperability across many platforms and solutions must be a reality. Today, the industry is just reaching the point that enterprises can move or migrate workloads within and between their own data centers. Data center operators are far from being able to have data and services seamlessly and securely scale beyond their borders to span public and private clouds when desired. Intel's cloud vision calls for a level of federation that enables the movement of workloads and data from one service provider to another; burst implementations between internal private cloud and public cloud providers if additional capacity is needed; and secure and reliable data flow across vendors, partners, and clients. The first element mentioned above, the one related to the "federated" clouds, will probably be the most difficult to develop, but this task will represent the primary area of interest of the Open Data Center Alliance^[3], who'll ultimately draw up the necessary standards for creating public and/or private clouds that can be easily interconnected without affecting their level of security and/or functionality. Plus, getting the various entities out there to share precious computing resources will also be a difficult task, especially if we're talking about direct competitors.

Automated means that cloud computing services and resources can be specified, located, and securely provisioned with very little or zero human interaction. Today, the industry faces many gaps in automation. According to IDC's Data Center Survey in 2009, virtualization thus far has failed to reduce complexity. The number of server instances that can be managed by the average systems administrator has increased from 37 to only 41 comparing nonvirtualized servers to virtualized servers. Moreover, virtual machines are generally statically provisioned rather than automatically responding to user needs. Data center management remains very manual today—patching of servers doesn't scale reliably. Intel's cloud computing vision calls for automation that dynamically allocates resources to agreed-upon service levels and optimizes the data center for maximum resource utilization and power efficiency. This includes automation of provisioning, resource monitoring, reporting of consumption for bill-back, and workload balancing.

Client-aware means that cloud-based applications are able to dynamically sense and take advantage of the capabilities of the end point device to optimize application delivery in a secure fashion while enhancing the experience of the end user. Today, there are certain frameworks that allow for some level of data center intelligence and scaling to support the client being served, but they are neither consistently applied nor ubiquitous. Many of today's Internet services default to the lowest common denominator device even if the user is accessing the service with a more capable device such as a PC. Conversely, other services are difficult to use on a handheld device because they were written for a PC. Intel's cloud computing vision calls for the data center and service provider to enable secure access and optimal experience across a range of devices, by enabling the cloud to sense and dynamically adjust to take advantage of the attributes and capabilities of the client device. These attributes include items

"Data center operators are far from being able to have data and services seamlessly and securely scale beyond their borders to span public and private clouds when desired."

"The number of server instances that can be managed by the average systems administrator has increased from 37 to only 41 comparing nonvirtualized servers to virtualized servers."

“Building open, interoperable solutions that embrace standards evolving the infrastructure to realize the full potential of cloud computing will not be easy.”

“The evolution of cloud computing requires open, interoperable solutions that embrace standards.”

such as a device’s remaining battery life, policies, and connectivity. At the same time, client device capabilities can affect the overall performance of cloud solutions: taking advantage of local performance on the client device can enable a better end-user experience, and security capabilities on the client device can ensure security policies are applied at the device.

Pillars to Enable Intel’s Open Cloud Vision

Building open, interoperable solutions that embrace standards evolving the infrastructure to realize the full potential of cloud computing will not be easy. It will require cooperative development and specific focus by many providers and customers across the IT landscape. We believe that to move towards this vision of cloud computing, individual organizations and the IT industry as a whole need to focus on three key areas:

- *Efficiency:* While the need for computing across different industries increases exponentially, resources are limited. These resources include space, power, cooling capacity, qualified IT professionals, money for infrastructure, and money for operations. To meet the increasing demand of computing with existing or available resources will require better efficiency both compute infrastructure and processes.
- *Simplification:* Generally, the growth of a system inherently increases complexity, and this is certainly true of IT infrastructures. Multiple architectures complicate management. Increased server utilization raises network bandwidth requirements. And systems from different vendors typically present integration complications. For cloud computing environments to deliver on their promise, simplification must underlie cloud architectures and practices.
- *Security:* Both business risk and compliance requirements make data security paramount. In an environment with abundant traditional security issues, cloud computing creates new challenges because it moves data in new ways, often outside of traditional physical boundaries. The successful implementation of cloud computing requires new security models to meet new challenges.

In addition to the three pillars described, to achieve the vision, delivery of open, interoperable solutions that embrace industry standards is essential. When multiple providers (of solutions, hardware, software, integration, or processes) act independently, poor interoperability and lack of flexibility are the natural results, which are in direct contradiction to the main promises of cloud computing. The evolution of cloud computing requires open, interoperable solutions that embrace standards.

Open Data Center Initiative and Intel Cloud Solutions

The Intel® Open Data Center Initiative is Intel’s comprehensive engagement with ecosystem partners and end customers to help speed the delivery of technology that enables more secure, efficient, and simplified cloud data

centers that preserve IT flexibility and choice. Intel is working directly with leaders in global IT for enterprise and service providers and is an advisor to the Open Data Center Alliance. The Alliance will define a roadmap of the highest priority usage models for cloud and next-generation data centers and lay out the requirements to address with multivendor, interoperable solutions that embrace standards. Intel responds to these usage models and others that we identify through our end-user engagements to deliver products and technologies that meet the requirements of the usage models. We then engage and rally leading systems and solution providers to deliver products and solutions and enable deployment via reference architectures and best practices through Intel® Cloud Builders^[4]. Intel Cloud Builders brings together leading systems and software solutions vendors to provide best practices and practical guidance on how to deploy, maintain, and optimize a cloud infrastructure based on Intel architecture. In short, Intel Cloud Builders provides the industry a central point for cloud innovation based on the IT requirements defined by the Open Data Center Alliance and other IT end users. Intel Cloud Builders publishes detailed reference architectures, success stories, and best practices that you can use right now to deploy and enhance your cloud. Using this guidance and interaction with cloud leaders, IT managers can begin utilizing proven solutions to improve cloud security and efficiency while simplifying data center management and operations.

While the Open Data Center Alliance will determine future requirements for cloud infrastructure, the Intel Cloud Builders program will help bring these requirements to life with full solutions. The program now has a total of 90 Cloud reference architectures and solutions with more on the horizon. It represents a community of the most critical providers of technology in the cloud, including Canonical, Cisco, Citrix, Dell, EMC, Huawei, Enomaly, Eucalyptus Systems, Neusoft, Gprox, HP, IBM, Joyent, Microsoft, NetApp, NetSuite, Novell, Parallels, Red Hat, and VMware.

The Open Data Center Alliance is an independent organization of leading global IT managers who have come together to amplify their collective voice to set data center requirements for today and the future that are best-of-breed and enable flexibility and choice. Their mission focuses on delivering next-generation data center and cloud requirements to meet the challenges facing IT today and tomorrow and delivering them in an open, industry-standard, and multivendor fashion. Intel serves as the technical advisor to this organization. Open Data Center Alliance is a coalition of more than 300 leading businesses that together represent more than 100 billion US dollars (USD) in annual IT investment and that have cloud research or projects underway. Alliance Steering Committee members include BMW, Capgemini, China Unicom, Deutsche Bank, Disney Technology Solutions and Services, J.P. Morgan Chase, Lockheed Martin, Marriott International, Inc., National Australia Bank, NTT Data, Terremark, and UBS. The alliance will lay out future hardware and software requirements that lead to more open and interoperable cloud and data center solutions. Intel plays a unique advisory role within the alliance, whose initial membership was purposely focused on end user companies rather than technology providers.

“Intel Cloud Builders provides the industry a central point for cloud innovation based on the IT requirements defined by the Open Data Center Alliance and other IT end users.”

“The Open Data Center Alliance is an independent organization of leading global IT managers who have come together to amplify their collective voice to set data center requirements for today and the future that are best-of-breed and enable flexibility and choice.”

“Many people believe that cloud computing means all the complicated computing tasks will be centralized on the server side in the cloud.”

Cloud 2.0: Balanced Cloud Architecture for Scalable Cloud Services

Many myths exist in the industry today about Cloud Computing. For example, many people believe that cloud computing means all the complicated computing tasks will be centralized on the server side in the cloud, while on the device side, all it takes is simple terminals with a browser, i.e. a typical thin client model which has been around for the last 15 years. As cloud data centers become bigger and bigger, it further enhanced the concept that cloud computing equals big data center build-out and concentrate compute tasks to data centers and thin client will be the future client compute model of choice in the era of cloud computing.

Table 1 below lists a number of myths about Cloud Computing in the industry along with the corresponding aspects of Cloud Computing which are actually happening.

Myth	Fact
Cloud computing is a technology.	Cloud computing is a new service model based on several technologies evolved over time, such as virtualization, service orientation, and dynamic resource management.
Cloud computing equals thin client.	Service orientation can be implemented using Web services, but using Web services will not necessarily result in a service orientation solutions.
Cloud computing means all services are from the cloud data center.	No quite true. There will still be services that are distributed along the network infrastructure or even at local storage devices when the network bandwidth is limited or not available.
Cloud computing services are homogenous, one-size-fits-all.	Cloud services have to be customized and diverse to meet the needs of different people, different devices, and different network connectivity conditions.
Cloud computing is one-way service flow from the data center to devices.	Services could be bidirectional and peer-to-peer in a cloud environment.
Endpoint devices in a cloud environment are all interactive with user interfaces to access services.	Endpoint devices do not have to be interactive. They could be one-way data gathering nodes without any user interface, such as surveillance cameras.
The world could be covered by a few big clouds from a few big cloud service providers.	The needs for cloud services are diverse, sensitive to culture and usage patterns, some private, some public. There is no way a few big companies can cover these diverse needs. There will be many clouds and they need to work together to provide the best user experience.

Table 1: Cloud Computing Myth

Will cloud computing lead us to thin client model? Will the Cloud lead us to the mainframe architecture? Are these myths representing the true meaning of cloud computing? The answer to these questions is clearly “No”.

The promise of cloud computing is to provide seamless computing experience, anytime, anywhere, on any devices. By definition, users’ needs are different with personal preferences, different contexts, and different locations. It will not be possible to have a one-size-fits-all device to meet all the needs. It will have to be multiple devices, whether it is a PC, TV, smart phone, tablet, or even embedded devices inside cars, boats, and smart buildings. In addition, there are great varieties of network connectivity; some devices have broadband connections and some have 3G wireless connections, while others may not have consistent connections at all. Cloud computing has to provide services to all types of devices, at different connectivity context, and to different people. There have been over 3 billion connected PCs already. With more and more smart phones, tablets, and embedded devices connected to the Internet, there will be over 15 billion connected devices using cloud services. Cloud implementation architecture has to keep this in mind to provide balanced compute experiences to best serve users in the most cost effective way—this is the essence of what we called Cloud 2.0—the next generation of intelligent cloud services infrastructure built on top of distributed and balanced compute systems architecture for scalable and customized service delivery to the masses (billions of users) with best user experience for given devices and network environment.

“Cloud computing has to provide services to all types of devices, at different connectivity context, and to different people.”

Figure 2 shows a high level architecture illustration of Cloud 2.0. The large centralized data centers that drive the growth of Cloud 1.0 will still play

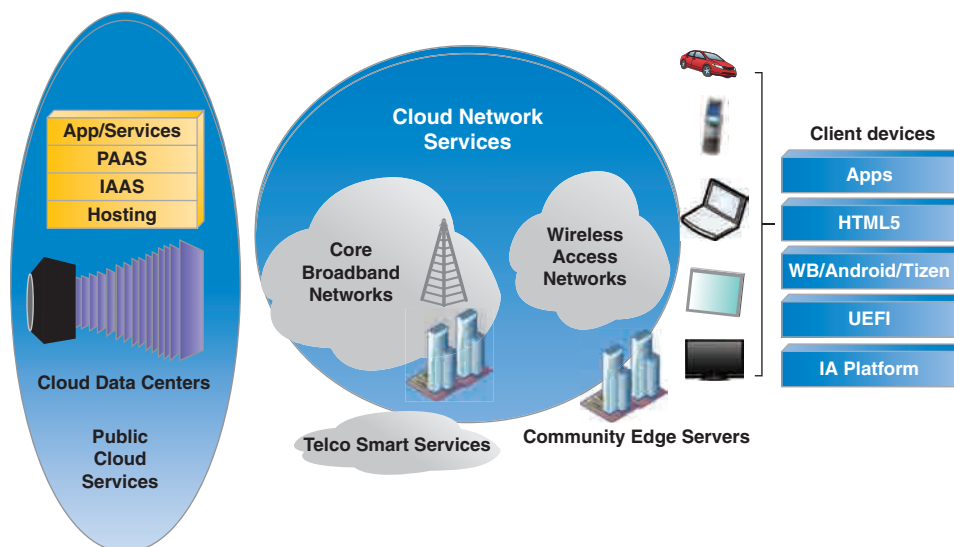


Figure 2: Cloud 2.0 architecture components
(Source: Intel Corporation, 2012)

“The grand vision of cloud computing will not be realized if network bandwidth is insufficient for Cloud services or scalable to millions of users.”

“A cloud service needs to know what the network connectivity condition at the point of service.”

a major role. That is where most cloud services are hosted and managed. However, Cloud 1.0 focuses only on compute, storage, and network resources at the data center. It assumes that the wide area network does not have much intelligence or service capabilities, nor does it emphasize the important roles client devices must play. Cloud 1.0 assumes the wide area network's bandwidth is abundant and can scale to all customer needs. This is not true, especially with a large number of users using different devices with diverse network conditions. The grand vision of cloud computing will not be realized if network bandwidth is insufficient for Cloud services or scalable to millions of users. The additional components of Cloud 2.0 are to enhance the network with distributed resources and intelligent services as an extension to the cloud services from the data center. This could be traditional data cache services, it could also be new services routing data intelligently to make the best use of given bandwidth and running applications or services in the network or at the edge of the network. In this way, not only user experience could be greatly improved, but also can optimal use of resources be achieved, all by having intelligence and compute resources distributed from data center throughout the network.

Specifically, to ensure good user experiences and optimal use of compute resources, Cloud 2.0 must be architected with the following requirements or capabilities in mind.

- User-aware: a cloud service needs to know who accesses the service and whether the person has the right authority, and to verify the user identity before the service is provided.
- Device-aware: a cloud needs to know what device is consuming the service, the capabilities of a device, whether a particular task associated with a service should be executed on a device should the service be customized for best user experience, and so on.
- Context-aware: a cloud service needs to know what the network connectivity condition at the point of service, whether the location or other context information can be used to provide better services, or other devices nearby to help in term of providing resources or data for more efficient service offering.
- Designed for manageability: cloud services, running over independent systems in the data center and across a network, delivered to different client devices, should be fully managed with standard interfaces.
- Designed for scalability: cloud services are meant to scale. They should facilitate any target number of different service consumers defined as required by service consumers. Architecturally, services should take

advantage of the latest programming technologies for multithreading and have clear definition of service boundaries (functionalities and interfaces). In addition, management facilities have to be in place to measure the quality and health of services. Automatically extend service capacity to ensure quality of services.

- Designed for federated solutions: cloud services may span typical boundaries, such as network administrative boundaries, organizational and operational boundaries, and the boundaries of time and space. Services will have the ability to cross corporate or transnational boundaries, requiring a high degree of built-in security, trust, and internal identity, so that they can negotiate and establish federated service relationships with other services following given policies administrated by the management system.
- Cohesiveness across services based on standards in a network of service is essential for service federation.
- In a federated environment, it is possible for technology to run ahead of legal and regulatory boundaries. For instance, although it might be possible to encrypt and package a data set for remote execution in another country to any desired level of tamperproof strength, laws in the originating country may make it illegal to move the data. In this case any legal statutes trump technological options. Flexibility from service orientation still allows outsourcing to countries where these operations are allowed.

“Cloud services may span typical boundaries, such as network administrative boundaries, organizational and operational boundaries, and the boundaries of time and space.”

Conclusion

All in all, we'd dare to say that Intel has some pretty big plans for the future of cloud computing, but the journey ahead will not an easy one. Intel is fully committed to making Intel's Open Cloud Vision a reality.

References

- [1] Cisco: Internet traffic to quadruple by 2015. http://news.cnet.com/8301-1023_3-20067979-93.html
- [2] Intel Cloud 2012 Vision: <http://www.intel.com/content/www/us/en/cloud-computing/cloud-computing-intel-cloud-2015-vision.html>
- [3] Open Data Center Alliance: www.opendatacenteralliance.org
- [4] Intel Cloud Builders: www.intel.com/cloudbuilder

Author Biographies

Kai X. Miao, PhD, principal engineer, Intel China Ltd., has held several technical and management positions at Intel in the past 14 years. Kai Miao received his MS and PhD in Electrical Engineering from University of Cincinnati and BS degree in Communications from Northern (Beijing) Jiaotong University.

Jackson He, PhD is General Manager of Intel APAC R&D Ltd., He has been with Intel for 17 years and has led many technical evangelizations. In the past a few years, Dr. He has been driving cloud technology innovations in both the US and China.

THE PAST, THE PRESENT, AND THE FUTURE OF CLOUD COMPUTING

Contributor

Reuven Cohen

Senior Vice President,
Virtustream

“No longer are users bound by the limitations of a single computing device, but instead are free to experience a multitude of devices, platforms, and mobility (both socially and physically).”

To best understand the current opportunities found in the rapid adoption of cloud computing technology, you must visit the past so you can adapt to the future. Author Reuven Cohen takes us on a guided journey through the cloud like no other person can. So sit back and enjoy the ride into the future of computing.

Introduction

To say cloud computing has entered the collective consciousness of the IT world would be putting it mildly. Over the last few years we’ve seen cloud computing emerge at the heart of a radical shift in the way we consume, deploy, and utilize computing technology within our digital lives. In this article I will explore the roots of the trend over several decades from desktop to mobile, to federated markets, as well as consider its future.

A Brief History

Cloud computing has been referred to as revolutionary, even magical. Like most trends in IT, cloud computing is a combination of a number of underlying trends that have long been in the works, a kind of evolutionary blend of our previous successes and failures. A key term driving the adoption of cloud computing has been the term “the cloud.” In essence the concept of “the cloud” is as a metaphor for the Internet as an operational environment where applications are utilized over the Internet rather than through more traditional means such as a desktop. No longer are users bound by the limitations of a single computing device, but instead are free to experience a multitude of devices, platforms, and mobility (both socially and physically).

To understand this trend we must follow its roots, ones that go back as far the 1960s as seen with Douglas F. Parkhill who first envisioned the coming trend. In his 1966 book *The Challenge of the Computer Utility*^[1], Parkhill, a Canadian electrical engineer, predicted that the computer industry would come to resemble a public utility “in which many remotely located users are connected via communication links to a central computing facility.” A primary tenant of today’s cloud platforms, Parkhill’s “Computing Utility” vision spoke directly to the coming shift we see taking place today.

For many years, Parkhill’s computing utility concept remained unrealized, in part because of the immaturity of the underlying networks and technologies themselves.

In the 1980s and 1990s with the advent of the Internet and later its massive adoption, we began to see the emergence of ISPs and then ASPs, which began to form the first hosted application services. These early service providers made it possible to not only host end user applications that before this point were limited to the realm of desktops and single user servers, but also freed enterprises from managing IT operations not essential to their core businesses. Now applications could be provided as a service over the Internet and managed by a third party. Just as outsourced services had quickly become part of many enterprise strategies, companies were able to outsource their IT in the same way.

Thirty years after Parkhill's book, in the mid-1990s, another trend began to emerge, one where service providers realized that hosting single tenant applications (applications hosted on single server or computer) was not an efficient use of computing resources. The emergence of *virtualization*, or the virtual representation of a computing resource, changed all this. With virtualization, computing resources become transient and adaptive, with the ability to adjust to demands of the macro environment in a near real-time fashion. A perfect storm was brewing with the Internet playing a central role.

Virtualization was the evolutionary missing link, one that gave computational resources a new found manageability and efficiency. For the first time "Virtual machines" would be able to not only scale horizontally (more resources added as needed) but vertically, whereby clones of application components could be replicated at will. This newfound freedom opened a world of possibilities. Freed from the constraints of the previous client/server models of the past, a new breed of service providers rose to take advantage of this flexibility.

The Present State of Cloud Computing

What does cloud computing look like today? Service providers have realized that access to computing capacity is the great equalizer. For the first time large scale computing had been democratized. What previously had been limited to the only the big players was now open to anyone, anywhere. The only the limitation was your imagination.

As important as the access to computing capacity is the data itself, and more importantly the information contained within. To understand this you need to think what the personal computer (PC) has done to information. What the PC revolution started and what the Internet supercharged is information creation. It could be said that more information is now being created in a few months than what was created in the hundreds of years before the Internet ever existed.

A recent article in *The Economist*^[2] points out "that the world contains an unimaginably vast amount of digital information which is getting ever vaster ever more rapidly. This makes it possible to do many things that previously could not be done: spot business trends, prevent diseases, combat crime and so on. Managed well, the data can be used to unlock new sources of economic value, provide fresh insights into science and hold governments to account."

"With virtualization, computing resources become transient and adaptive, with the ability to adjust to demands of the macro environment in a near real-time fashion."

"For the first time large scale computing had been democratized. What previously had been limited to the only the big players was now open to anyone, anywhere."

“Whether big or small, data has become the oil powering the information age.”

For the most part the majority of the information humankind has created has not been accessible. Most of this raw data or knowledge has been sitting in various silos—be it a library, a single desktop, a server, database, or even data center. The most successful companies of the last decade have discovered how to tap into this raw data. They are better at analyzing, mining, and using this mountain of data and turning a useless raw resource into something much more useful: information.

Whether big or small, data has become the oil powering the information age.

The aforementioned *Economist* article also puts the idea of data as power into perspective. “When the Sloan Digital Sky Survey started work in 2000, its telescope in New Mexico collected more data in its first few weeks than had been amassed in the entire history of astronomy. Now, a decade later, its archive contains a whopping 140 terabytes of information. A successor, the Large Synoptic Survey Telescope, due to come on stream in Chile in 2016, will acquire that quantity of data every five days.”

Cloud computing has become an established approach to the management and deployment of applications within a large and growing number of businesses. The reality is that most new software is developed using cloud as the central architectural tenant. By 2015, over 2.5 billion people with more than 10 billion devices will access the Internet^[3].

Gazing into the Future

Looking into the immediate future, an ever-increasing variety of choices are beginning to shape a market that is no longer made up of a few select providers, but is instead a globally expanding selection of regional and industry-specific clouds, each cloud built for the particular requirements of a business with even more particular verticals. Furthering this trend is a burgeoning array of open source software stacks powered in part by a multitude of devices all connected to the cloud that will eventually connect anything and everything.

“An ongoing transition to the cloud is fueling a massive cloud infrastructure build, projected to be worth some 82.9 billion US dollars (USD) by the year 2016^[4].”

An ongoing transition to the cloud is fueling a massive cloud infrastructure build, projected to be worth some 82.9 billion US dollars (USD) by the year 2016^[4]. This new market will feature thousands of vertically focused clouds of all shapes and sizes, ranging from business, infrastructure and developer offerings, to consumer, mobile and gaming services.

Over the next few years a key opportunity within the cloud industry will be the creation of federated cloud ecosystems. These marketplaces will be defined by interoperability among multiple competing cloud computing providers and platforms using agreed-upon standards and application interfaces. One of the key attributes that will help facilitate this shift is the concept of “Global Web Scale” whereby distributed components (services, nodes, applications, users, virtualized computers) come together to form a massive global environment. These federated clouds may consist of hundreds of thousands of computing nodes from a variety of distributed resources both internal and external, and

federating them together will lead to a massive global scale. This scale will require new technologies and techniques.

The management of application content and components will be an important part of this change. Driven in part by the ever-changing resource demand patterns and a lack of cooperation among end-user's applications, particular sets of resources will get swamped by excessive workloads, which significantly undermines the overall utility delivered by the system. The ability to manage application content will become a critical part of future cloud management. How application components adapt to a constantly fluctuating environment will be an important aspect in future cloud deployments and architectures. Improved dynamic capacity allocation methods and components will enable applications to leave and join the system at will.

Finding commonality and standards within measurement, security, and power will help manage and define the growing cloud computing space. Transparency among providers and platforms will also be a significant aspect within globally federated environments. The ability to audit and define trust within a variety of clouds using common security procedures will become paramount. The ability to not only trust but verify the integrity of the entire stack is essential.

Establishing a common unit of measure, just like the kilowatt for electricity, the ability to understand in real time how much aggregate resources are required to get something done, and moreover how much those resources will cost on the smallest of increments as possible will be an important factor. The correlation of economics and measurement of resources will be a significant forthcoming trend. The ability to adapt workloads based on the amount of power consumed or required will be an even more important part of the equation as the cloud becomes larger and more global in its deployments and consumes greater amounts of energy.

In the coming years anything that can either collect or display information will. This will open a world of opportunities to connect anything and everything to the cloud. How this "cloud of things" is managed and maintained will be an increasingly important aspect of both personal and professional computing environments. The line between a personal computer and personal computing will quickly become blurred as everyday things start replacing the more traditional ways we interact with computing technology.

While some may see globalization as the key, the true opportunities will be found on a local basis. Localization will also drive an important part of the landscape for cloud services in the future as we see more and more emphasis placed on emerging markets for growth with varied business sectors. Look to Sub-Saharan Africa and Asian markets such as China and India to lead the way in the use and adoption of cloud technology. These markets are uniquely positioned in that they have green field opportunities with a lack of legacy infrastructure in place. This lack of a legacy infrastructure creates the ideal environment to skip the technologies of the past and move directly to the more efficient cloud-centric ones of the future. This will become a technological advantage especially within high growth markets.

"Finding commonality and standards within measurement, security, and power will help manage and define the growing cloud computing space."

"While some may see globalization as the key, the true opportunities will be found on a local basis."

In Summary

Over the last forty years we have witnessed change at an ever increasing pace from Parkhill's concept of "connecting remotely located users" to complex horizontally and vertically scaled virtual machines. With the supercharged information creation taking place, this pace will only continue accelerating as we enter the next stage of our connected future and federated cloud systems. Those who embrace the cloud will flourish enabling the next generation of technology.

Author Biography

Reuven is an early innovator in the cloud computing space as the founder of Toronto based Enomaly in 2004 (acquired by Virtustream in 2012). Enomaly was among the first to develop a self-service infrastructure as a service (IaaS) platform (ECP) circa 2005. As well as SpotCloud (2011) the first commodity style cloud computing Spot Market.

Apart from his current role as Senior Vice President at Virtustream, Reuven writes "The Digital Provocateur" column for *Forbes Magazine*, is the co-founder of CloudCamp (100+ Cities around the Globe). CloudCamp is an unconference where early adopters of Cloud Computing technologies exchange ideas and is the largest of the "barcamp" style of events. He is also the co-host of the DigitalNibbles Podcast sponsored by Intel.

You can learn more about Reuven at his website <http://ruv.net> or on Twitter at www.twitter.com/ruv

- [1] Parkhill, Douglas F. 1966. *The Challenge of the Computer Utility*. Addison-Wesley Publishing Company, ASIN: B000O121OS
- [2] The Economist, February 25, 2010, Special report from the print edition, http://www.economist.com/node/15557443?story_id=15557443 IDC, 2011, Intel's Vision of the Ongoing Shift to Cloud Computing. Sources: IDC "Server Workloads Forecast" 2009; and IDC "The Internet Reaches Late Adolescence" Dec 2009
- [3] Visiongain, November 7th 2011. Cloud 2011: Moving into the Realm of an Essential IT Strategy, <http://www.visiongain.com/Report/643/Cloud-2011-Moving-into-the-REALM-of-an-Essential-IT-Strategy>

CHINA: A BOOMING MARKET FOR CLOUD COMPUTING

Contributors

Jackson He

Intel Corporation

Li Jiang

Centrin Data Systems

Jerry Xie

China Telecom Cloud Computing Corp.

“Strong government initiatives to drive, the momentum toward cloud computing has gone beyond technical innovations and become a part of the national strategy as part of the next five-year economic plan.”

China is one of the most vibrant and active markets in the world. Unlike IT technology transitions of the past, for cloud computing and the Internet of things (IOT), the Chinese government has put a lot of effort in championing this round of IT infrastructure build-out. Cloud and IOT technologies have been positioned as having strategic importance for the transformation of its domestic economy. China has already started cloud pilot programs in five major cities and makes information technology, including cloud and IOT, one of the key areas of focus in its five-year plan (2011–2015). China is shoulder to shoulder in this round of IT technology transformation with the rest of world.

Introduction

Over the last few years, China has emerged as the largest Internet market with more than 500 million Internet users, the largest PC market, and the largest smart phone market. China has many Internet companies such as Baidu, Alibaba, Tencent, similar to their “sister” companies in the US (Google, MSN, eBay, and so on). With the strong consumer interest for mobile Internet and strong government initiatives to drive, the momentum toward cloud computing has gone beyond technical innovations and become a part of the national strategy as part of the next five-year economic plan.

In the current five-year plan (2011–2015), it was clearly stated that “acceleration of the economic structure by cultivating a new business model, new energy, new material, new bio medicines, high-end manufacturing, new energy sources for automobiles, triple play, cloud computing, the Internet of things (IOT) pilot projects, and so on.” Over the last few years, the government-driven investments in next generation Internet (IPv6), upgrading broadband infrastructure, network security, digital TV and satellite communications was greatly increased. Cloud and IOT pilot projects in five major cities (Beijing, Shanghai, Hangzhou, Shenzhen, and Wuxi) have already started.

At the same time, cloud and mobile Internet services driven by the private sector, championed by leading Internet service companies such as Baidu, Tencent, Sina, Alibaba, Shanda, and Sohu, have taken off and become mainstream in China. Baidu has taken over 80 percent of the Internet search market in China, which attracts both mobile users and traditional PC users; Tencent has more registered users than Facebook; Alibaba/Taobao has more transactions and cash flow than the biggest department stores in China. The momentum of cloud and mobile Internet services is embraced and fueled by enthusiastic user communities that could amount to over 1 billion people!

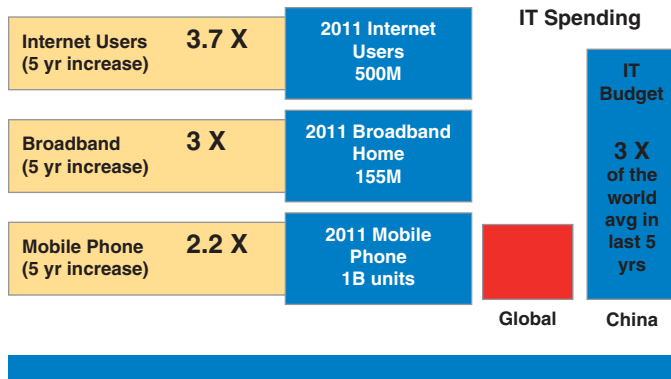


Figure 1: Booming cloud computing market in China
(Source: Intel Corporation, April 2012)

As shown in Figure 1, the Internet market in China has grown much faster than other parts of the world. China has the largest Internet user base—overall 500 million Internet users. The IT spending in the last five years is three times more than the world average. This sets a great foundation for a booming cloud computing market.

In the latest government report this year, RMB 660 Million (USD 100 Million) were allocated by the central government for cloud infrastructure as seed funds, and more matching funds from local government are expected. It was clearly pointed out that the government will accelerate the investment in next generation Internet (IPV6), digital TV network, satellite network, and high-speed, high volume network backbone. The government set the following cloud computing directions:

- Nationwide coordination of cloud computing business and related service development
- Balance the relationship of government-driven and market-driven initiatives
- Leverage cloud computing technology to boost the construction of e-Government, e-Commerce, and optimized distribution of social resources based on information technologies.

At the same time as the strong central government initiatives, local governments have become strong proponents of cloud computing as well. Over 20 cities in China have developed cloud computing plans. There have been special incentives for land, tax, and investment for cloud projects. For example, the “Cloud Sea (云海) Project” in Shanghai is targeted to focus on small business, financial services, healthcare, and media service, while in Beijing there is the “Harmony Cloud (祥云) Project,” targeting to improve information services in Beijing area for e-Government and small businesses. Hundreds of local IT businesses are participating in this type of local government cloud program. According to MIIT (Ministry of Industry and Information Technology), cloud computing is positioned as the fourth IT revolution after Mainframe, PC, and Internet. In the next three years, cloud computing could amount to 200 Billion RMB (USD 30 Billion) revenue in China.

“The government will accelerate the investment in next generation Internet (IPV6), digital TV network, satellite network, and high-speed, high volume network backbone.”

“The “Cloud Sea (云海) Project” in Shanghai is targeted to focus on small business, financial services, healthcare, and media service.”

“It is foreseeable that in the coming years, there will be more systematic and coordinated efforts toward building cloud computing infrastructure in China.”

“With favorable government policies and incentive programs, many Internet services companies have actively started their own cloud initiatives.”

The impact of cloud computing in China is not limited to information services. There are profound new requirements for the whole solution stack from computer hardware, systems software, middleware, to application software, and so on. There are also great many opportunities for local businesses to capture the cloud computing momentum to build vibrant ecosystems from backend data centers to diverse mobile devices, as well as compelling services on top of them. It is foreseeable that in the coming years, there will be more systematic and coordinated efforts toward building cloud computing infrastructure in China. The level of government investment and industrial policy guidance for the IT industry has been stronger in China than in other parts of the world. China could be the “land of opportunity” for cloud computing.

Public Cloud Infrastructure and Services

Back in the late 1990s China was about five years behind the last wave of Internet build-out compared with the western market. However, the growth of Internet companies and Internet services has been phenomenal. China has developed its own complete Internet service ecosystem, parallel to that of the US. Alibab/Taobao, the biggest online store in China followed the similar pattern of growth to eBay in the US, while by 2011, Taobao had over twenty thousand servers handled transactions valued at more than 400 Billion RMB (USD 70 Billion). On a peak day, there could be overall 130 million independent visitors per day using Taobao. Baidu has been leading the Internet search market and took over three times the market share of Google in China as the local market leader. It has become the most popular search engine for both mobile and desktop platforms in China.

When it comes to this round of cloud build-out, Chinese companies are on the same page with their US counter parts. With favorable government policies and incentive programs, many Internet services companies have actively started their own cloud initiatives. Alibaba has started its AliCloud division to focus on cloud infrastructure and service build out for its subsidiaries, including Taobao. Baidu has ventured into mobile Internet services and developed its branded mobile devices. Tencent has built large data centers and developed mobile Internet services to serve up to 700 million registered customers.

Meanwhile, three big telecommunications vendors in China, China Telecom, China Mobile, and China Unicom, positioned cloud and related mobile Internet services as the next strategic frontier for building their future competitive advantages. China Telecom started a dedicated cloud company to oversee its IDC (Internet data center) business and build out its cloud infrastructure for additional service opportunities. China Mobile has spent a significant amount of research resources and developed big cloud solutions since 2010. China Unicom is leading the local cloud customer requirement definitions as one of the Steering Committee members of ODCA (Open Data Center Alliance).

On the software and solutions front, local players are also very active. The Chinese government has clear directions on open source policy. They want local business to use open source solutions where possible.

While most of the PC users are still using Microsoft Windows* OS, all the local Internet service providers use Linux* for their data center and server systems. Like in other parts of the world, Android* has gained strong momentum in the smart device OS market. Android in China is also the preferred OS for most local smart device vendors, despite the strong presence of Apple and iOS* products.

When it comes to virtualization and cloud middleware, open source is by far the most preferred choice. Most of the leader local vendors (such as Baidu, Tencent, Alibaba, Neusoft, and Huawei) have developed their own VM management cloud middleware based on open source virtualization VMM like Xen or KVM. Popular proprietary solutions from VMware and Microsoft are targeted for evaluation and small-scale implementations, but are not the first choice for large-scale production deployments. On the other hand, there are so many competing cloud middleware solutions based on underlying open source virtualization solutions that they have caused confusion to cloud operators trying to select the most appropriate solutions for their business. Further standardization and consolidation are needed to drive up the quality and drive down the confusion of open source cloud middleware solutions.

Vertical Cloud Computing Opportunities

At the same time the public cloud services are driven by leading Internet service providers and next generation network infrastructure are built by telecommunications companies, cloud services for vertical markets are gaining more momentum as well. A lot of the forecasted cloud revenue could be service revenues from SMB (small and medium business). There are about 30 to 40 million SMB in China. In the past, the level of IT technology usage has been very low due to the complexity of hardware and software deployment and management. Cloud computing presents a more feasible solution for scalable and manageable way to provide services. The Chinese government has made information technology applications for SMB a key driver for cloud infrastructure build-out. The goal is that in the next three years, 25 percent of service providers for SMB will leverage cloud technology.

Another vertical market that attracts large government investments is education. The Chinese government has pledged to invest 4 percent of its GDP (approximately US 200 billion according to 2011 GDP) for education by 2015. Of the total education investments, 5 to 6 percent will be for information technology. The government has just recently released the guidelines for education information systems for the next decade, in which cloud technology was clearly called as a key to developing and deploying education services to K-12 schools, university research, and continuous education for workforces. Many universities and research institutes are devoted to education cloud services research. New education usages and methodologies are emerging based on cloud and mobile Internet technologies. According

“While most of the PC users are still using Microsoft Windows OS, all the local Internet service providers use Linux* for their data center and server systems.”*

Android has gained strong momentum in the smart device OS market. Android in China is also the preferred OS for most local smart device vendors, despite the strong presence of Apple and iOS* products.*

“In the past, the level of IT technology usage has been very low due to the complexity of hardware and software deployment and management.”

“As China becomes an aging society (by the United Nations definition, more than 10 percent of population is older than 60 years old), healthcare and preventive medical services are becoming more and more important.”

“China has its special security regulations that require local vendors to play more active roles in developing localized security solutions that serve this diverse market with billions of customers.”

to the education information systems guidelines issued by the Ministry of Education (MOE), by 2015 the education cloud foundation should be developed that covers major cities and universities.

As China becomes an aging society (by the United Nations definition, more than 10 percent of population is older than 60 years old), healthcare and preventive medical services are becoming more and more important. The Chinese government continues its focus on healthcare information systems. In the current five-year plan, the government programs have expanded the healthcare information systems more toward the preventive side. According to the Ministry of Civil Affairs (MOCA), China will invest up to 500 billion RMB (USD 80 billion) in the next 10 years in research and development of senior-friendly communities where sensors and IOT technologies will be used for frontend and cloud infrastructure and services for the backend.

China also has its own sophisticated banking and financial systems. The financial communities are looking for ways to consolidate their IT services by outsourcing part of their services to secured data center operators. Moreover, with the financial and B2B and B2C services becoming more customized and personal in the cloud context, more devices, more data, more customers present more service opportunities. At the same time, security and privacy requirements for financial services, as well as for the entire mobile cloud computing segment at large, are becoming more stringent than ever. More innovative hardware and software combined security solutions are needed. Moreover, China has its special security regulations that require local vendors to play more active roles in developing localized security solutions that serve this diverse market with billions of customers.

Key Technology Challenges and Opportunities Ahead

Given the tremendous cloud market opportunities in China, the technical challenges are equally formidable. While China is facing many of the same technology challenges as the rest of the world, the customer requirements and market conditions in China could present some unique technology development opportunities that will continue to brew a dynamic cloud ecosystem. For the first time in computer industry history, China is at the same level with the western world in terms of market opportunities and usage model innovations. There are few “mature” technologies or usage models to copy or buy from other markets. The following is a summary of some of the key challenges we are facing in the industry; some of them are the same challenges we are facing in other countries as well:

- Security: this is the most important for cloud services in all markets. The Chinese government has special regulations on cryptography algorithms and encryption products. Many international security solutions may not be applicable “as-is.” They have to be adapted, integrated, and deployed through local solution providers. Moreover, there are many new security

requirements that are unique to market in China, such as banking and payment systems integrations.

- **Connectivity:** the cloud computing model depends on interconnected computing devices from servers to smart devices. There are many networking technologies ranging from 10G Ethernet, to InfiniBand*, FCoE (Fiber Channel over Ethernet), Wi-Fi*, and WiDi. Most of these technologies that worked well in other markets should be applicable in China. On the other hand, China has run out of IPV4 addresses and has decided to have broad IPV6 deployment. This will be the first IPV6 deployment of this magnitude in the world. The technology and implications to new applications and services are still evolving. With IPV6, the way we operate data centers and manage connected devices could be quite different. Interoperating between IPV4 and IPV6 networks also presents interesting challenges.
- **Green data center:** energy efficiency of the datacenter has become a universal problem that goes along with the boom of cloud data centers. In China, this problem has become more important, as data centers in China are traditionally not very efficient. With the increasing number of cloud data centers, the Chinese government has set specific goals to improve data center energy efficiency by 40 percent in the next five years. There are two aspects of green data centers: a) at design time, optimize data center layout and use the most energy-efficient servers, storage devices, and network infrastructure designed for centralized management; b) during the operation time of the data center, the central management tools can dynamically monitor the data center workload and resource utilization so that data center can be managed in the most energy-efficient manner by enforcing management policies through management tools. There are challenges for both data center design and operations in China.
- **Cloud manageability:** Besides power management at the data center, cloud resources and services management are essential to successful cloud businesses. A large cloud operation typically requires more than one data center. Cloud management tools have to be able to cover cross-data-center management of both virtual and physical resources, which include cross-data-center workload migration, disaster recovery, service quality assurance, billing, and customer intelligence. In China, open source cloud management solutions are not ready for prime time; proprietary solutions are either homegrown or too expensive. The lack of robust cloud management tools become a major challenge for broad cloud deployment.
- **Cloud software ecosystem:** besides the hardware infrastructure required, cloud computing and related services are built on top of a sophisticated software stack. This is not something a single company can fulfill. A healthy software ecosystem that covers all layers of the cloud stack and different market segments is needed. In addition, the Chinese government strongly promotes open source based solutions. Many software companies in China have built their businesses on open sources software at different levels. With the favorable government policy and established open source community,

“China has run out of IPV4 addresses and has decided to have broad IPV6 deployment. This will be the first IPV6 deployment of this magnitude in the world.”

“In China, open source cloud management solutions are not ready for prime time; proprietary solutions are either homegrown or too expensive.”

“With the vast market and large user population, China needs experienced cloud systems integration vendors to put cloud solutions and services together.”

““Fast second” strategy, which means to learn and follow what works in other market first and then apply these technologies to the vast market opportunities and build out solutions unique to China.”

open source cloud software ecosystem has a good chance to emerge in China.

- System integration services: cloud services require much systems integration effort to put together. Without “the last mile” integrations, there will not be complete solutions with a compelling user experience. The promise of cloud computing will not delivered. With the vast market and large user population, China needs experienced cloud systems integration vendors to put cloud solutions and services together. This could be an organic part of a cloud company. It could also be independent cloud systems integration services provided by a third party.
- Cross-smart-device usages/applications: systems integration could also include application development and porting to support diverse smart devices for different segments and usages. As HTML5 becomes popular, an HTML5-based user experience becomes the key. China has over 100,000 developers in the HTML5 interest groups. There is a good foundation to focus on the culture and behavior of Chinese users to define and develop unique HTML5-based smart applications for cloud services that span all kinds of smart devices from PCs and laptops to tablets and smart phones. On the other hand, we have to deal with the security challenges inherent from HTML5 and make sure that HTML5 runtime engines are optimized to protect user data and deliver the best performance for given devices.

In addition, there should be new strategic thinking about cloud services development in China. Granted that there are still technology gaps to realize the grand vision of cloud computing, China is still behind the western countries in many areas. Nevertheless, as stated in the early part of this article, China is a close follower in cloud solutions, and even a leader in some areas with its strong government programs. One of the strategies could be the “fast second” strategy, which means to learn and follow what works in other market first and then apply these technologies to the vast market opportunities and build out solutions unique to China. Over time, cloud services in China could be leaders in the world. To achieve this, we need broad collaboration with international players at the same time focusing on original innovations to solve unique challenges in China. Chinese cloud solution vendors should embrace international standards where possible and play active roles in standard organizations, such as W3C, DMTF, IETF, and so on.

Summary and Conclusion

China has emerged as a leading market for cloud computing. With its large user population, vibrant market dynamics, and comprehensive government programs, China presents unique business opportunities to grow cloud service infrastructure. While China faces some of the same technical challenges as other markets, it also has unique requirements in many areas. China has the opportunity to closely follow the technology trends and develop cloud services for China. This may be an opportunity for the Chinese IT industry to lead.

References

- [1] A White Paper of China Cloud Application Model Projects: CCID Consulting, October, 2011
- [2] Cloud Computing Weekly References, China Computing Association, 01/11/2012, 01/18/2012, 02/08/2012, 02/15/2012
- [3] IT International Express, MIIT Center for International Exchanges, 03/19/2012
- [4] Cloud Computing 2012 Whitepaper, China Academy of Telecommunication Research, MIIT, April 2012
- [5] Smart Terminal 2012 Whitepaper, China Academy of Telecommunication Research, MIIT, April 2012

Author Biographies

Jackson He, Ph.D. is General Manager of Intel APAC R&D Ltd. He has been with Intel for 17 years and has led many technical evangelizations. In the past a few years, Dr. He has been driving cloud technology innovations in both the US and China.

Jiang Li is Senior Vice President of Centrin Data Systems leading cloud services innovations and driving Chinese cloud government cloud initiatives architecture designs. Prior to joining Centrin Data Systems, he worked at IBM for 17 years and Bank of Hawaii for 5 years. Mr. Li also serves on several government expert committees on cloud computing and IT policies.

Jerry Xie, Ph.D. is the CEO of China Telecom Cloud Computing Corporation, a newly established subsidiary of China Telecom. Prior to that, Dr. Xie worked 20 years in the US and held engineering and management positions with Sony Electronics, Intel, and Sun Microsystems.

SERVICE SECURITY AND COMPLIANCE IN THE CLOUD

Contributors

Raghu Yeluri – Intel Architecture Group

Intel Corporation

James Greene – Intel Architecture Group

Intel Corporation

Sudhir Bangalore – Intel Architecture Group

Intel Corporation

One of the biggest barriers impeding broader adoption of cloud computing is security—the real and perceived risks of providing, accessing and control services in multitenant cloud environments. IT managers need higher levels of assurance that their cloud-based services and data are adequately protected as these architectures bypass or reduce the efficacy and efficiency of traditional security tools and frameworks. The ease with which services are migrated and deployed in a cloud environment brings significant benefits, but they are a bane from a compliance and security perspective. IT managers are looking for greater assurances of end-to-end service level integrity for these cloud-based services. This article explores challenges in deploying and managing services in a cloud infrastructure from a security perspective, and as an example, discusses work that Intel is doing with partners and the software vendor ecosystem to enable a security enhanced platform and solutions with security anchored and rooted in hardware and firmware to increase visibility and control in the cloud.

Introduction

The cloud computing approach applies the pooling of an on-demand, self-managed virtual infrastructure, consumed as a service. This approach abstracts applications from the complexity of the underlying infrastructure, which allows IT to focus on the enabling of business value and innovation. In terms of cost savings and business flexibility, this presents a boon to organizations. But IT practitioners unanimously cite security, control, and IT compliance as primary issues that slow the adoption of cloud computing. These results often denote general concerns about privacy, trust, change management, configuration management, access controls, auditing, and logging. Many customers also have specific security requirements that mandate control over data location, isolation, and integrity that typically use legacy solutions that rely on fixed hardware infrastructures.^[27]

Under the current state of cloud computing, the means to verify a service's compliance with most of the aforementioned security challenges and requirements are labor-intensive, inconsistent, nonscalable, or just not possible. For this reason, many corporations only deploy less critical applications in the public cloud and restrict sensitive applications to dedicated hardware and traditional IT architectures.^[28] For business-critical applications and processes and sensitive data, however, third-party attestations of security controls usually aren't enough. In such cases, it is absolutely critical for organizations to be able to verify for themselves that the underlying cloud infrastructure is secure enough for the intended use.^[29]

“Many corporations only deploy less critical applications in the public cloud and restrict sensitive applications to dedicated hardware and traditional IT architectures.”

This requirement drives the next frontier of cloud security and compliance: building a level of transparency at the bottom-most layers of the cloud by developing the standards, instrumentation, tools, and linkages to monitor and prove that the IaaS clouds' physical and virtual servers are actually performing as they should and meet defined security criteria. Today, security mechanisms in the lower stack layers (for example, hardware, firmware, and hypervisors) are almost absent.

Cloud providers and the IT community are working earnestly to address these requirements, enabling cloud services to be deployed and managed with confidence, with controls and policies in place to monitor trust and compliance of these services in cloud infrastructures. Specifically, Intel Corporation and other technology companies have come together to enable a cloud infrastructure that is highly secure and based on a hardware root of trust, which provides tamper proof measurements of key physical and virtual components in the computing stack, including the hypervisors. These organizations are collaborating to develop a framework to integrate the secure hardware measurements provided by the hardware root of trust into adjoining virtualization and cloud management software. The intent is to improve visibility, control, and compliance for cloud services. For example, having visibility into the trust and integrity of cloud servers allows cloud orchestrators to provide improved controls on onboarding services for their more sensitive workloads—offering more secure hardware and subsequently better controlling the migration of workloads and meeting security policies.

We will discuss how cloud providers and organizations can use the hardware root of trust as the basis for deploying secure and trusted services. In particular, we'll cover Intel® Trusted Execution Technology (Intel TXT) and the Trusted Compute Pool usage models, and envision the necessary ecosystem for implementing them.

Security in the Cloud

Security is a key barrier to the broader adoption of cloud computing. The real and perceived risks of providing, accessing and controlling services in multitenant cloud environments can slow or preclude the migration to services by IT organizations. In a non-virtualized environment, the separation provided by physical infrastructure is assumed to provide a level of protection for applications and data. In the cloud, this traditional physical isolation between applications no longer exists. Cloud infrastructure is multi-tenant, with multiple applications utilizing a shared common physical infrastructure. This provides the benefit of much more efficient resource utilization. However, because the physical barriers between applications have been eliminated, it is important to establish compensating security controls to minimize the potential for malware to spread through the cloud. This section covers the security challenges in the cloud, and provides a set of requirements that have to be addressed for cloud security.

“Today, security mechanisms in the lower stack layers (for example, hardware, firmware, and hypervisors) are almost absent.”

“The real and perceived risks of providing, accessing and controlling services in multitenant cloud environments can slow or preclude the migration to services by IT organizations.”

“Considered a disruptive technology, cloud computing has the potential to enhance collaboration, agility, efficiency, scaling, and availability; it provides the opportunity for cost reduction through optimized and efficient computing.”

Cloud Concepts

Cloud computing moves us away from the traditional model where organizations dedicate computing power (and devices) to a particular business application, to a flexible model for computing where users access applications and data in shared environments.^[3] Cloud computing is a model for enabling ubiquitous, on-demand network access to a shared pool of convenient and configurable computing resources (such as networks servers, storage, applications, and services). Considered a disruptive technology, cloud computing has the potential to enhance collaboration, agility, efficiency, scaling, and availability; it provides the opportunity for cost reduction through optimized and efficient computing.

Many definitions attempt to address cloud computing from the perspective of different roles—academicians, architects, engineers, developers, managers, and consumers. For this article we’ll focus on the perspective of IT network and security professionals; more specifically, for the security architects at service providers and enterprises in their quest to provide a more transparent and secure platform for cloud services.

The National Institute of Standards and Technology (NIST) defines cloud computing through five essential characteristics, three cloud service models, and four cloud deployment models.^{[14][30]} They are summarized in visual form in Figure 1.

Cloud service delivery is divided among three archetypal models and various derivative combinations. The three fundamental classifications are often

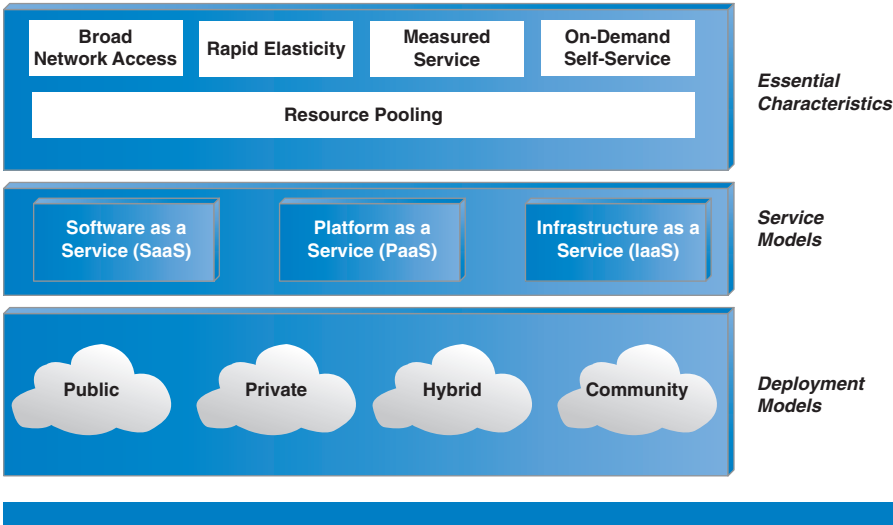


Figure 1: NIST cloud computing dimensions^[14]
(Source: NIST Special Publication 800-53, “Recommended Security Controls for Federal Information Systems and Organizations,” revision3, 2010)

referred to as the SPI Model, where SPI refers to Software, Platform, or Infrastructure (as a Service), respectively defined thus^[5]:

- *Software as a Service (SaaS)* - The capability where applications are hosted and delivered online via a web browser offering traditional desktop functionality, such as Google Docs, Gmail, and MySAP.
- *Platform as a Service (PaaS)* - The capability provided to the consumer is to deploy onto the cloud infrastructure consumer-created or acquired applications developed using programming languages and tools supported by the provider. The consumer does not manage or control the underlying cloud infrastructure including network, servers, operating systems, or storage, but has control over the deployed applications and possibly application hosting environment configurations.
- *Infrastructure as a Service (IaaS)* - The capability where a set of virtualized computing resource, such as compute and storage and network are hosted in the cloud; customers deploy and run their own software stacks to obtain services. The consumer does not manage or control the base, underlying cloud infrastructure but has control over operating systems, storage, deployed applications, and possibly limited control of select networking components (such as host firewalls).

In support of these service models and the NISTs deployment models (public, private, and hybrid), many efforts are centered around the development of both open and proprietary APIs that seek to enable things such as management, security, and interoperability for cloud computing. Some of these efforts include the Open Cloud Computing Interface Working Group, Amazon EC2 API, VMware's DMTF-submitted vCloud API, Sun's Open Cloud API, Rackspace API, and GoGrid's API, to name just a few. Open, standard APIs will play a key role in cloud portability, federation, and interoperability as well as common container formats such as the DMTF's Open Virtualization Format (OVF).^[5]

The architectural mindset used when designing solutions has clear implications on the future flexibility, security, and mobility of the resultant solution, as well as its collaborative capabilities. As a rule of thumb, perimeterized solutions are less effective than de-perimeterized solutions in each of the four areas. Careful consideration should also be given to the choice between proprietary and open solutions for similar reasons.

The NIST definition emphasizes the flexibility and convenience of the cloud, which allows customers to take advantage of computing resources and applications that they do not own for advancing their strategic objectives. It also emphasizes the supporting technological infrastructure, considered an element of the IT supply chain that can be managed to respond to new capacity and technological service demands without the need to acquire or expand in-house complex infrastructures.

Understanding the dependencies and relationships between the cloud computing deployment and service models is critical to understanding

“Many efforts are centered around the development of both open and proprietary APIs that seek to enable things such as management, security, and interoperability for cloud computing.”

“As a rule of thumb, perimeterized solutions are less effective than de-perimeterized solutions.”

“The speed of adoption depends on how security and trust in the new cloud models can be established.”

“Multitenant environment; how do you ensure that systems and applications are appropriately and sufficiently isolated?”

cloud security risks and controls. With PaaS and SaaS built on top of IaaS, as described in the NIST model (above)^[14] inherited capabilities introduce security issues and risks. In all cloud models the risk profile for data and security changes, and is an essential factor in deciding which models are appropriate for an organization. The speed of adoption depends on how security and trust in the new cloud models can be established.

Cloud Security, Trust, and Assurance

There is a significant amount of focus and activity across various standards organizations and forums to define the challenges, issues, and a solution framework to address these drivers. The Cloud Security Alliance, NIST, and the Open Cloud Computing Interface (OCCI), are examples of organizations promoting cloud security standards. Following are the key drivers for cloud security^{[17][12]}:

- Visibility, compliance, and monitoring: providing seamless access to the security controls, conditions, and operating states within the cloud’s virtualization and hardware layers for auditability and at the bottom-most infrastructure layers of the cloud security providers. The measured evidence would enable organizations to comply with security policies and with regulated data standards and controls such as FISMA and DPA.^[13]
- Data discovery and protection: cloud computing places data in new and different places—not just user data, but also the application and VM data (source). Key issues include data location and segregation, data footprints, backup, and recovery.
- Architecture: standardized infrastructure and applications lead to more opportunity to exploit a single vulnerability many times—the BORE (Break Once, Run Everywhere) principle. Considerations for the architecture include: 1) Protection; how do you protect against attacks with standardized infrastructure when the same vulnerability can exist at many places, due to the standardization? 2) Multitenant environment; how do you ensure that systems and applications are appropriately and sufficiently isolated? 3) Security policies; how do you ensure that security policies are accurately and fully implemented across cloud architectures?
- Identity management: identity management (IdM) is described as the management of individual identities, their authentication, authorization, roles, and privileges/permissions within or across system and enterprise boundaries with the goal of increasing security and productivity while decreasing cost, downtime, and repetitive tasks.^[25] From a cloud security perspective, questions like, “how do you control passwords and access tokens in the cloud?” and “how do you federate identity in the cloud?” are very real and thorny questions for cloud providers and subscribers to address.
- Automation and policy orchestration: the efficiency, scale, flexibility, and cost-effectiveness that cloud computing brings is because of the automation; the ability to rapidly deploy resources, scale up and scale

down with processes, applications, and services provisioned securely “on-demand.” A high degree of automation and policy evaluation and orchestration are required so that security controls and protections are handled correctly with very minimal scope of errors and with minimal intervention.

This article focuses on the first set of security drivers—visibility, compliance, and monitoring. Before we delve into these, we should start with a baseline definition of some key security concepts. These terms will form the foundation of what visibility, compliance and monitoring would entail. Let us start with baseline definitions for the terms security, trust, and assurance^[23]:

- *Security.* Concerns the confidentiality, availability and integrity of data or information. Security also includes authentication and nonrepudiation.
- *Trust.* Revolves around the assurance and confidence that people, data, entities, information or processes will function or behave in expected ways. Trust may be human to human, machine to machine (for example, handshake protocols negotiated with in certain protocols), human to machine (for example, when a consumer reviews a digital signature advisory notice on a Web site), or machine to human. At a deeper level, trust might be regarded as a consequence of progress towards security or privacy objectives.
- *Assurance.* Provides the evidence or grounds for confidence that the security controls implemented within an information system are effective in their application. Assurance can be obtained by: 1) actions taken by developers, implementers, and operators in the specification, design, development, implementation, operation, and maintenance of security controls; 2) actions taken by security control assessors to determine the extent to which the controls are implemented correctly, operating as intended, and producing the desired outcome with respect to meeting the security requirements for the system.

While the cloud provides organizations with a more efficient, flexible, convenient, and cost-effective alternative to owning and operating their own servers, storage, networks, and software, it also erases many of the traditional, physical boundaries and controls that help define and protect an organization’s data assets. Physical servers are replaced by virtual ones. Perimeters are established not by firewalls alone but also by highly mobile virtual machines. As virtualization proliferates throughout the data center, the IT manager can no longer point to a specific physical node as being the home to any one critical process or data, because virtual machines (VMs) move around to satisfy policies for high availability or resource allocation. Public cloud resources usually host multiple tenants concurrently, increasing the need for isolated and trusted compute infrastructure as a compensating control. However, mitigating risk becomes more complex as the cloud introduces ever expanding, transient chains of custody for sensitive data and applications. Regulatory compliance for certain types of data would similarly become increasingly difficult to enforce in such models.

“This article focuses on the first set of security drivers—visibility, compliance, and monitoring.”

“Public cloud resources usually host multiple tenants concurrently, increasing the need for isolated and trusted compute infrastructure as a compensating control.”

“Most organizations are already leasing computing capacity from an outside data center to host noncritical workloads such as Web sites or corporate e-mail.”

For this reason, the vast majority of data and applications handled by clouds today isn't business critical and has lower security requirements and expectations, tacitly imposing a limit on value delivered. Most organizations are already leasing computing capacity from an outside data center to host noncritical workloads such as Web sites or corporate e-mail. Some gone a small step further and have outsourced business functions such as sales force management to providers in the cloud. If their workloads were compromised or the business processes became unavailable for a short period of time, the organization might be highly inconvenienced, but the consequences would probably not be disastrous.

Higher-value business data and processes, however, have been slower to move into the cloud. These business-critical functions—for example, the cash management system for a bank or patient records management within a hospital—are usually run instead on in-house IT systems to ensure maximum control over the confidentiality, integrity, and availability of those processes and data. Although some organizations are using Cloud for higher value information and business processes, they're still reluctant to outsource the underlying IT systems, because of concerns about their ability to enforce security strategies and to use familiar security controls in proving compliance.

Security and Compliance Challenges

The four basic security and compliance challenges to organizations are:

- *Governance.* Cloud computing typically increases an organization's reliance on the cloud providers' logs, reports, and attestations in proving compliance. When companies outsource parts of their IT infrastructure to cloud providers, they effectively give up some control over their information infrastructure and processes, even as they are required to bear greater responsibility for data confidentiality and compliance. While enterprises still get to define how information is handled, who gets access to that information, and under what conditions in their private or hybrid clouds, they must largely take cloud providers at their word or their SLA that security policies and conditions are indeed being met and may be forced to compromise to a capability that the provider can deliver. The organization's ability to monitor actual activities and verify security conditions within the cloud is usually very limited and there are no standards or commercial tools to validate conformance to policies and SLAs.^{[4][12]}
- *Co-Tenancy and Noisy or Adversarial Neighbors.* Cloud computing introduces new risk resulting from co-residency, which is when different users within a cloud share the same physical requirement to run their virtual machines. Creating secure partitions between co-resident VMs has proven challenging for many cloud providers, ranging from the unintentional, “noisy-neighbor” (a workload that consumes more than its fair share of compute, storage or I/O resources, therefore “starving” other virtual tenants on that host), to the deliberately malicious; such as when

“The organization's ability to monitor actual activities and verify security conditions within the cloud is usually very limited.”

malware is injected into the virtualization layer, enabling hostile parties to monitor and control any of the VMs residing on a system. Researchers at UCSD and MIT were able to pinpoint the physical server used by programs running on the EC2 cloud and then extract small amounts of data from these programs, by placing their own software there and launching a side-channel attack.^{[4][20]}

- *Architecture and Applications.* Cloud services are typically virtualized, which adds a hypervisor layer to the traditional IT services stack. This new layer in the service stack introduces opportunities for improving security and compliance, but also creates new attack surfaces and potential exposure to risks. Organizations must evaluate the new monitoring opportunities and the risks presented by the hypervisor layer and account for them in policy definition and compliance reporting.^{[4][20]}
- *Data.* Cloud services raise access and protection issues for user data and applications, including source code. Who has access, and what is left behind when you scale down a service? How do you protect data from the virtual infrastructure administrators and cloud co-tenants? Encryption of data—at rest, in transit, and eventually in use—would become a basic requirement. But encryption comes with a performance cost. If we truly want to encrypt everywhere, how do we do it cost effectively and efficiently? Finally, one area that is least discussed is “data destruction.” There are clear regulations on how long data has to be saved (after which it has to be destroyed) and how to handle data disposal. Examples of these regulations include the Sarbanes-Oxley Act (SOX), Section 802 (7 years)^[22], HIPAA, 45 C.F.R. § 164.530(j) (6 years)^[17], and FACTA Disposal Rule.^[8]

Given that most organizations are using cloud services today for applications that are not mission critical or are of low value, security and compliance challenges seem manageable—but this is a policy of avoidance. These services don’t deal with data and applications that are governed by strict information security policies such as health regulations, FISMA regulations, and the Data Protection Act in Europe. The security and compliance challenges mentioned above would become central to cloud providers and subscribers once these higher-value business functions and data begin migrating to private cloud and hybrid clouds, creating very strong requirements for cloud security to provide and prove compliance. Industry pundits believe that cloud value proposition will increasingly drive the migration of these higher-value applications and information and business processes to cloud infrastructures. And as more and more sensitive data and business-critical processes move to cloud environments, the implications for security officers in organizations would be very wide-ranging to provide a transparent and deep compliance and monitoring framework for information security.

Cloud Service Security Requirements

This section addresses the key Security requirements as they pertain to launching Services in a Cloud Infrastructure.

“If we truly want to encrypt everywhere, how do we do it cost effectively and efficiently?”

“The implications for security officers in organizations would be very wide-ranging to provide a transparent and deep compliance and monitoring framework for information security.”

“Trusted computing uses cryptographic and measurement techniques to help enforce a selected behavior.”

“Machine authentication provides a user a higher level of assurance, as the machine is indicated as known and authenticated.”

Trust in the Cloud

One of the pillars of security in the cloud is trust.^[12] A trusted computing system will consistently behave in expected ways, and hardware and software will enforce these behaviors. Trusted computing uses cryptographic and measurement techniques to help enforce a selected behavior because it authenticates the launch and authorized processes. This authentication allows someone to verify that only authorized code runs on a system. This typically covers initial booting and may also cover applications and scripts. Usually, the establishment of trust of a particular component implies the ability to establish the trust for that component with respect to other trusted components. This trust path is known as the chain of trust, with the first component known as the root of trust.^{[4][9]} It is implied that the root of trust be a trusted set of functions that are immune from physical and other attacks. Since an important requirement for trust is to be tamper-proof, cryptography or some immutable unique signature that identifies a component is used. For example: the hardware platform is usually a good proxy for a root of trust, since for most attackers the risk and difficulty of tampering directly with hardware exceeds the potential benefits. With the use of hardware as the initial root of trust, one can then measure software (such as hypervisor or operating system) to determine whether unauthorized modifications have been made to it. In this way, a chain of trust relative to the hardware can be established.

Trust techniques include hardware encryption, signing, machine authentication, secure key storage, and attestation.^{[9][6]} Encryption and signing are well-known techniques, but these are hardened by the placement of keys in protected hardware storage. Machine authentication provides a user a higher level of assurance, as the machine is indicated as known and authenticated. Attestation provides a means for a third party (also called trusted third party) to affirm that firmware and software that are loaded are correct, true, or genuine. This is particularly important to cloud architectures based on virtualization.

Cloud Governance, Risk, and Compliance

Centralized controls and the specialized expertise of cloud services providers will enable security technologies for the computing infrastructure to be deployed far faster and more efficiently in cloud environments than if those same technologies were deployed in traditional enterprise IT environments. In fact, we believe by the end of this year, cloud providers will be able to introduce the first IaaS clouds built on measured trust environments. These new, more secure clouds will give organizations more flexible, affordable and efficient alternatives for shifting high-value business processes and data into private clouds.

Although not every organization will need the high security afforded by a trusted computing environment, every organization using cloud services could benefit from the vastly improved control and transparency that a measured chain of trust enables. Simply being able to verify conditions in the cloud services stack down through the hypervisor is a huge step forward in providing visibility into actual states and activities within the cloud and in better

regulating how cloud resources are managed. Internal and private clouds built on a measured chain of trust will:

- Strengthen an organization's ability to enforce differentiated policies in private clouds.
- Enhance monitoring for compliance at all layers within the cloud.
- Streamline the auditing process.
- Allow for more flexible usage and billing for secure computing resources.

Hardware Root of Trust: Building Security from the Ground Up

Organizations that are using (or desire to use) cloud services are starting to require cloud service providers to better secure the hardware layer and provider greater transparency into the system activities within and below the hypervisor.^{[4][6]} This means that cloud providers should be able to:

- Give organizations greater visibility into the security states of the hardware platforms running the IaaS for their private clouds.
- Produce automated, standardized reports on the configuration of the physical and virtual infrastructure hosting customer virtual machines and data.
- Provide policy-based control based on the physical location of the server where the virtual machines are and control the migration of these virtual machines onto acceptable locations based on policy specifications (such as some FISMA and DPA requirements dictate).
- Provide measured evidence that their services infrastructure complies with security policies and with regulated data standards.

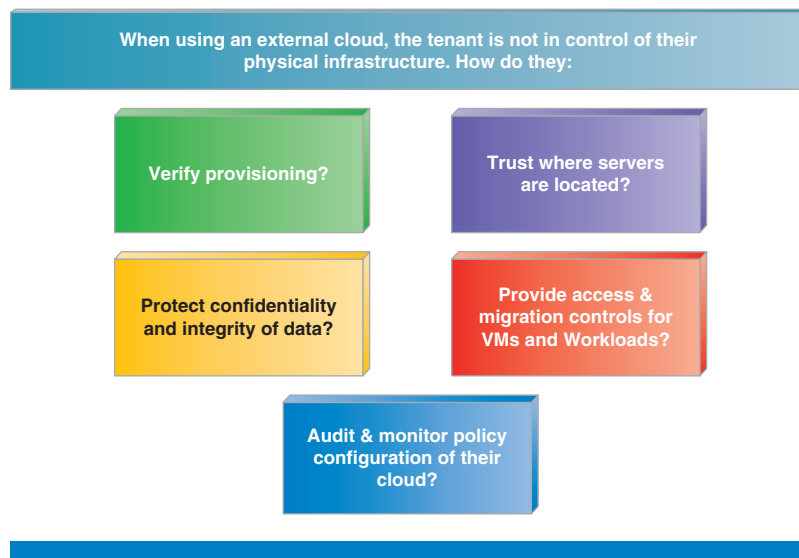


Figure 2: Summary of the top five trust issues from a cloud subscriber's perspective^[17]

(Source: Orrin, S. Information Security and Risk Management Conference, 2011 ISACA, Session 241: Building Trust and Compliance in the Cloud)

“A growing ecosystem of technology companies is collaborating to develop a new interoperable trusted computing infrastructure.”

What is needed are a set of building blocks for the development of “Trustworthy Clouds.” These building blocks can be summarized as^[17]:

- Creating a chain of trust rooted in hardware that extends to include the hypervisor.
- Hardening the virtualization environment using known best methods.
- Providing visibility for compliance and audit.
- Using trust as part of the policy management for cloud activity.
- Leveraging infrastructure and services to address data protection requirements.
- Using automation to bring it all together and achieve scale and management efficiency.

Cloud providers and other members of the IT community are preparing to address this need. A growing ecosystem of technology companies is collaborating to develop a new interoperable trusted computing infrastructure. The goal of this emerging infrastructure is to eliminate attacks such as virtual rootkits and to provide the foundation for a hardware root of trust, which establishes a bottom-up security structure based on hardware.

The Advantages of Cloud Services on a Trusted Computing Chain

The advantages of building cloud services on a trusted chain of computing resources include^{[4][12][26]}:

- Improving co-residency security by ensuring the launch of only trusted code. Protecting against untrusted software isn’t just about malware; it also applies to more benign conditions, such as the improper migration or deployment of virtual machines. To illustrate, if load-balancing software or a cloud administrator attempts to move virtual machines from an unsecured computing platform to a secure, trusted one, the management software would prevent the incoming VMs, since it originated from an unsecured platform.
- Preventing the unsafe transit of secure virtual machines. In the same way that VMs coming from an unsecured platform would not be allowed to move to secured platforms, VMs originating on secured platforms would not be allowed to move to unsecured ones. If, for instance, an administrator attempted to transfer a secured VM onto a new server, the virtualization management console would first perform a policy check on the outgoing VM and then measure the security configurations of the new server against accepted standards. If the new server couldn’t meet the secure standards required to host the VM, the virtualization management console or security policy engine would block the VM’s move and log the attempt.
- Maximizing operational efficiency by creating trusted pools of systems. Once platform trustworthiness can be measured, cloud providers can put such measurements to use in building trusted pools of systems, all with

identical security profiles. Hypervisors can then make more efficient use of secure clouds, moving VMs with similar security profiles within zones of identically secured systems for load balancing and other administrative purposes—all while protecting data in conformance with regulated standards and policies.

- Building secure clouds customized to comply with the most rigorous requirements. The secure cloud's ability to map high-trust zones of systems will enable organizations and cloud providers to customize their clouds to comply specifically with PCI DSS, HIPAA, or other highly controlled information standards. Then, trusted pools of cloud-based resources—all compliant with the same set of information standards—could be dynamically allocated to optimize workloads. Such a scenario would extend the cloud's efficiency and scalability benefits to even the most strictly controlled business processes and heavily regulated industries. Furthermore, cloud services could be fine-tuned to provide different levels of data security. For instance, two clouds could be proven HIPAA-compliant, with one cloud tuned to provide lower-level security at a lower cost for data such as patients' insurance information. The other HIPAA-compliant cloud, handling sensitive health information such as patient medical histories, could be tuned for maximum security. By tailoring cloud service levels, security and pricing to the value of information handled within each cloud, organizations provisioning private clouds can buy only what they need, making the cost benefits and business case for moving into the cloud even more compelling.^[1]

“The secure cloud’s ability to map high-trust zones of systems will enable organizations and cloud providers to customize their clouds.”

Intel® TXT and Enabling Cloud Security: A Case Study

The value of Intel® Trusted Execution Technology (Intel® TXT) is in establishing this root of trust, providing the necessary underpinnings for reliable evaluation of the computing platform, and its level of protection.^[6] This root is optimally compact, extremely difficult to defeat or subvert, and allows for flexibility and extensibility to measure platform components during boot and launch environment including BIOS, OS loader, and virtual machine managers (VMMs), also known as hypervisors. In today's environment, with stringent security requirements, a system cannot blindly trust its execution environment, given the current nature of malicious threats including rootkits, hyper-jacking, and reset attacks.

Intel TXT reduces the overall attack surface for individual systems and compute pools. The principal function provided by Intel TXT is a signature, essentially a hash computation or checksum done in secure TXT memory in the processor of the launch environment, to enable a trusted software launch and to execute system software. The protection of the launch environment ensures the cloud IaaS has not been tampered with. Additionally, security policies based on a trusted platform or pool status can then be set to restrict (or allow) the deployment or redeployment of VMs and data to platforms

“Intel TXT reduces the overall attack surface for individual systems and compute pools.”

with a known security profile. Rather than relying on the detection of malware using “blacklisting” techniques, Intel TXT works by building trust into a known software environment, assuring the software being executed hasn’t been compromised—a whitelisting-based approach. This advances security to address key stealth mechanisms used to gain access to parts the data center, to access or compromise information. Intel TXT works with Intel® Virtualization Technology™ (Intel® VT) to create a trusted, isolated environment for virtual machines. Intel TXT also provides an attestable infrastructure that allows the platform to be challenged and the trust status securely disclosed.

“Knowledge truly is power.”

Secure Onboarding and Trusted Pools

Knowledge truly is power. Through the use of trust technologies, one can learn about the relative security of hosts in a cloud—which have had a verified launch and which have not. Using this knowledge, we can make better data management and security decisions—as the platforms can be segregated based on this attribute of trust, and that segregation provides opportunity. A variety of usage models can be instantiated above this secure platform.

One of the leading use models is to aggregate trusted systems and segregate them from untrusted resources—much as many IT shops seek to separate their higher-value, more sensitive workloads from their commodity supplications and data. This is the concept of *trusted pools*. Such pools allow IT to gain the benefits of the dynamic cloud environment, while still enforcing higher levels of protections for their more critical workloads. In the trusted pools model, policies are defined such that security-sensitive virtual machines can only be launched on trusted platforms or migrated to other trusted platforms. This enables the creation and maintenance of pools of platforms with trusted hypervisors, and the control of virtual machine migration across resource pools is governed by security engines plugged into the virtualization or cloud management infrastructure.

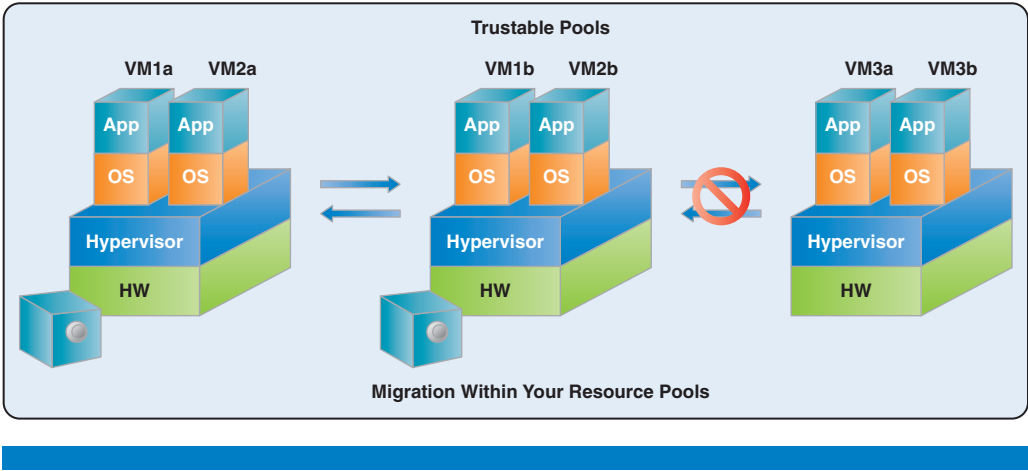


Figure 3: Migration in trusted pools^[9]
(Source: Intel Corporation, 2012)

Once trust is established at the time the hypervisor is launched, appropriate policy and compliance activities can be applied to make migration and deployment decisions, and to manage the operation and migration of workloads within the cloud. Figure 3 depicts the notion of trusted pools and the migration policy enforcement. A virtualization or cloud orchestration framework would provide the mechanisms to define the migration policy and would enforce the policy over the life time of that service running in its infrastructure.

Secure on boarding usage, as depicted in Figure 4, shows that as organizations are migrating their services to the cloud, they can request that their services are only onboarded onto servers that are “attested to” for a secure launch of the BIOS and the hypervisor environments (as indicated by the green dots). The cloud provider would provide a dashboard and a compliance report that shows that the service has always launched and run on trusted hardware.

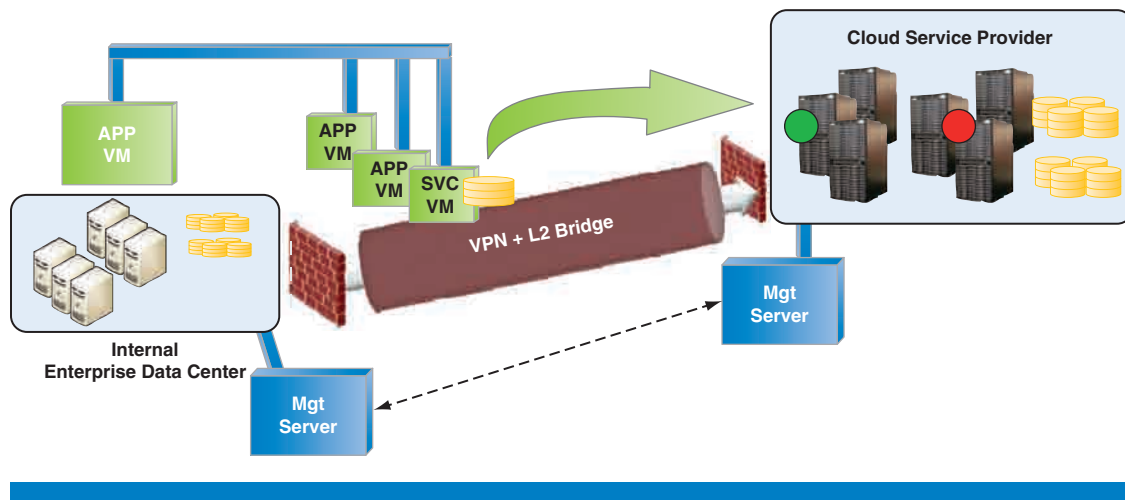


Figure 4: Secure onboarding of virtual machines to trusted clouds^[26]

(Source: Yeluri.R, and Moreau. D., “Building Trust and Compliance in the Cloud”, VMworld 2011 Technical Session.)

Cloud onboarding and cloud bursting describe the notion of either dynamically or manually deploying or migrating entire application components or parts of application (packaged as virtual machines) that normally run on internal organizational compute resources to another (internal or external) cloud to address a spike in demand, for business continuity, or for capacity optimization. The general premise of cloud bursting is to allow the cloud to act as overflow resources in the event 1) an organization’s infrastructure becomes overloaded (by demand spikes), or 2) if it is cheaper for some of the applications to be executing on the overflow capacity, or 3) for disaster recovery and failure mitigation.^[26]

Secure onboarding is complementary to trusted pools use models as they are both predicated on having visibility into the trust status of the host and making

data management decisions designed to match risk to workload sensitivity through policy enforcement and virtualization and cloud management tools.^[26]

Intel® Trusted Execution Technology (Intel® TXT)

This section provides a high level architectural overview of Intel Trusted Execution Technology, the capabilities and principle of operation, and a snapshot of the TXT Solution eco-system and its maturity.

Intel® TXT Architectural Overview

Intel TXT is a set of enhanced hardware components designed to protect sensitive information from software-based attacks. Intel TXT features include capabilities in the microprocessor, chipset, I/O subsystems, and other platform components. When coupled with an enabled operating system, hypervisor, and enabled applications, these capabilities provide confidentiality and integrity of data in the face of increasingly hostile environments.

Intel TXT incorporates a number of secure processing innovations, including:

- Trusted extensions integrated into silicon (processor and chipset). These instructions allow for the orderly quiescence of all activities on the platform such that a tamper-resistant environment is enabled for the measurement and verification process and allows for protection of platform secrets in the case of “reset” and other disruptive attacks.
- Authenticated code modules (ACM). Platform-specific code is authenticated to the chipset and executed in an isolated environment within the processor and the trusted environment (authenticated code mode) enabled by AC Modules to perform secure tasks.
- Launch control policy (LCP) tools - LCP provides the local definition of the “list” of “known good” configurations and components—this provides the foundational definition of that the platform vendor or owner will consider their trusted platform.

Some of the required components for the Intel TXT secured platform are provided by third parties, including: 1) Trusted Platform Module (TPM) 1.2 (third party silicon): A hardware device defined by the Trusted Compute Group^[13] that stores authentication credentials in platform configuration registers (PCRs), which are issued by Intel Trusted Execution Technology. 2) Intel TXT-enabled BIOS, firmware, operating system, and hypervisor environments.

Intel® TXT Capabilities

The capabilities of Intel® TXT include:

- Protected execution. Lets applications run in isolated environments so that no unauthorized software on the platform can observe or tamper with the operational information. Each of these isolated environments executes with the use of dedicated resources managed by the platform.

“Intel TXT features include capabilities in the microprocessor, chipset, I/O subsystems, and other platform components.”

“Some of the required components for the Intel TXT secured platform are provided by third parties.”

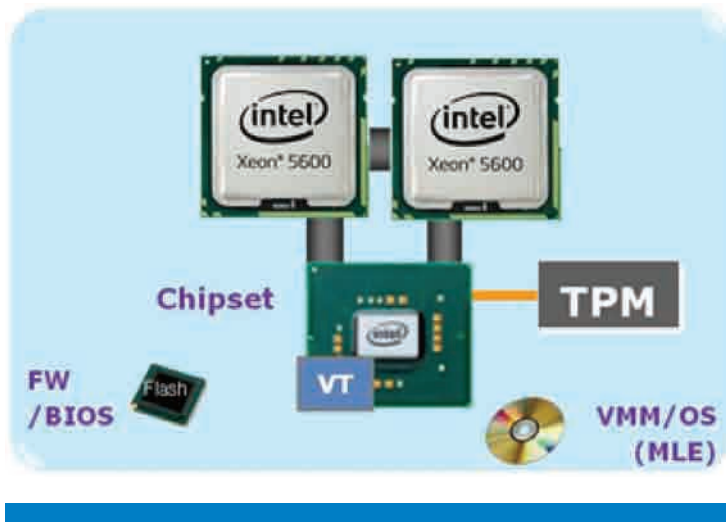


Figure 5: Intel® Trusted Execution Technology components^[9]
(Source: Intel Corporation, 2012)

- Sealed storage. Provides the ability to encrypt and store keys, data, and other sensitive information within the hardware. This can only be decrypted by the same environment that encrypted it.
- Attestation. Enables a system to provide assurance that the protected environment has been correctly invoked and to take a measurement of the software running in the protected space. The information exchanged during this process is known as the attestation identity key credential and is used to establish mutual trust between parties.
- Protected launch. Provides the controlled launch and registration of critical system software components in a protected execution environment.
- Intel® Xeon® processor 5600 series and the more recent Xeon Processor E3, Xeon Processor E7 and forthcoming Xeon Processor E5 series processors support Intel TXT, which is designed to address such software-based attacks.

Intel® TXT versus Alternatives

Historically, root of trust solutions have not been widely deployed, but interest is growing rapidly to deal with new threats. The solutions in the market have been largely limited to software-based models or static solutions rooted in a defensible component such as a BIOS boot block. On the plus side, some of these will take advantage of some hardware capabilities—particularly of the Trusted Platform Module. For such uses, the Trusted Computing Group has defined Static Root of Trust security models that set standards for implementing trust solutions. These provide some better protections and some of the same attestation and reporting capabilities that will be useful for providing integrity assurances that can also allow better data management and audit practices. The capabilities that these solutions can't provide is the processor- and chipset-enforced tamper resistance of the launch environment

“Attestation. Enables a system to provide assurance that the protected environment has been correctly invoked.”

“Intel TXT creates a cryptographically unique identifier for each approved launch-enabled component.”

and the ability to execute a trusted launch dynamically with a trust event evoked outside of the platform reset vector. While it is good that there are multiple ways to implement a solution, stronger protections against the challenge of the sophisticated adversary is a powerful advantage.

Intel® TXT: Principle of Operation

Intel TXT works through the creation of a measured launch environment (MLE) that enables an accurate comparison of all the critical elements of the launch environment against a known good source. Intel TXT creates a cryptographically unique identifier for each approved launch-enabled component and then provides a hardware-based enforcement mechanism to block the launch of the code that does not match that which is authenticated or alternately indicate when an expected trusted launch has not happened. This hardware-based solution provides the foundation on which IT administrators can build trusted platform solutions to protect against aggressive software-based attacks and to better control their virtualized or cloud environments.

Figure 6 illustrates two different scenarios. In the first, the measurements match the expected values, so the launch of the BIOS, firmware, and VMM are allowed. In the second, the system has been compromised by a root-kit

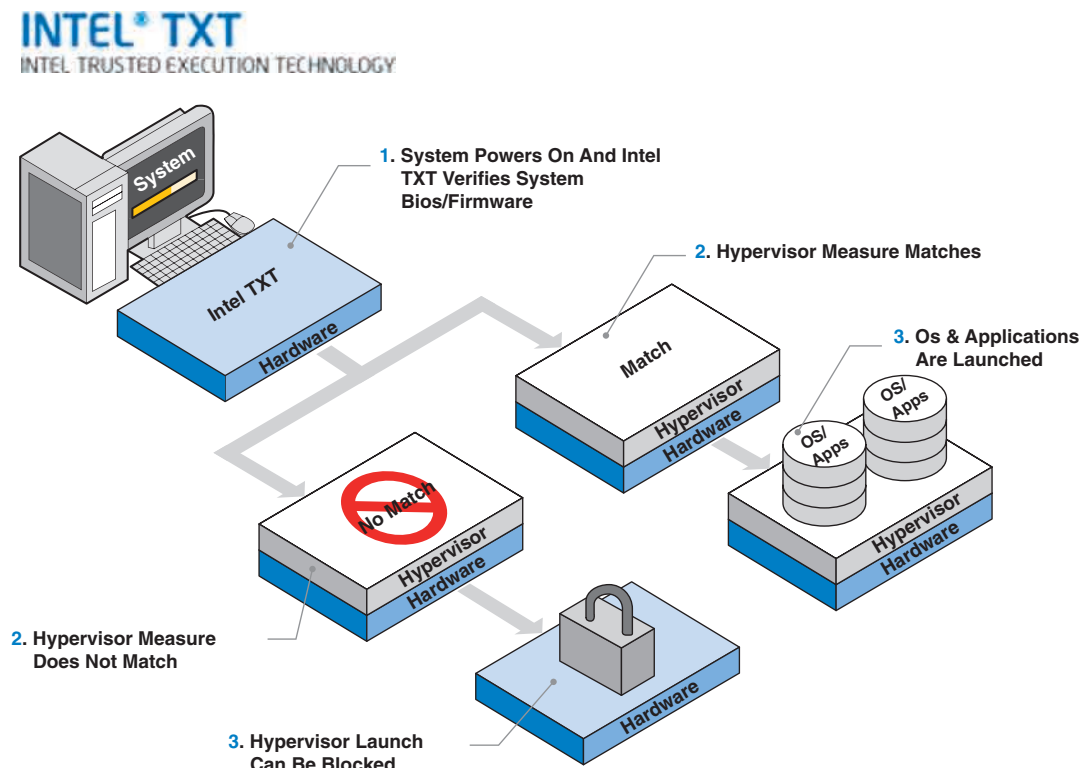


Figure 6: How Intel® Trusted Execution Technology protects the launch environment^[9]
(Source: Intel Corporation, 2012)

hypervisor, which attempts to install itself below the hypervisor to gain access to the platform. In this case, the Intel TXT-enabled, MLE-calculated hash system measurements will differ from the expected value, due to the insertion of the root-kit. Therefore, the measured environment will not match the expected value and, based on the launch policy, Intel TXT could abort the launch of the hypervisor or report an untrusted launch into the virtualization or cloud management infrastructure for subsequent use.

Intel® TXT Solution Stack and the Ecosystem Support

As you have seen from the use models above, Intel TXT provides a powerful foundation that extends deep into the cloud and across businesses, with numerous touch points. This means that it requires a great deal of software and services to take the base measurements of the hardware and firmware components of Intel TXT and allow them to be shared across organizations, to be used for implementing workload placement controls and for policy audit, monitoring and assessments. It helps to look at the ecosystem needed for the Intel TXT use models in layers, in order to best visualize the components, their roles, and the participants that typically would provide them, as shown in Figure 7.

“Intel TXT provides a powerful foundation that extends deep into the cloud.”

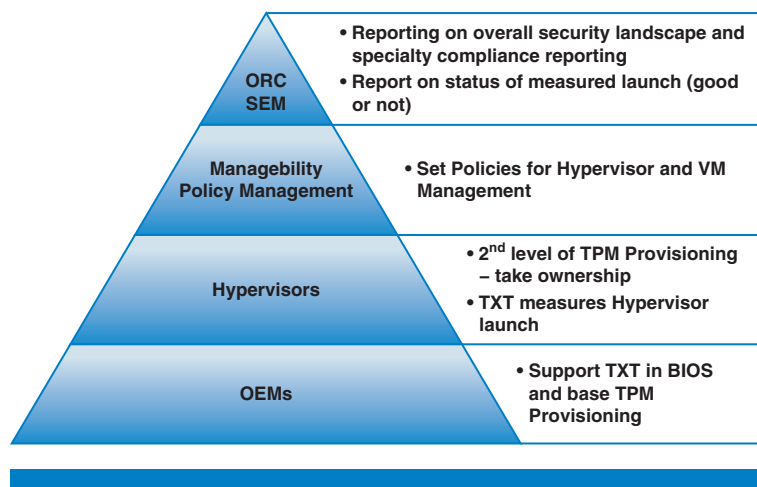


Figure 7: Intel® Trusted Execution Technology ecosystem solution stack
(Source: Intel Corporation, 2012)

The base layer of the solution stack for Intel TXT is the enables server platforms. These will be provided by a broad number of system vendors with platforms based on Intel Xeon processor family servers. Not all server systems will have the capability to support Intel TXT—pending the requisite platform components listed in the previous sections of this article—but Intel TXT enabled systems have been shipping from vendors such as Dell, NEC, and others since mid-2010 and by mid-2012 enabled systems will be quite broadly available from a wide array of suppliers. The system vendors will provide the required hardware, but also instrument their BIOS such that it can be

“Intel has maintained Intel TXT-enabled versions of open source hypervisors such as Xen and then KVM* for a number of years.”*

“Provides a first foundation for creating security-based segregation of our otherwise undifferentiated resource pools—the basis for our trusted pools.”

measured and verified as part of the chain of trust. The system vendors will also provide the base provisioning of the Trusted Platform Module (TPM)—the critical hardware component that will protect the measurements of the trust process and the mechanism for reporting this for subsequent use at upper layers of our use model solutions.

This brings us to the next critical element in the solution stack—the virtualization layer. Specifically, the key component here is the hypervisor. To play a role in our Trust use models, a hypervisor must have the capability to invoke the secure measurement process and to further provision the TPM for ownership and reporting purposes. With these functions, we gain the ability for the Hypervisor to allow itself to be measured; its trust ability to be established and the results of the trust verification process to be asserted into upper layers of the stack. Intel has maintained Intel TXT-enabled versions of open source hypervisors such as Xen* and then KVM* for a number of years, and the commercial ecosystem is coming along as well. Since 2011, VMware has been a leader in demonstrating the capabilities of a trusted hypervisor. Red Hat, SuSE, Canonical, and Enomaly have also released versions of their products that enable Intel TXT. You can expect this list to grow through 2012 and beyond as more customers demand increased trust, visibility, and control in the cloud.

With the growing interest and demand for security in virtualized and cloud architectures, it was natural that the management and policy tools that control these environments would take on the role, and these are the next layers in our solution stack. These tools will be the ones that:

- Use attestation to gather the results of platform launch processes and identify trusted or untrusted systems for subsequent use.
- Create policies that constrain access or workloads based on platform trust attributes.

As such, it is this layer that provides most of our initial visibility and control for our virtualized resources. It provides a first foundation for creating security-based segregation of our otherwise undifferentiated resource pools—the basis for our trusted pools. Here again we have seen early leaders enabling these powerful aspects of the trusted pools use models, with VMware providing capabilities early in their vCenter* console and HyTrust* providing the first trust-aware policy engine for enforcing security controls into virtualized environments. Other vendors are working to follow in the footsteps of these leaders and extend new capabilities on top of those already described.

At the peak of the ecosystem lie security-specific tools, such as governance, risk, and compliance (GRC) and security information and event management (SIEM) solutions. These tools are critical to mainstreaming trust and elements of cloud security into overall corporate security management systems. This is a crucial requirement as IT managers to not want a new suite of tools for managing cloud security, they would very much rather see existing tools extended to include the new cloud and virtualized architectures as they adopt them.

The GRC tools will be able to set requirements for platform trust and integrity based on workload requirements and security standards and then query the environment to determine the security controls in place and dashboard the actual versus policy to determine compliance. SIEM tools will allow trust events to be captured, reported, logged and processed for correlation to determine responses or heuristics to see if a larger attack is underway. RSA was one of the first ecosystem participants to seize upon the value of platform trust for GRC uses, with a joint Intel-VMware-RSA demonstration at the RSA security show in 2010. Intel is working with a number of other providers in these market segments to provide customers with an ample choice of solutions and capabilities, with availability in 2012 and beyond.

As you have probably already determined, not all use models will require tools or capabilities from each layer, though all do require enabled solutions at the OEM and hypervisor layers.

Conclusion

The use models we've discussed in this article are early-stage implementations to address requirements that customers and industry bodies are defining now. However, these models do provide a foundation for enhanced security that can evolve with new technologies from Intel and others in the hardware and software ecosystem.

There are no “silver bullets” for security, where a single technology solves all problems—security is too multifaceted for such a simplistic approach. But it is very clear that a new set of security capabilities are needed, and it is best to start at the most foundational elements. Trusted platforms provide such a foundation. Such platforms provide:

- Increased visibility into the operational state of the critical controlling software of the cloud environment through attestation capabilities; and
- A new control point, capable of identifying and enforcing local “known good” configuration of the host operating environment and reporting the resultant launch trust status to cloud and security management software for subsequent use.

Each of these capabilities complements the other as they address the joint needs for visibility and control in the cloud. Of equal importance, these attributes can be available to both consumers of cloud services and the cloud service providers, thanks to common standards for key functions such as attestation, but also due to the work for the ecosystem to enable solutions are many layers. It is only through the integration of trust-based technologies into the virtualization and security management tools in traditional IT environments (tools such as security event information management (SEIM) or governance, risk, and compliance (GRC) console) that will deliver the required scale and seamless management that will help customers realize the benefits of cloud computing.

“SIEM tools will allow trust events to be captured, reported, logged and processed.”

“There are no “silver bullets” for security.”

References

- [1] Amazon Web Services, “Overview of Security Processes,” August 2010.
- [2] Barros, A. and Kylau, U., “Service Delivery Framework—An Architectural Strategy for Next-Generation Service Delivery in Business Network, Proceedings of the 2011 Annual SRII Global Conference, pp. 47–37, 2011.
- [3] Demirkan, H., Harmon, R.R., Goul, M., “A Service Oriented Web Application Framework,” IT Professional, Vol. 13, no. 5, 15–21, 2011.
- [4] Curry S., Darbyshire J, Fisher Douglas, et al., “RSA Security Brief”, March 2010.
- [5] Cloud Security Alliance, “Security Guidance for Critical Areas of Focus in Cloud Computing v2.1,” 2009.
- [6] Cloud Security Alliance Group, “CSA-GRC Stack,” <www.cloudsecurityalliance.org/grcstack.html>, Accessed January 2012
- [7] E. Castro-Leon, E., Golden, B., Yeluri, R., Gomez, M., Sheridan, C., Creating the Infrastructure for Cloud Computing: An Essential Handbook for IT Professionals, Chapter 4, Intel Press, 2011.
- [8] “FACTA Disposal Rule goes into Effect Jun1 2005,” Report on Federal Trade Commission Website. Retrieved February 2012 from <http://www.ftc.gov/opa/2005/06/disposal.shtm>
- [9] Intel Corporation. “Intel TXT,” white paper, 2012, retrieved from <http://www.intel.com/technology/security/downloads/arch-overview.pdf>
- [10] ISACA (auditor’s perspective journal) <http://www.isaca.org/Journal/Past-Issues/2009/Volume-6/Pages/Cloud-Computing-An-Auditor-s-Perspective1.aspx>.
- [11] Kang, K., Strong, R., Fang, H., Chen, T., Rhodes, J., Zhou, R., “Complex service management in a hybrid cloud,” Proceedings of the 2011 Annual SRII Global Conference, pp. 34–46, 2011.
- [12] Khan, K.M, Malluhi, Q., Establishing Trust in Cloud Computing, IEEE IT Professional, Vol. 12-5, pp. 20–27, September/October 2010.
- [13] NIST, “Recommended Security Controls for Information Systems,” Special Publication 800-31-1, National Institute of Standards and Technology, 2005.
- [14] NIST Special Publication 800-53, “Recommended Security Controls for Federal Information Systems and Organizations,” revision3, 2010.
- [15] NIST Special Publication 800-37 Guide for Applying the Risk Management Framework to Federal Information Systems, 2010.

- [16] NIST Special Publication 800-88 Guidelines for Media Sanitization, Computer Security, 2008.
- [17] Orrin, S. Information Security And Risk Management Conference, 2011 ISACA, Session 241: Building Trust and Compliance in the Cloud.
- [18] Research paper—"Cloud Computing and Security - A Natural Match," http://www.trustedcomputinggroup.org/resources/cloud_computing_and_security_a_natural_match
- [19] Research paper—"Security Issues and Solutions in Cloud Computing" <http://wolfhalton.info/2010/06/25/security-issues-and-solutions-incloud-computing/>
- [20] Ristenpart, T., Tromer, E., et al., "Hey, You, Get Off of My Cloud: Exploring Information Leakage in Third-Party Compute Clouds," CCS'09, Chicago, Illinois, USA.
- [21] SOX 802 Final Rule: Retention of Records Relevant to Audits and Reviews. Retrieved February 2012 from <http://www.sec.gov/rules/final/33-8180.htm>
- [22] Technical Report from RAND Corporation, Europe, "Handbook of Legal Procedures of Computer and Network Misuse in EU Countries, RAND" 2005 edition of the CSIRT Legal Handbook. Retrieved February 2012 from http://www.rand.org/pubs/technical_reports/2006/RAND_TR337.pdf
- [23] Wesam Dawoud, Ibrahim Takouna and Christoph Meinel, "Infrastructure as a service security: Challenges and solutions," in 2010 The 7th International Conference on Informatics and Systems, 2010, pp. 1–8.
- [24] Wikipedia—Cloud computing security; Retrieved February 2012 from http://en.wikipedia.org/wiki/Cloud_computing_security
- [25] Yeluri, R. and Moreau, D., "Building Trust and Compliance in the Cloud," VMworld 2011 Technical Session.
- [26] Kaufman, L.M., Potter, B., Can Public-Cloud Security Meet Its Unique Challenges? IEEE Security & Privacy, July/August 2010.
- [27] Weinberg, N, Gartner: "Private Clouds Are a Last Resort," Network World, October 19, 2011 quoting Garner's Daryl Plummer; retrieved February 25, 2012 from <http://www.networkworld.com/news/2011/101911-gartner-private-cloud-252151.html>
- [28] Wenbo Mao, Attestation as a Service—Local Attestation for Cloud Security vs. Remote Attestation for Grid Security, EMC Community Network, retrieved February 25, 2012 from <https://community.emc.com/blogs/WenboMao/2010/04/25/attestation-as-a-service-local-attestation-for-cloud-security-vs-remote-attestation-for-grid-security>

- [29] Mell, P, Grance, Timothy, The NIST Definition of Cloud Computing, NIST Special Publication 800-145, September 2011.

Author Biographies

Raghu Yeluri is a Principal Engineer in the Intel Architecture Group at Intel with focus on virtualization, security and cloud architectures. He is responsible for understanding enterprise and data center needs, developing reference architectures and implementations aligned with Intel virtualization, security and cloud related platforms and technologies. Prior to this role, he has worked in various architecture and engineering management positions in systems development, focusing on service-oriented architectures in engineering analytics and information technology. He has multiple patents and publications, and has co-authored an Intel Press book on Cloud Computing – “Building the Infrastructure for Cloud Computing, an essential handbook for IT Professionals”.

James Greene is currently in a product marketing role for Security Technologies at Intel, responsible for the definition of products and usage models for datacenter and cloud security solutions. He came to Intel when the company purchased Conformative Systems in 2005. At Conformative, James led all product marketing activities for the XML processing appliance portfolio. Prior to that, James held several leadership roles in marketing, strategy, market development and business and technology planning for Compaq and Hewlett Packard enterprise server, workstation and storage business units.

Sudhir S Bangalore is a Senior Systems Engineer in Intel Architecture and Systems Integration (IASI) group, which is part of Intel Architecture Group, and is focused on developing solutions to enable virtualization and cloud security, with focus on Intel Architecture and associated ingredients. He is responsible for understanding enterprise and data center needs, developing reference implementations and innovative solutions to meet these needs with Intel technologies. Prior to this role, he has worked as a architect and a key engineer on Intel’s Enterprise Access Management framework and implementation. Sudhir has a Master’s degree in Computer Science, and has been with Intel for more than 10 years.

ON THE CONCEPT OF METADATA EXCHANGE IN CLOUD SERVICES

Contributors

Enrique Castro-Leon

Intel Architecture Group,
Intel Corporation

Mrigank Shekhar

Intel Architecture Group,
Intel Corporation

John M. Kennedy

Intel Architecture Group,
Intel Corporation

Jerry Wheeler

Intel Architecture Group,
Intel Corporation

Robert R. Harmon

School of Business Administration,
Portland State University

Javier Martinez Elicegui

Telefónica Digital,
Telefónica S.A.

Raghu Yeluri

Intel Architecture Group,
Intel Corporation

The concept of metadata is essential in a number of disciplines, including Database Architecture and Library Science. Likewise, we anticipate that the service metadata exchange, that is the standardized exchange of information describing a service between a potential service provider and subscriber during service setup and operations will also become an essential component in the establishment of a cloud service ecosystem. When an application is implemented using outsourced cloud components there is inherently less transparency about the service components when services cross organizational boundaries as in private clouds, or even more so company boundaries. This circumstance makes it difficult to implement manageability policies for a composite application made of outsourced service components. This paper reports an initial exploration of geolocation and power management examples for service metadata elements.

Introduction

The notion of metadata is central to many disciplines such as the Dewey Decimal System for cataloguing books in library and information science, the EXIF data recorded with every frame by many digital cameras.^[3] The notion of metadata is also essential to the business model of the leading social media companies like Facebook, Google, and Twitter. It can be said that the revenue from metadata is so valuable to these companies that it allows them to deliver the actual content to their user community essentially for free or at nominal cost. Reflecting this value, metadata is also at the core of intellectual property ownership and privacy controversies in the industry today. Cloud metadata constitutes an emerging field. The cloud metadata literature typically refers to metadata management for information stored in the cloud, about performance issues, replication policies and the optimal configurations for specific applications.^{[1] [2] [17]}

In the context of databases, the notion of metadata is commonly associated with information about stored data to organize the data and make it easier to search and retrieve. Metadata access and replication policies are important factors in the information retrieval performance and, when improperly architected, can become a bottleneck. Metadata is an important consideration in a company's data strategy. Metadata is essential in defining the structure of warehouse repositories aggregating information from multiple sources.

Needed in this environment are mechanisms for service metadata exchange, information about the service itself. Not much has been reported on metadata for cloud services proper. This article captures an initial exploration on the subject.

Defining Cloud Service Metadata

David Linthicum introduces the notion of metaseervices^[10] as being data about services. The analogy here is that metaseervices are descriptive of services in the same sense that metadata is descriptive of data stores. For the purposes of this work we need to take this initial definition one step beyond, more in alignment with the business process management (BPM) conceptual view described by R. Schmidt^[12] and applicable specifically to the description of cloud-based services, as illustrated in Figure 1.

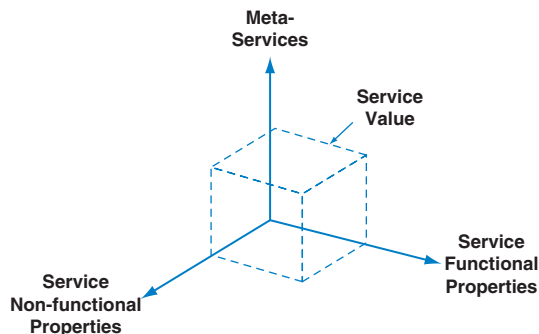


Figure 1: The three dimensions of service value
(Source: Intel Corporation 2012)

Schmidt posits that services are usually defined in terms of functional properties, such as the service's API and access methods. Yet this description is incomplete for practical implementations.^[11] There are two more necessary dimensions beyond the obvious functional capabilities that a service offers: nonfunctional properties and metaseervices, and a service's value is determined by aggregate (goodness), that is, the volume encompassed by the three dimensions as shown in Figure 1.

For instance, for most applications a cloud service would be of limited value if it does not come with a description about its reliability and security aspects. Qualitative assertions about a service's reliability and security and vague promises are not likely to satisfy service consumers. Unintended deviations from the stated reliability and security aspects are possible and because of this, the service is not complete without enforcement provisions, preferably in real time and in an automated manner. In other words any exceptions or deviations are to be handled programmatically instead of relying on manual intervention. Manual processes would be too onerous, again negating the value of the cloud. Enforcement is provided as set of actions under a metasevice. The implementation of these enforcement actions can be automatic in the form of machine-to-machine interactions, such as rebates if certain conditions are not met or manual, in the case, for example, of complaint, arbitration, or lawsuit processes. Because of the expense involved, manual processes should be reserved for truly exceptional occasions. Other actions, such as billing or even search and discovery actions can be construed as metasevice functions.

“For most applications a cloud service would be of limited value if it does not come with a description about its reliability and security aspects.”

“Schmidt claims the three axes cover the service space completely, but it is entirely possible that new aspects may appear and become relevant with practice evolution and increasing process maturity.”

“From a solution integration perspective, the transactions involved in integrating a service component are as relevant as the services rendered.”

A quintessential example of metadata is pricing, namely the pricing of a service itself. For composite services, the service provider can determine the cost of providing the service through the application of a formula combining the cost of the subsidiary services with the provider's direct costs. These calculations can be made in real time without the need to do offline accruals at the end of each billing period.

For the purposes of this discussion, we will use the term service metadata to refer to the nonfunctional aspects of a cloud service (Schmidt's axis 2). Metadata consists of nouns or declarations, essentially a set of property sheets describing the service. Metadata needs to be accompanied with a set of verbs or actions to make metadata actionable or enforceable (axis 3.) Schmidt claims the three axes cover the service space completely, but it is entirely possible that new aspects may appear and become relevant with practice evolution and increasing process maturity.

The meta-aspects are both relative and recursive. The aspects are relative in the sense that the meta- designation depends on the particular context. A service provider may decide to outsource all or portions of the meta-functions, for instance relying on a third party service provider for issuing certificates. Such certificate authority provider would be a service provider on its own right, beholden in turn to its particular metadata and metaservices. Hence someone's metadata may be someone else's main product. Metadata is recursive in the sense that a metaservice, in addition to possibly being a service on its own right, can also carry its own metadata and metaservice attributes. The manageability aspects of a service map directly to a cloud service's metadata and metaservice where manageability monitoring is captured under metadata, and the manageability control aspects are enforced under a metaservice.

A number of high-level service metadata concepts, albeit under a different name, are covered in the Open Data Center Alliance Usage: Service Catalog usage model prescription [22]. This represents a vast, yet to be explored area. The main challenges are not technical, but in achieving a common agreement and mechanisms to automate cloud service setup. These challenges will be eventually addressed through a number of commonly agreed standards. The next section covers examples of efforts in this area.

Motivation for Developing Metadata Concepts

It can be said that the main factor limiting the applicability of cloud services is the lack of maturity of the nonfunctional aspects of the cloud. It makes it difficult for solution architects to estimate the security and service level implications of the extended application. From a solution integration perspective, the transactions involved in integrating a service component are as relevant as the services rendered. By their nature, services are subject to service level agreements, whether implicit or explicit. The World Wide Web as initially conceived was an open framework for humans to access applications over the Internet. The human-to-machine interactions eventually evolved into Web services with machine-to-machine interactions over the Internet. We expect a

similar evolution for services, with labor-intensive negotiations for service setup today giving way to automated processes not too far into the future. The force driving this change is lower transaction costs^[21].

Service level agreements (SLAs) are considered to be the complete definition of the service offered by the provider and consumed by the customer, including all functional and nonfunctional parameters of the service. Without machine-readable SLAs, service consumers are forced to manually investigate and negotiate offerings from multiple service providers, and for scalability and manageability reasons service providers cannot offer personalized SLAs. Opportunities for offering better value services to customers are lost, as are opportunities to consolidate infrastructure.

This problem of metadata has been subject to extensive research past and present. SLA@SOI^[18], for example, is a European Commission funded research project dedicated to developing and proving the concept of machine-readable service level agreements. SLA@SOI has provided a solution for comprehensive SLA management supporting arbitrary service types and covering the complete service lifecycle. The project has developed a holistic reference architecture, a harmonized set of models for SLA-related management activities, and a complete reference implementation of the architecture published as open-source. The project has delivered a complete implementation of SLA management at the infrastructure layer.^[7]

Of particular interest to this discussion is perhaps the SLA model that SLA@SOI has developed. The SLA*^[6] model in particular is a generic service level agreement model that is rich, comprehensive, extensible, and format-independent. Arbitrary SLAs can be defined using domain-specific vocabularies. Developed by the SLA Management and Foundations focus area of the project, the most recent version of the model is described in that focus area's latest project deliverable^[19]. Software to support the SLA model has been published open source and is documented on the SLA@SOI SourceForge^[20] project wiki.

The SLA* model is focused on syntax. The FI-Ware project^[16], a European Commission funded public-private partnership project, is building on SLA* to develop a model that also comprehends semantics. The intent is that the resulting model can be contributed to W3C's Unified Service Description Language (USDL) Incubator Group^[13] for publication as a standard.

SLA@SOI has also played a key role in the creation of OGF's Open Cloud Computing Interface (OCCI) standard^[9]. Version 1.1, published in 2011, is a RESTful API designed to allow cloud resources to be managed: provisioned, reprovisioned, queried, and torn down. An extension has recently been proposed to support SLAs. As mentioned above, if we look at previous patterns of technology evolution, the role of metadata goes beyond service introspection and transparency. Today, even when a specific IT application is outsourced to a service provider, applications such as payroll, expense reports, or CRM, there is a lengthy and manual negotiation process upfront.

“Service level agreements (SLAs) are considered to be the complete definition of the service offered by the provider and consumed by the customer.”

“If we look at previous patterns of technology evolution, the role of metadata goes beyond service introspection and transparency.”

“The transparency brought about by metadata will make it easier to build IT applications with best-of-breed service components.”

“This new environment will foster the creation of new classes of service providers, more horizontally focused in a rich and agile application environment.”

Machine-readable SLAs and standardized APIs will provide a means to automate much, if not all, of this process.

These applications are treated as unbreakable monoliths precisely because it is very difficult to infer the behaviors of the whole based on the characteristics of the individual components. The availability of metadata will allow making predictable QoS assessments for a whole application based on the analysis of the metadata of the service components. The transparency brought about by metadata will make it easier to build IT applications with best-of-breed service components, likely smaller and more fungible than the fully fledged services of today. In particular, cloud services functioning as building blocks to other services are instances of servicelets^[8]. It will be possible to fine tune the delivered QoS through the manipulation of the individual servicelets. The assumption is that the means will exist to estimate the delivered QoS in a provable way, that is, through the application of mathematical formulas using servicelet metadata.

For instance a solution architect for a provider may determine that in order to reach the advertised QoS the backend storage availability needs to be increased from 98 percent to 99 percent. An in-house solution might require reengineering the storage subsystem in the company-owned data center. A more agile, service-oriented solution might involve leasing additional storage capacity, perhaps from another storage provider and deploying a mirror copy of the data, or perhaps switching to another provider altogether with a different and more stringent QoS profile. All these assessments would be done on the basis of available metadata. If the consumer is unsure about the validity of the provider metadata, it could sign up the services of a third-party provider in the business of monitoring storage QoS. This is an example of a metaservice.

In principle this process is not much different than building a portfolio of investments in finance with a target set of behaviors by integrating a set of instruments with desired characteristics and aided by the services of ratings companies. This capability is a function of technology maturity. We can expect machine-to-machine negotiations to become more prevalent. Negotiations that used to require human-to-human interactions between service providers and consumers with lengthy and protracted negotiations taking months to close will morph to humans interacting with portals, with time constants shrinking to days, and eventually machine-to-machine, with transactions closed in a matter of minutes.

This new environment will foster the creation of new classes of service providers, more horizontally focused in a rich and agile application environment. One aspect that can't be emphasized enough is that metadata is money, with significant strategic implications to suppliers and consumers alike. It enables Google, Amazon, Facebook, Netflix, and Apple services. The economic benefit of getting a hold on metadata explains the viability and success of these companies to the extent that they can offer a number of products for free or at much lower prices than would otherwise be possible.

The Dynamic Enterprise Perimeter for Cloud-Based Applications

The boundary of a set of IT resources under which a common metadata and metaservice applies is said to define an enterprise perimeter. This concept was introduced by Dournaee for security analysis, but it is actually applicable to any aspect of manageability^[4]. Within the enterprise perimeter a set of assertions (meta-assertions) holds true. One example assertion would be: “the servers inside this perimeter support Intel® Node Manager technology.” The presence of this capability would enable the carrying out of common power management policies within the perimeter for the precise optimization of power and energy consumption.

Now assume that the initial set of IT resources is augmented with outsourced resources through a cloud bursting arrangement. If principles of software engineering were to be applied to this situation toward the goal of limitless scalability, the augmented resources would be transparent to the original set so the same methods and operational procedures apply uniformly. If this condition is met, the net effect of cloud bursting would be that of expanding the initial enterprise perimeter into a virtual enterprise perimeter (dynamic enterprise perimeter in Dournaee’s definition) encompassing the original and the expanded resources.^[4]

For this scenario to take place, this would require that the metafeatures of a service, namely the metadata and metaservices, be a compatible superset of the metafeatures in the internal resources.

A metadata attribute likely to attain prominence in the near future is that of *geolocation*, that is the ability of a service provider to attest to the location of its resources, both direct resources *and* the location of any delegated resources. Workload geolocation compliance is an emerging requirement for cloud service providers. Lack of trust due to the absence of geolocation metadata in a service offering can represent a barrier for the acceptance of the service, forcing operators to run these functions in house at a possibly higher cost. As an example, a FISMA-regulated workload cannot leave the US. Publications released from NIST provide explicit geolocation compliance guidelines for the federal government and information technology industry. There are similar requirements from Europe defined by the European Network and Information Security Agency (ENISA) and the Directive 95/46/EC, specifying restrictions on the movement of personal information outside the European Economic Area, including healthcare information.

Because the metafeatures are predetermined by service providers and these rarely match and exceed the service consumer requirements, establishing true virtual enterprise perimeters is not feasible, and hence the norm today is that the use of cloud bursting implies a degradation of metafeatures in the form of degraded QoS or security. Unfortunately, even if the features are present in the servers used, there is no common method to convey these capabilities, let alone consume them. The lack of transparency in cloud services relative to QoS

“The boundary of a set of IT resources under which a common metadata and metaservice applies is said to define an enterprise perimeter.”

“Workload geolocation compliance is an emerging requirement for cloud service providers.”

“Breaking up resources into granular components during the service setup would allow a potential service consumer to discover, request and allocate components a la carte to assemble a service fitting a target metafeature profile.”

“Simple contracts can be negotiated directly between the provider and the consumer.”

and security represents a commonly quoted roadblock today toward a wider adoption of cloud computing.

A piecewise approach represents a possible solution to this conundrum given that it is not feasible for service providers to implement the superset of features that would satisfy all present and future customers. Breaking up resources into granular components during the service setup would allow a potential service consumer to discover, request and allocate components a la carte to assemble a service fitting a target metafeature profile. In the example above, a potential service consumer processing electronic health records may be obligated to in-source the capability, forgoing the cost advantages of cloud computing because of the lack of providers with a capability to pass on service geolocation metadata.

A bursting environment allowing the service requester to transparently extend current operational policies to leased resources enabled by a metadata exchange capability would bring significant operational advantages. One advantage would be economies of scale from uniformly carrying out operational procedures to both internal and external resources. It puts the service consumer in the driver's seat and builds credibility and trust for the service provider. Because of this the service consumer can quantify the risks more accurately, and in so doing, becomes less demanding and dependent on the assurances of the service provider. It's a lot easier for the service provider to offer assurances enabling the consumer to enforce certain policies than it is to offer absolute guarantees about service features.

The cloud challenges the traditional notion of a security perimeter. Perhaps the shortcomings of firewalls have been present all along, but it's only recently that it's become evident how woefully insufficient this approach is. The availability of security-related metadata can help by bringing in a quantitative, data-driven approach to system design and operations.

Metadata in Cloud Service Architecture

Augmenting an existing set of resources with cloudburst resources requires the execution of a carefully choreographed dance starting with a discovery process. If the provider features look promising a detailed metadata exchange takes place.

If the set of metafeatures matches the service consumer requirements and the requester decides to proceed, a contract negotiation ensues, followed by configuration requests and culminating with the binding of new resources with the original resource set. At this point the new resources become available to the consumer.

Simple contracts can be negotiated directly between the provider and the consumer. More complex contracts may require the intervention of neutral third parties functioning as brokers and providing escrow services.

These negotiations are normally carried out through metaselects. In particular these negotiations underscore the need for metadata exchange capabilities for practical service transactions. The metadata exchange proper can take any of multiple possible forms. It can be as simple as exchanging an XML formatted message with the service description, or it can be brokered through the third parties with audit capabilities ensuring that QoS assurances are met and independently verified.

Continuing the geolocation example, users of cloud computing deploy VMs on a cloud provider's infrastructure without having to maintain the hardware. Differences in laws governing issues such as privacy, information discovery, compliance, and audit require that some cloud users restrict VM locations to certain jurisdictions or countries.

In a cloud environment services are more likely than not to be composites of more primitive services, that is, composites of servicelets mentioned in the earlier section, "Motivation for Developing Metadata Concepts." Service providers can elect to pass through metadata from constituent services. The pass through may not be supported if the provider retains this information as a trade secret or to replace presumably fungible servicelets without notice. Service consumers wishing for more transparency and control may push back or refuse to do business with providers withholding metadata from critical components.

Beyond the geolocation example, metadata can cover almost any item: if it's about a storage service, it may include availability, reliability, power, and energy consumption statistics. If it's about processing resources, it might also provide information on whether a resource is exclusive or subject to multitenancy, or whether it is physical or virtualized. The metadata exchange can be a two-way process when the discovery process is included. This would help enforcing locality requirements or requirements for exclusive use of resources.

One initial benefit of a metadata exchange capability is the implementation of a "certificate of origin" or brand authentication capability for constituent services as part of the setup handshake allowing the prospective service consumer to determine the pedigree of constituent servicelets, not just at the server level, which would be closest to the hardware, but for any constituent service.

The attestation of the certificate of origin can be implemented as a metaselect from the third-party broker or by using the hardware root-of-trust capability implemented by the Intel® Trusted Execution Technology (Intel TXT) or both. A service consumer looking for a service featuring a third-party storage servicelet can retrieve the certificate of origin for the storage servicelet of the particular service and verify that it is in effect provided by that third party. The metadata exchange process allows service consumers to realize the brand promise of any of the constituent servicelets.

"In a cloud environment services are more likely than not to be composites of more primitive services."

"The metadata exchange process allows service consumers to realize the brand promise of any of the constituent servicelets."

“The current state of affairs is the result of the desire of providers wanting to have it both ways, leading essentially to a stalemate situation.”

“The provider will ensure that certain virtual machines stay in the specified geographic region or predefined enterprise perimeter.”

Metadata Security

Blogger Chris Wolf has lamented that cloud providers, such as Terremark (now part of Verizon), IT Structures, EngineYard, and Logica emphasize their easy-to-use management model, but none has come forward to explain how to address regulatory and security compliance^{[14][15]}. He further states that this type of compliance remains a primary barrier for enterprise cloud adoption. He proposes the following tiered-security model for cloud service providers:

- Level A: Dedicated physical and virtual infrastructure, including dedicated server and networked storage assets.
- Level B: Dedicated virtual and physical server infrastructure, shared/ logically zoned storage infrastructure (clients receive dedicated LUNs, but data traverses a shared physical SAN).
- Level C: Shared virtual and physical infrastructure, isolation provided by dedicated virtual security appliances (such as VM firewalls, IDS, IPS).
- Level D: Shared virtual and physical infrastructure, no appliance-based segmentation or isolation (isolation provided via VLANs).

This tiered model can pass muster with enterprise security auditors, paving the way for a broader enterprise adoption of cloud services. Unfortunately, the likely reason for the silence coming from providers is implementation cost.

That’s because the current offerings are essentially monolithic. A Level A provider would be at a cost disadvantage relative to a Level D provider. The current state of affairs is the result of the desire of providers wanting to have it both ways, leading essentially to a stalemate situation.

The piecewise approach mentioned above with transparency through metadata availability would offer a way out of this apparent contradiction. A service consumer wishing to assemble an application with Level A resources would search for physical resources knowing that there is a price premium for exclusive use of physical resources, whereas a cost-driven service consumer would host their application on VMs running on a shared physical host. The hardware for the physical hosts can be the same, but the usage models are different. This difference is reflected in the service metadata.

Geolocation as Service Metadata

Elaborating on the geolocation example introduced above, this section elaborates one possible method for the setup of a service with geolocation as metadata capability.

An organization processing geographically restricted data in the cloud will look for a service provider capable of attesting the location of the resources used as part of a service level agreement (SLA) to be negotiated. As part of the service guarantees, the provider will ensure that certain virtual machines stay in the specified geographic region or predefined enterprise perimeter.

Also, these cloud subscribers and their customers will need certifiable mechanisms to conduct audit and verify geolocation information at runtime to ensure that the location restrictions in their cloud SLAs are being met.

A typical service setup exchange is as follows:

- Service subscriber A looks in a service directory B, possibly through a broker for an IaaS instance where the geolocation attribute is present.
- B has done prior introspection on the metadata retrieved from each of the services in its repository and catalogued the geolocation metadata capability.
- Service consumer A initiates a cloud bursting request to IaaS provider C for N servers and requests the service metadata for the negotiated package. The metadata is essentially a dictionary containing a description of the location of each of the servers in the service package. These capabilities include information such as the platform unique ID and IP addresses for the servers as well as the actual APIs needed to carry out geolocation attestation queries.
- The cloudburst server resources get linked up to the service consumer through a Layer 2 tunnel, and after the setup transaction is complete, the new contingent of servers may be placed in the same subnet as the local servers, all compliant with the geolocation requirements.

Security is an inseparable consideration with metadata exchange. Because of the profit considerations involved and the ease with which compliant servicelets can be replaced with cheaper substitutes such as servers outside the predefined enterprise perimeter, temptations exist for unscrupulous operators to hide information about the true location of the servers involved. Likewise, cloud providers may attempt to move customer VMs to cheaper data center locations in violation of the negotiated SLAs. Under these circumstances, physical asset isolation and remote attestation are critical for high security data center facilities.

Asset management is another aspect requiring geolocation information in almost all data centers. Small but significant numbers of dead or “zombie” servers become part of the cost of doing business. Zombies are powered servers connected to the network but not doing any useful work. Once these servers are singled out and their location verified, they can be decommissioned.

Managing and planning data center capacity often requires determining location of the available capacity. Establishing a baseline for processing, networking, and storage capacity is required for long-term capacity planning. The ability to schedule resources server resource peaks by making capacity available in an agreed-upon advance notice is a common practice in a cloud environment.

Data center airflow and cooling costs can be optimized if a thermal profile can be accurately created. Without precise server location, creating a meaningful data center thermal profile is very difficult. The ability to locate a server within

“Security is an inseparable consideration with metadata exchange.”

“Zombies are powered servers connected to the network but not doing any useful work.”

“In practice carrying an inventory is costly in terms of the labor involved and can’t be done casually.”

“Once management across local and remote nodes has been demonstrated to work the cloud bursting setup will be re-implemented on top of the OpenStack framework.”

a pod, row, and aisle, as well as building information, is critical for workload placement for power and thermal-aware workload balancing.

Server location in a data center is maintained in a configuration management database, or CMDB. If updates to the CMDB are done manually, errors unavoidably creep in. In theory it is possible to generate a new baseline by conducting an inventory. In practice carrying an inventory is costly in terms of the labor involved and can’t be done casually.

Intel/Telefónica Cloud Bursting and Multi-Cloud Power Management Project

Intel Corporation and Telefónica Digital initiated a joint project to build a proof point of the concepts described in the previous section. The project is still under execution at the time of this writing. There are three main phases for this project:

- Demonstrate the feasibility of cloud bursting across two companies. An application running on servers in Folsom, California requests external resources, namely servers hosted by Telefónica running in Madrid. A carefully choreographed *pas de deux* ensues, facilitated by a cloud fabric implemented under Citrix Netscaler* VPX. When the dance completes, the remote Telefónica servers (“nodes”) end up in the same subnet as the initiating servers in Folsom, California.
- Carry power monitoring and power limiting policies over the network using the IPMI protocol. Since all servers are in the same subnet, the execution of these policies should be transparent. Once management across local and remote nodes has been demonstrated to work the cloud bursting setup will be re-implemented on top of the OpenStack framework, specifically through the use of the OpenStack Compute Nova API, extended as appropriate, to support machine-to-machine negotiation. The auto-negotiation will also encompass the metadata exchange that would allow a cloud bursting initiator to retrieve information about the power management capabilities of the servers hosted by the service provider (Telefónica), including the endpoint of the API to carry out power management operations, such as real-time power consumption monitoring and establishing power usage limits. A security layer, using Intel® Trusted Execution Technology, may be added to implement a metadata attestation mechanism.
- All of the handshake processes above will be documented and submitted as a reference implementation or standardized usage model to the OpenStack community and any appropriate industry-recognized standards body.

In the remainder of this section we provide a detailed description of the cloud bursting setup orchestration. In the initial state, as shown in Figure 2, we see two resources (servers) functioning independently. The Intel data center hosts an application capable of cloud bursting (that is, Intel is the service consumer or initiator), whereas the Telefónica data center offers hardware for rent as a traditional cloud IaaS service.

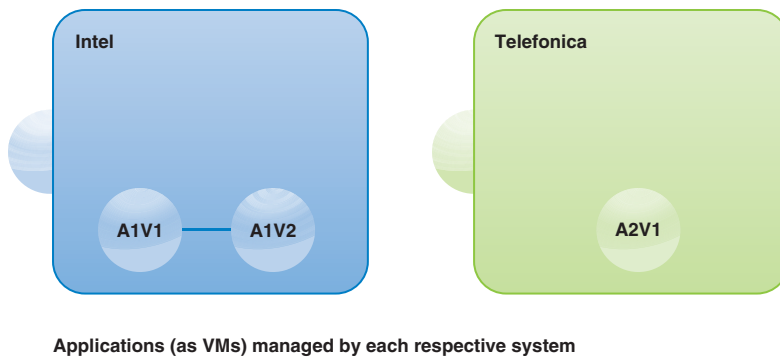


Figure 2: Initial state before cloud bursting
(Source: Intel 2012)

Now assume that customer A1, currently running two virtual machines in the Intel host, requests that a third virtual machine be instantiated, as shown in Figure 3. Furthermore assume that under current policies there is no more room for virtual machines at the Intel host.

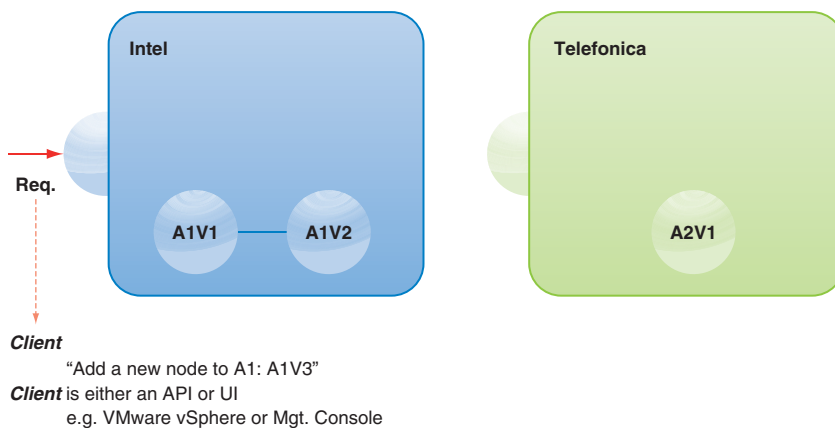


Figure 3: The application receives a request to add a third virtual machine
(Source: Intel 2012)

The infrastructure manager orchestrating resources for the Intel application, which can be a service in its own right, makes a decision to outsource the hosting of the new virtual machine to an external IaaS provider and queries a resource broker, essentially a directory service. The broker returns a URL for the extended resource, a handle for the Telefónica data center provisioning API. The requester uses the handle to initiate a process to qualify and bind the Telefónica service to the application.

“As a matter of policy the application is compelled to compute the carbon footprint for all resources it touches.”

At a higher level this orchestration can be accomplished through the OpenStack or the OCCI frameworks. Figure 4 shows some of the OpenStack Compute API (Nova API) calls. At a lower level these calls get mapped into Citrix Netscaler operations.

Assume that as a matter of policy the application is compelled to compute the carbon footprint for all resources it touches, and through the Nova API it retrieves the Telefónica host metadata dictionary. After some parsing it extracts the power management object, listing the server’s power management features. In our proof-of-concept this will confirm the presence of Intel® Node Manager Technology and the access methods (API) to elicit these features. At this point the API can be bound to the power management policies already in effect, allowing operations such as calculating the energy expended by the operations performed on the Telefónica node.

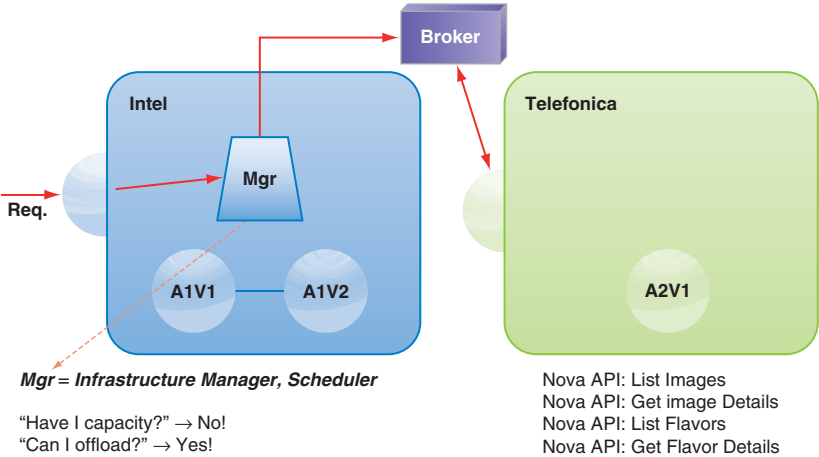


Figure 4: Current policy prevents new VM from being created at the Intel host. The Infrastructure Manager queries a broker for an external resource for a cloud bursting target.
(Source: Intel 2012)

The final configuration after the automated setup is shown in Figure 5.

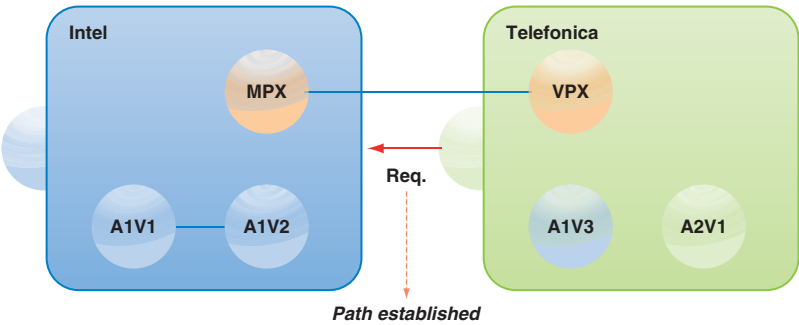


Figure 5: Physical host and virtual machine bound to originating service
(Source: Intel 2012)

When a number of services are connected in this manner to build a composite application, the result is a multi-cloud, and hence the moniker *multi-cloud management*.

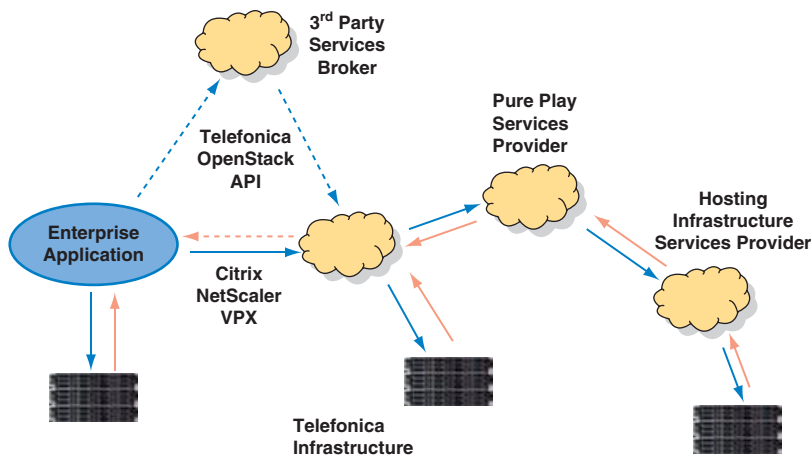


Figure 6: A Multi-Cloud Composite Application.

(Source: Intel 2012)

The multi-cloud handshaking process happens in this manner: An enterprise application monitoring the application's performance senses in-house resources running short of an anticipated peak demand and goes through a third-party broker to identify a cloud provider to land the extra demand. The broker recommends the Telefónica service and goes through the provider/consumer setup described earlier. The choreography of this setup is fairly complex, facilitated by Citrix NetScaler VPX.

In order to manage risk, Telefónica also balances in-house and subsidiary cloud providers. Some of these providers may be pure play providers, relying on IaaS providers for their infrastructure needs as shown in Figure 6.

For each of the recursive provider/consumer relationships represented by the blue arrows in Figure 6, there is metadata flowing in the opposite direction, represented by the red arrows.

The original enterprise cloud bursting requester now has the option of requesting exclusive access to the hardware and can manage this hardware as if it were a local resource if the application demands it. Beyond that, the metadata exchange mechanism can in principle perform introspection on the service to identify all the intervening services to verify their reputation. Conversely, any service provider can advertise their presence and leverage their brand name.

The “in principle” is an important qualifier. How much metadata is exposed is very much a business decision among the parties involved. Some providers will make their particular combination of services their secret sauce and choose not to disclose it, where others may decide that disclosing a complete portfolio is in

“How much metadata is exposed is very much a business decision among the parties involved.”

“It is very likely that the metadata revolution that took place in consumer space social networks will take place in enterprise space in the next few years.”

“Metadata exchange depends on mutually agreed protocols between the contracting parties in order to work.”

their interest. Conversely, some service providers who want to make themselves known may require providers upstream to provide their identity as a condition for doing business.

The OpenStack effort has resulted in potential enhancements to OpenStack being identified, and in several research areas being identified for future collaboration. Further investigation is ongoing to identify a suitably scalable approach for secure metadata exchange that might be of interest to the OpenStack community.

Conclusions

Economics is driving the adoption of the cloud, and within this context, the authors believe that economics will also drive the need for metadata. It is very likely that the metadata revolution that took place in consumer space social networks will take place in enterprise space in the next few years.

With the status quo and a one-size-fits-all strategy, a service product would be overprovisioned or too complex for customers with the lowest SLA requirements, resulting in higher OpEx, and there would be opportunity costs: customers not signing up because a service offering does not meet their requirements profile, being either under- or over-specified. The service metrics enabled by the availability of service metadata reduces uncertainty and adds predictability to service transactions, increasing the perceived value from the perspective of the service consumer, and enabling the service provider to offer more targeted, differentiated products.

The geolocation example described in this article is only one example of new cloud service capabilities enabled by service metadata exchange. We will be reporting the outcomes of actual experiments in future publications.

Another area that requires additional exploration is the security aspects of metadata exchange. Since the ability to advertise certain capabilities may determine whether or not a specific service gets hired, these decisions will have a revenue impact on service operators. Therefore the temptation exists for unscrupulous operators to alter metadata records.

Hence it is necessary to implement methods to deter this potential tampering. An example could be service spoofing: a service operator may claim to be using a storage service from a highly regarded provider, when in reality the operator vectors storage functionality to a lower cost and presumably less reliable provider. Service metadata brings a needed degree of transparency in cloud transactions, a first step toward bringing predictable QoS that in turn will enable enforceable SLAs. This degree of unpredictability has been one of the main barriers toward cloud adoption in the industry.

Metadata exchange depends on mutually agreed protocols between the contracting parties in order to work. To this end we are actively working with open source projects and open standards organizations to introduce and showcase such support into appropriate points in the cloud computing stack.

The work reported here represents only the beginning of what we believe is a rich area for research, very relevant to cloud computing because of the revenue and cost implications to service providers and consumers alike. Platform power management features are being used in this project as a proof of feasibility for the metadata exchange.

Support for metadata exchange is an optional component in a service offering. Service providers will support the capability if it's in their interest, for instance if it helps in building a reputation with positive revenue impact. In any case, when a service provider chooses to expose certain metadata, the service consumer needs to have a reasonable assurance that the information is truthful.

There are also technical motivations for metadata adoption: this framework can be extended into a general framework for a two-way auto-negotiation where a service consumer queries or imposes certain requirements on the service provider, for instance the type of hypervisor to be used in a service request.

Areas to be further explored in the future are the concept of a metadata management infrastructure, whether brokered by third parties or as two-party exchanges such as the recursive aspects of metadata exchange and metadata filtering. We will report on these discoveries as data becomes available.

References

- [1] Ganesan Shankaranarayanan, Adir Even, "The Metadata Enigma," Comm. ACM, Vol. 49-2, February 2006.
- [2] J. Wang, P. Varman, C. Xie, "Middleware Enabled Data Sharing on Cloud Storage Services," Proc. ACM MW4SOC'10, Bangalore, November 29, 2010.
- [3] Wikipedia, Metadata, Retrieved February 5, 2012 from <http://en.wikipedia.org/wiki/Metadata>.
- [4] Dournae, B., "Taking Control of the Cloud for your Enterprise," Intel Corporation white paper, 2010. Retrieved from http://info.intel.com/rs/intel/images/Taking_Control_of_the_Cloud.pdf.
- [5] Harmon, R. R., Auseklis, N., "Sustainable IT Services: Assessing the Impact of Green Computing Practices," PICMET 2009 Proceedings, IEEE, pp. 1707-1717, 2009.
- [6] Kearney, K.T.; Torelli, F.; Kotsokalis, C.; "SLA★: An abstract syntax for Service Level Agreements, Grid Computing (GRID)," 2010 11th IEEE/ACM International Conference, pp. 217–224, 2010, http://www.sla4d-grid.de/sites/default/files/SLAWS_Grid-2010_SLAatSOI_Model.pdf
- [7] Kennedy, J; Edmonds, A., Bayon, V., Cheevers, P. Lu, K., Stopar, M., Murn, D., "SLA-Enabled Infrastructure Management," Chapter 7 in *Service Level Agreements for Cloud Computing*, Springer, 2011.

- [8] E. Castro-Leon, E., He, J., Chang, M., Peiravi, P., *The Business Value of Virtual Service Oriented Grids*, Chapter 2, Intel Press, 2008.
- [9] Open Cloud Computing Interface, Retrieved February 20, 2012 from <http://occi-wg.org/>.
- [10] Linthicum, D.S., *Cloud Computing and SOA Convergence in Your Enterprise, A Step-by-Step Guide*, Addison-Wesley, 2010.
- [11] Schmidt, R., "Perspectives for Moving Business Processes into the Cloud," I. Bider et al. (Eds.), BPMDS 2010 and EMMSAD 2010, pp. 49–61, Springer-Verlag Berlin Heidelberg 2010.
- [12] Schmidt, R., "A Service-System Based Identification of Meta-services for the Service-Oriented Enterprise Architecture," Proc. 15th IEEE Int. Enterprise Distributed Object Computing Conference, pp. 293–300, 2011.
- [13] The W3C Incubator Activity. Retrieved February 20, 2012 from <http://www.w3.org/2005/Incubator/usdl/>.
- [14] Wolf, C., "Security and Compliance in the Cloud," 2009. Retrieved February 5, 2012 from <http://www.chriswolf.com/?p=230>
- [15] Wolf, C., "The Cloud Mystery Machine: Metadata Standards," 2010. Retrieved February 12, 2012 from <http://www.chriswolf.com/?p=490>
- [16] "The Future Internet Core Platform." Retrieved February 20, 2012 from <http://www.fi-ware.eu/>.
- [17] Verma, A., Venkataraman, S., Caesar, Campbell, R., "Efficient Metadata Management for Cloud Computing Applications," Technical Report, University of Illinois, 2010. Retrieved from <https://www.ideals.illinois.edu/handle/2142/14820>
- [18] SLA@SOI Project Web Site in the European Union. Retrieved February 20, 2012 from <http://www.sla-at-soi.eu>.
- [19] Foundations for SLA Management – Evaluated Framework, SCLA@SOI Deliverable D.A5a, Retrieved February 20, 2012 from <http://sla-at-soi.eu/wp-content/uploads/2011/08/D.A5a-M38-SLA-Foundations-and-Management.pdf>.
- [20] A Collection of Resources Relating to the SLA Model, SLA@SOI SourceForge Project Wiki. Retrieved February 20, 2012 from <http://sourceforge.net/apps/trac/sla-at-soi/wiki/SlaModel>.
- [21] Coase, R. H., "The Nature of the Firm," *Economica*, Vol. 4 No. 16. (Nov. 1937), pp. 386-405.
- [22] Open Data Center Alliance Usage: Service Catalog, Open Data Center Alliance (2011)

Authors' Biographies

Enrique Castro-Leon is an enterprise architect and technology strategist with Intel Corporation working on technology integration for highly efficient virtualized cloud data centers to emerging usage models for cloud computing.

He is the lead author of two books, *The Business Value of Virtual Service Grids: Strategic Insights for Enterprise Decision Makers* and *Creating the Infrastructure for Cloud Computing: An Essential Handbook for IT Professionals*.

He holds a BSEE degree from the University of Costa Rica, and M.S. degrees in Electrical Engineering and Computer Science, and a Ph.D. in Electrical Engineering from Purdue University.

Mrigank Shekhar is an enterprise architect with Intel Corporation working on platform solution architecture for highly efficient cloud data centers for cloud computing. He has also worked for Siemens and Biotronik where he led the research and design team responsible for implantable defibrillator products. He holds several patents in medical electronics and server architecture. He also successfully founded a point-of-sale company. He holds a Master's degree in Computer Engineering from the University of Southern California, Los Angeles.

John Kennedy is a senior researcher in the Cloud and Services Lab of Intel Labs Europe. He is focusing on open standards and open source enablers to facilitate more dependable cloud computing. John is particularly active in Intel's European research initiatives, engaging in collaborative research at local, national, and international levels. He has a B. Eng (hons) in Electronic Engineering and M. Sc. in Computing. John joined Intel in 1997 and has held roles of systems analyst, developer, solution architect, and applied researcher in Intel's Manufacturing, IT, and Intel Labs organizations.

Robert R. Harmon is Professor of Marketing and Technology Management and Cameron Research Fellow in the School of Business at Portland State University. His research is focused on service innovation, sustainable IT services, and the strategic migration of manufacturing companies to service enterprise business models. His research has been funded by the National Science Foundation, Intel Corporation, IBM, Tata Consultancy Services, the Semiconductor Industry Association (SIA), and Semiconductor Equipment and Materials International (SEMI), among others. He has a Ph.D. in marketing and information systems from Arizona State University.

Raghu Yeluri is a Principal Engineer with Intel Architecture Group at Intel with focus on cloud and security architectures. Raghu has worked with Intel for 15+ years with roles in both Information Technology and Systems Architecture, and has a MS in Computer Science and a BS in Electrical Engineering.

Francisco Javier Martínez Elicegui holds degrees in Computer Science from the Technical University of Madrid, Master in Expert Systems from the University of Salamanca and he is currently carrying Ph.D. research on Auditory Computation & Psychacoustics. Javier is a System Business Consultant at Telefónica Digital with an extensive computer engineer experience in several industrial sectors including manufacturing, bank, space industry, public administration and telecommunications. His areas of expertise include GUI Interfaces, telecommunication infrastructure management, Business Intelligence, Service Oriented Architecture, Semantic technologies, Mobile Financial Services and Cloud Computing. He is reviewer for the Distributed Computing and Artificial Intelligence annual congress (DCAI) and he holds publications in different international congresses.

OPENSTACK AND INTEL TECHNOLOGY BEST PRACTICE

Contributors

Billy Cox

Software and Services Group,
Intel Corporation

Jiangang Duan

Software and Services Group,
Intel Corporation

Fleming Feng

Software and Services Group,
Intel Corporation

Das Kamhout

IT, Intel Corporation

Susie Li

Software and Services Group,
Intel Corporation

Tonny Liu

IT, Intel Corporation

As IT moves into the era of “new IT”, cloud and related technologies are moving quickly. Many factors including cost constraints and consumerization of IT are driving these changes. Open source is leading the innovation tide and playing an indispensable role in shaping the future of new IT. OpenStack, as an emerging open source software stack for creating and managing clouds, has gained significant interest from both academia and industry leaders. In this article we illustrate Intel IT’s OpenCloud project, which implemented an open source private cloud in a production environment based on OpenStack. Furthermore, we present how Intel® Node Manager and Intel® Trusted Execution Technology can help improve the power efficiency and security of a cloud environment.

Introduction

The “new” IT continues to emerge from the shadows of traditional IT. A few challenges are pushing IT to not only consider changes, but to do so quickly. The forces include reducing IT cost while significantly increasing agility, improving security and related controls, and the consumerization of IT.

The pressure to reduce the cost of IT is not new. But now, in addition to being more efficient, IT is being asked to be a key partner in driving the business forward. This role places IT squarely in the position to not just host content and applications, but to show the business how to use modern tools to advance the business.

Security remains a top concern for the use of clouds, especially public multitenant clouds. Security includes things like access control and logging, encryption, single sign-on, and trust.

Consumerization is a recent trend often associated with “bring your own.” In reality, the trend is much more than just, “bring your own,” because the new users of IT simply have different expectations. They look at tools like Dropbox* as compared to traditional enterprise sharing and expect the simple functionality of Dropbox. They expect to use a tool and see immediate results, as well as rapid updates. And as the industry innovation continues, there is an expectation that IT will keep up.

These are fundamental forces driving IT to change—not merely adjust.

While traditional IT would have delivered email, new IT is expected to also deliver full collaboration with large files and cross-group editing. Traditional IT delivers each user a laptop; new IT empowers the employee to use whatever

device they choose. Traditional IT would provide a structured database; new IT provides large scale means to analyze unstructured data.

The quest to develop solutions for the new IT have involved numerous players in many parts of the industry and have demanded innovations across the field. In many ways, open source is leading the way delivering key technologies. As an example, the well-known use of Xen in the Amazon Web Services public cloud would not have been possible without the hard work of a dedicated community of contributors.

OpenStack is a new and emerging open source software stack for creating and managing clouds. At its heart, it is about building an open, extensible, framework for managing the various resources in cloud environment (compute, network, storage, and so on). The project mission is “To produce the ubiquitous open source cloud computing platform that will meet the needs of public and private cloud providers regardless of size, by being simple to implement and massively scalable.”^[1]

OpenStack was originally launched by Rackspace in collaboration with NASA in July 2010. It gained significant interest in a short time from both academic and industry leaders. As of this writing, there are more than 170 companies^[2] that have joined the project. It is the strong vibrant community of OpenStack contributors that is the key to the strength of OpenStack. All of the code of OpenStack is freely available under the Apache 2.0 license.

As shown in Figure1, OpenStack is composed of a set of interrelated projects, which make up the various components of a cloud computing platform. The latest stable release named Essex^[3] includes five key projects:

- OpenStack Compute (codenamed Nova) provides a tool to provision and manage large number of virtual machine instances and networks. It is similar in scope to Amazon EC2* and Rackspace Cloud Servers*. OpenStack Compute is designed to be both hardware and hypervisor agnostic.
- OpenStack Object Storage (codenamed Swift) provides a distributed, eventually consistent virtual object store. Swift has built-in redundancy and failover management and allows scaling to multiples of petabytes, billions of objects distributed across nodes.
- OpenStack Image Service (codenamed Glance) provides discovery, registration, and retrieval services for virtual machine images.
- OpenStack identity (codenamed Keystone) provides a unified authentication across all OpenStack projects, including Token, Catalog, and Policy services.
- OpenStack Dashboard (codenamed Horizon) provides administrators and users the ability to manage their infrastructure using a simple web interface.

“OpenStack is a new and emerging open source software stack for creating and managing clouds.”

“There are more than 170 companies^[2] that have joined the project.”

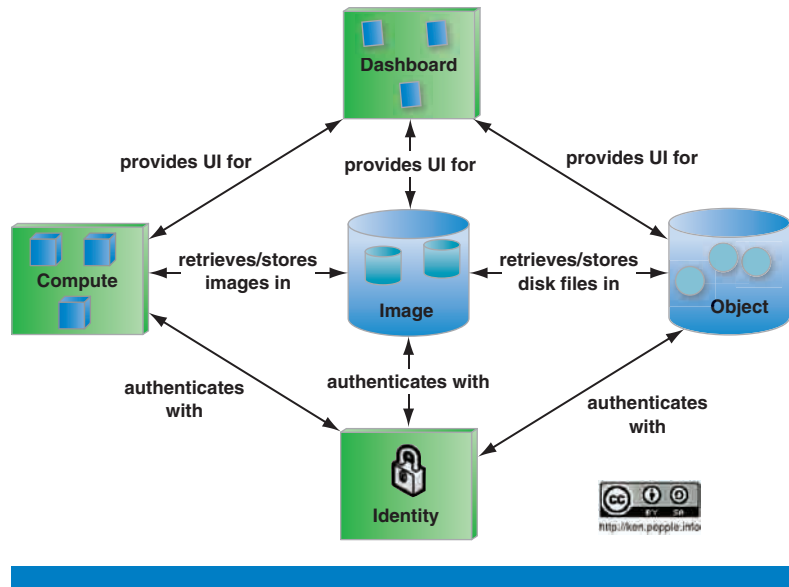


Figure 1: Key OpenStack projects

(Source: www.openstack.org/Documentation/os-compute-adminguide.pdf, the document itself is under Apache 2.0 License)

In the balance of this article we examine in more detail a real-world implementation of OpenStack in Intel IT and extensions to OpenStack for support of trust and power management.

Intel IT OpenCloud

In late 2010, Intel IT implemented their first private cloud solution focused on delivering compute IaaS to its internal application developers and application owners. This was predominantly built through a combination of existing enterprise manageability tools and solutions coupled with integration software and databases. The goal and results were aimed to provide a cohesive solution that brought together compute, storage, and network resources to expose compute IaaS to Intel IT’s end users. The success of the work was paramount in instituting rapid instantiation of new capacity for application developers (from 90 days to acquire a server to under three hour SLA with most happening in under 45 minutes), to establishing a federated capacity allowing increased sharing through multitenancy and resource pooling of assets, and to deliver numerous improvements in the automation and data transparency available to the operations team.

“In late 2011, Intel IT decided to augment their cloud with an open source approach.”

In late 2011, Intel IT decided to augment their cloud with an open source approach. This approach allows the team to get the latest technology the open source community has developed, focus only on the specific challenges they faced, and also provide feedback to the community with Intel’s most advanced technologies. The new design and implementation based on open source is referred to as the Intel IT OpenCloud.

Foundation of the Intel IT OpenCloud

For Intel IT OpenCloud the team had an opportunity to revisit a number of design decisions and seek out areas allowing them to work in the open community for cloud acceleration.

The team analyzed numerous options based on performance, reliability, cost, and features, which led to an architecture consisting of Intel 5600 series blade servers for compute servers, utilizing 10GBe for network fabric, and Intel 5600 2U servers for the storage nodes. The modular nature of the underlying platform with large network pipes for connectivity is intended to allow for significant scale out, to achieve the necessary resiliency and performance required for a wide range of workloads with a primary focus on cloud-aware applications.

For the cloud operating environment, the team explored numerous solutions and closed in on OpenStack for its vibrant community and its ability to flexibly handle the operations for compute, network, and storage. The team implemented the 1.0 version of Intel IT OpenCloud based on OpenStack Diablo release. The environment is built to handle rolling upgrades to allow for no impact implementation of new infrastructure versions, therefore the OpenStack Essex version could be integrated no later than three months after release.

A key aspect of the foundation is the usage of Nova security groups to allow for automated logical segmentation amongst tenants in the environment as well as between VM roles inside of a single tenant. The utilization of security groups enables rapid configuration of IP tables at instantiation time, which significantly lowers the effort expended in ensuring reasonable segmentation in a multitenant resource pool.

Monitoring and Managing the OpenCloud

In the existing Intel IT private cloud, significant focus is placed on ensuring that we could monitor the entire environment, including compute, storage, and network resources, simultaneously in a single view. It is also critical that operators can manage all resources in an automated fashion for significant scale with minimal interaction. For the Intel IT OpenCloud, the team chose to implement a solution based on the open source monitoring tool Nagios by taking advantage of its extensive list of pre-built monitors, its support of multiple operating systems, and its ability to monitor many of the resources beyond OS (load balancers, firewalls, network switches, and so on).

Monitoring alone is not that useful in an automated environment. The team decided to make extensive use of a real-time configuration management system to control the environment as it scales. The team focused on the open source tool Puppet to handle most of configuration management actions. However to complete the monitoring and managing circle to enable full automation the team architected in a small but important component that handles business logic rules.

“For the cloud operating environment, the team explored numerous solutions and closed in on OpenStack.”

“The team decided to make extensive use of a real-time configuration management system to control the environment as it scales.”

“One of the goals of the team was to show how an open source infrastructure could integrate well with existing solutions that run an enterprise.”

“The design allows for a very flexible approach to what causes alerts, what causes auto-remediation.”

The basic aspects of the architecture and implementation are that the watcher (Nagios) sends alerts onto the message bus to which the event handler is subscribed, and the event handler makes decisions based on real-time configuration data of the infrastructure and application layout. As examples:

- If an event is sent onto the bus about a specific node having issues inside a given application scale unit (combination of server instances for scaling), the event handler will decide to destroy that node and instantiate a replacement.
- If a more catastrophic failure happens that affects the entire data center, the event handler can make a choice to disable a scale unit inside the data center, or even remove the data center completely from the global load balancer list of DNS end points.

Currently the team focused on three levels of automated remediation. The three levels are: destroy and create node; remove scale unit from load balancer (in some situations this means 20 servers removed from the load balancer); and remove data center from global load balancer pool. These three levels allow for cloud-aware applications to operate effectively in an active/inactive implementation dispersed across multiple data centers to ensure high reliability, which is a key goal of the Intel IT OpenCloud team.

Integration with Enterprise Systems

Intel IT has investments in numerous enterprise technologies from our service management tools to our authentication and entitlement tools. One of the goals of the team was to show how an open source infrastructure could integrate well with existing solutions that run an enterprise. The team picked a few areas to focus on first, and in this article we outline service management.

The integration into the service management system was paramount as the Intel IT team was in the midst of transforming into a complete Information Technology Information Library (ITIL) environment. The architecture and design goals were to have the systems themselves provide the necessary data. Therefore the utilization of the configuration management system coupled with the monitoring system and correlation engine allowed for provision time correlation of resources, which is fed onto the message bus and then imported into the service management tool. The monitor/watcher is also fed information at provision time to ensure that the resources are immediately monitored and so alerts on those resources can be easily ingested into the service management tool again through the use of the message bus. This allows for automated remediation to happen in a self-contained fashion and only exceptions requiring an operator to receive a ticket for problem management. By utilizing a message bus model with publish/subscribe methods, the design allows for a very flexible approach to what causes alerts, what causes auto-remediation, and what generates a ticket for operator analysis.

The team expects to continuously improve the environment, and the next areas of focus on are orchestration, block storage, auto-scaling policies, complex application deployment, as well as providing the foundation for the Intel IT PaaS solution.

Intel Technology Adding Value in Cloud

Intel IT OpenCloud provides a success story to use OpenStack to build up a viable private compute IaaS solution. Though this is a good foundation, there is still much work ahead to make cloud computing optimized. Security and energy efficiency are undisputedly among the top cloud computing challenges. In the following section, we will introduce Intel Node Manager and Intel Trusted Execution Technology, and what additional value they can bring to cloud and make the cloud environment more power-efficient and secure.

“Security and energy efficiency are undisputedly among the top cloud computing challenges.”

Server Power Management

Server power consumption is often an afterthought in data centers. For example, in many facilities the utility bill is bundled with the overall building charge which reduces the visibility of the data center cost.

Even though servers have become much more efficient, packaging densities and power have increased much faster. As a result, power and its associated thermal characteristics have become the dominant components of operational costs.^[4]

“Power and its associated thermal characteristics have become the dominant components of operational costs.”

Power and thermal challenges in data centers include:

- Increased total operational costs due to increased power and cooling demands.
- Physical limitations of cooling and power within individual servers, racks, and data center facilities.
- Lack of visibility into actual real-time power consumption of servers and racks.
- Complexity of management components and subsystems from multiple vendors with incompatible interfaces and management applications.

These challenges to manage data centers can be translated into the following requirements:

- Power monitoring and capping capabilities at all levels of the data center (system, rack identification, and data center). What can be done at an individual server level becomes much more compelling once physical or virtual servers are scaled up significantly.
- Aggregation of the power consumed at the rack level and management of power within a rack group to ensure that the total power does not exceed the power allocated to a rack.
- Higher level aggregation and control at the row or data center level to manage power budget within the average power and cooling resources available.
- Optimization of productivity per watt through management of power at the server, rack, row, and data center levels to optimize TCO.
- Application of standards-based power instrumentation solutions available in all servers to allow management for optimal data center efficiency. Extension of instrumentation to enable load balancing or load migration

based on power consumption, and close coupled cooling for the management of pooled power and cooling resources.

Intel® Node Manager

Intel Node Manager is a smart way to optimize and manage power and cooling resources in the data center. This server power management technology extends component instrumentation to the platform level and can be used to make the most of every watt consumed in the data center.^[5]

“Intel Node Manager is designed to address typical data center power requirements.”

Intel Node Manager is designed to address typical data center power requirements such as described above. It is implemented on Intel server chipsets starting with Intel® Xeon® processor 5500 series platforms. It provides power and thermal monitoring and policy based power management for an individual server and is exposed through a standards based IPMI interface on supported Baseboard Management Controllers (BMCs). Intel Node Manager requires an instrumented power supply conforming to the PMBus standard.

“Intel Xeon processors regulate power consumption through voltage and clock frequency scaling.”

Intel Xeon processors regulate power consumption through voltage and clock frequency scaling. Reducing the clock frequency reduces power consumption, as does lowering voltage. The scale of reduction is accomplished through a series of discrete steps, each with a specific voltage and frequency. Voltage and frequency scaling also impacts overall system performance and therefore will constrain applications. The control range is limited to a few tens of watts per individual microprocessor. This may seem insignificant at the individual microprocessor level; however, when applied to thousands or tens of thousands of microprocessors typically found in a large data center, the potential power savings amount to hundreds of kilowatt hours per month.

Intel Node Manager is a chipset extension to the BMC for supporting in-band and out-of-band power monitoring and management at the node (server) level. Some of the key features include:

- Real-time power monitoring
- Platform (server) power capping
- Power threshold alerting

Figure 2 shows the Intel Node Manager server power management closed control loop.

The typical benefits brought by Intel Node Manager are illustrated in the following use cases.

Power Management Use Case 1: Real-Time Server Power Monitoring

Power monitoring is a critical capability that enables us to characterize workloads and identify opportunities and hotspots to increase data center energy efficiency. This use case needs to have Intel Node Manager-enabled hypervisor hosts, which can be a combination of open source Xen or KVM. Real-time power utilization of the server is shown in Figure 3.

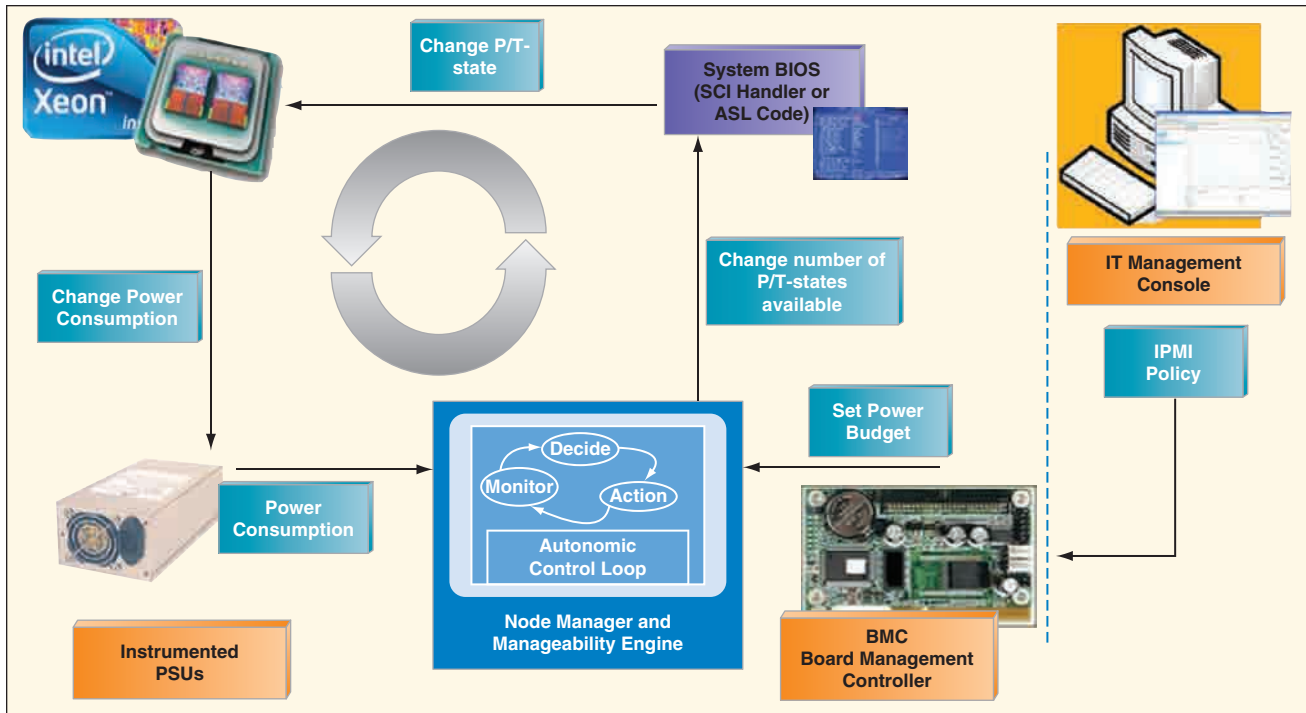


Figure 2: Intel® Node Manager closed control loop
(Source: Intel Corporation, 2012)

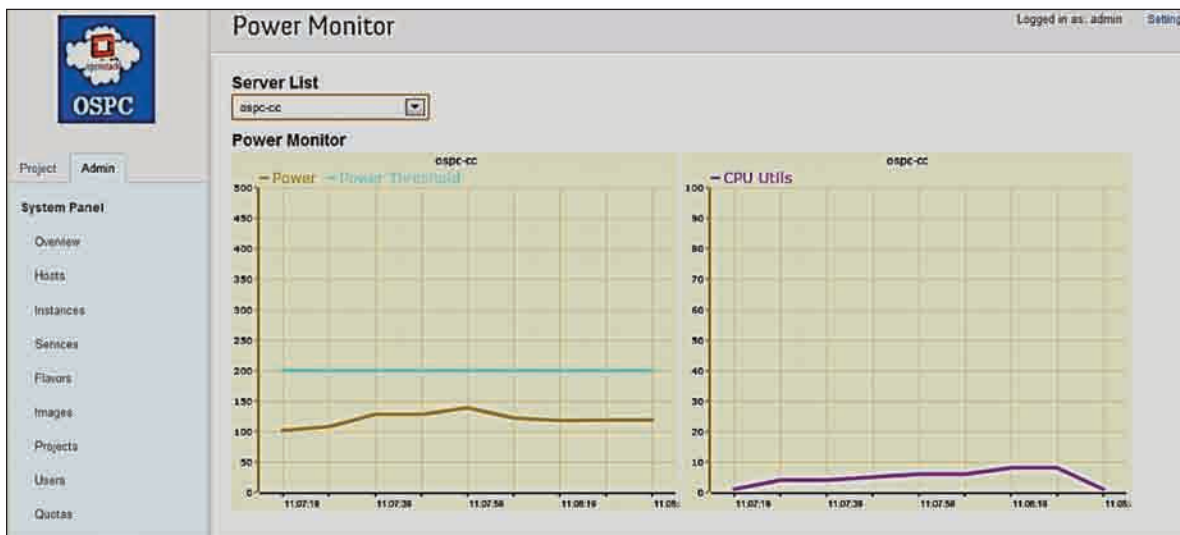


Figure 3: Real-time power consumption monitoring
(Source: OSPC)

Power Management Use Case 2: Policy-Based Resource Distribution Using VM Migrations

Real-time power consumption data allows us to perform power-aware resource distribution. Virtual machines that run the workload can be relocated

to optimize and rebalance power margins based on measurements. The virtual machines can be relocated from power constrained systems to unconstrained systems within the cluster or across different clusters for better system utilization and performance. Figures 4 and 5 show VM migration based on real-time power consumption data. Once the server power goes beyond the threshold, the VM has been migrated out of the server on to other unconstrained server within the cluster. Therefore, the power consumption of the original server has dropped down.

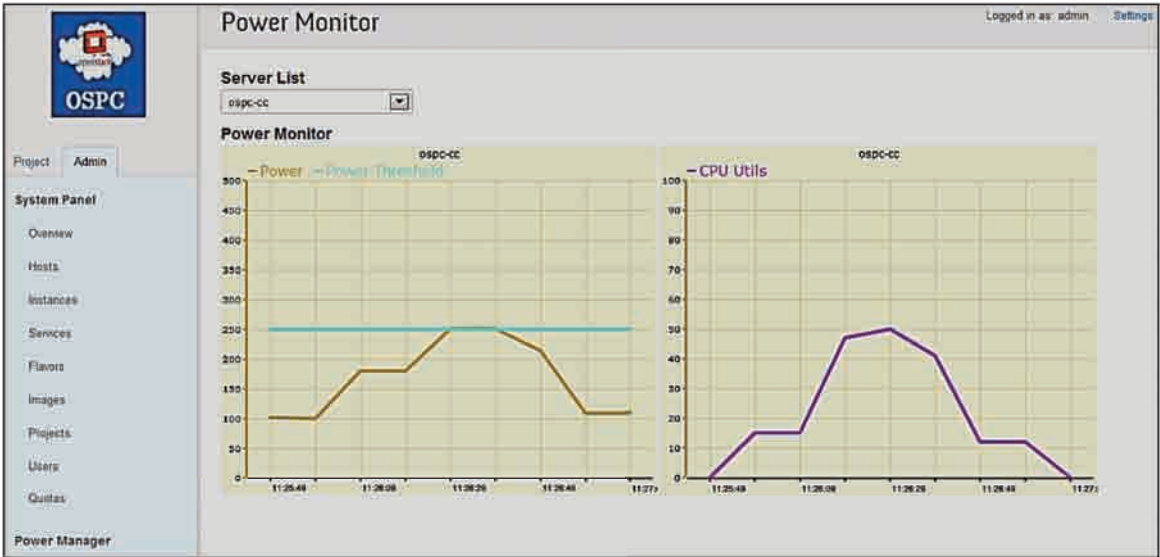


Figure 4: Policy-based VM migration—monitoring
(Source: OSPC)

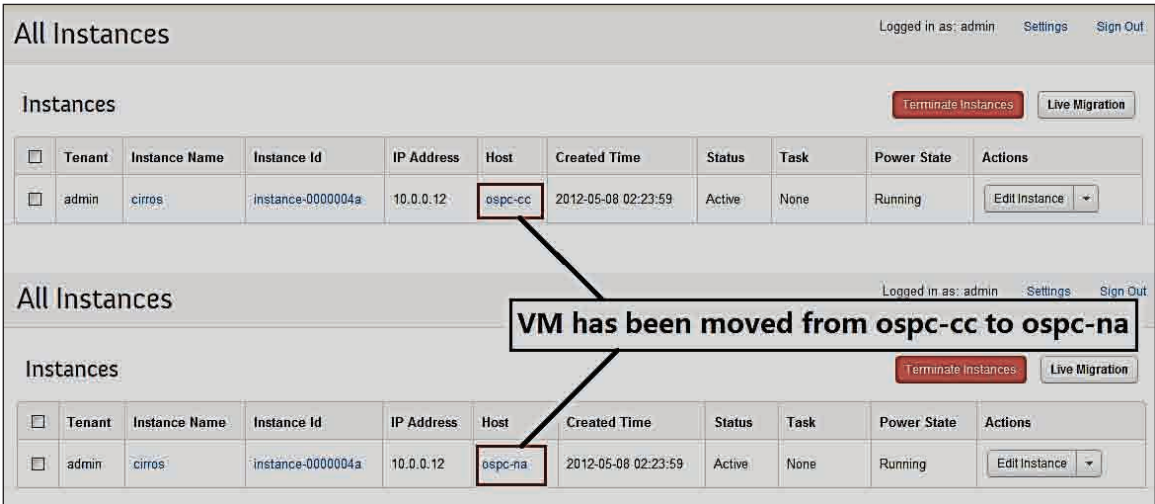


Figure 5: Policy-based VM migration—VM instances
(Source: OSPC)

Power Management Use Case 3: Optimize Rack Density (Policy-Based Power Capping)

Traditionally, we can only estimate the power consumption data from the manufacturer's specifications. This requires the allowance of a hefty safety margin, and thus results in overprovisioned data center power, overcooling of IT equipment, and increased TCO.

The availability of power monitoring data allows management by numbers, which tightly matches servers by power quotas to available data center power. The use case is useful in older data centers underprovisioned for power and in host settings with power quotas in effect. Therefore, it can optimize the rack utilization and increase the data center density.

Figure 6 shows that the power utilization of the server stays within the threshold limit set from the server, which is 160 watts. In the earlier use case we observed that without any power capping, the power utilization of the server went up to about 200 watts

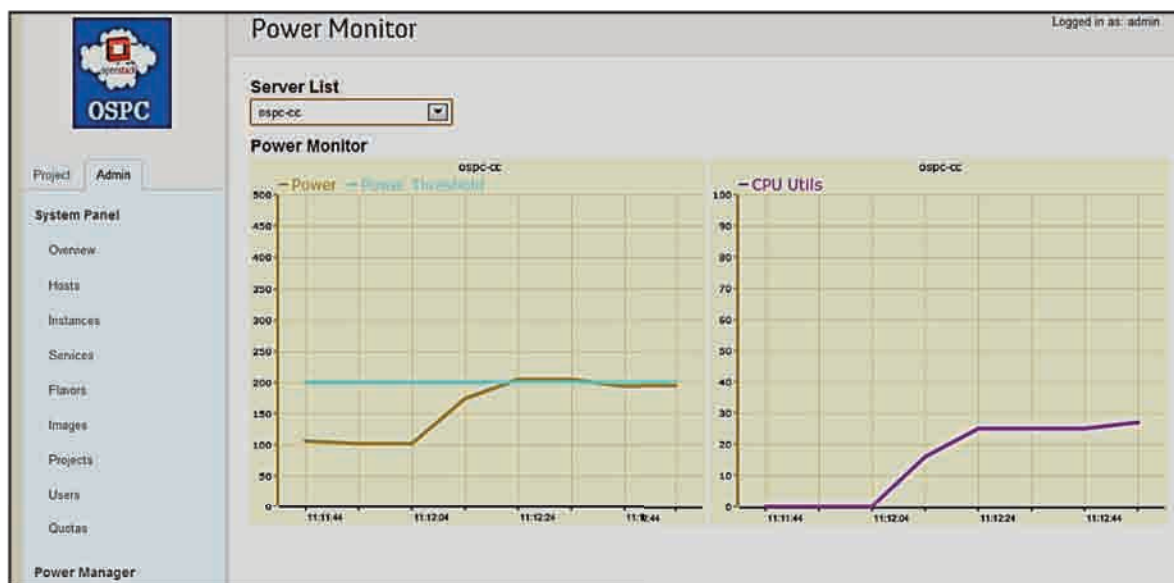


Figure 6: Policy-based power capping
(Source: OSPC)

Security

Recent cloud computing customer surveys unanimously cite security, control, and IT compliance as primary issues that slow the adoption of cloud computing. Many customers have specific security requirements that must assure data location and integrity, and today they're using legacy solutions that rely on fixed hardware infrastructures. However, the means in legacy solution to verify a service's security compliance are labor-intensive, inconsistent, and non-scalable. For this reason, many businesses only deploy non-core applications in the public cloud and restrict sensitive applications to dedicated hardware.

“By providing controls to ensure only a trustable hypervisor is running on a platform, Intel TXT helps protect a server.”

Comprehensive security requires an uninterrupted chain of control from the application user's interfaces to the underlying hardware infrastructure.^[6] Any gaps in this trust chain render them vulnerable to attacks. Intel® Trusted Execution Technology (Intel® TXT) provides hardware-based technologies to establish a root of trust that provides the necessary underpinnings for successful evaluation of the computing platform and its protection. It is specifically designed to harden platforms from the emerging threats of hypervisor attacks, BIOS, or other firmware attacks, malicious root kit installations, or other software-based attacks. Intel TXT gives IT and security organizations important enhancements to help ensure: more secure platforms; greater application, data, or virtual machine isolation; and improved security or compliance audit capabilities.

By providing controls to ensure only a trustable hypervisor is running on a platform, Intel TXT helps protect a server. While this basic protection and enhanced control is good on individual systems, it becomes even more powerful when one considers aggregated resources and dynamic environments such as cloud environment. Intel TXT can help create something known as *trusted computing pools*. In this usage model, a pool of trusted hosts can be created, each with Intel TXT enabled and by which the platform launch integrity has been verified. Data center administrators can set a policy that VMs with high requirements on host trustiness can be only scheduled on or migrated between the hosts in the trusted computing pool. This enables data center administrators to restrict confidential data or sensitive workloads to platforms that are better controlled and have had their configurations more thoroughly evaluated through the use of Intel TXT-enabled platforms.

We have enhanced OpenStack to use Intel TXT to implement the trusted computing pool usage model. See Figure 7 for the detail flow:

- In a cloud computing environment, a subscriber, who requires his VM to run on a trusted platform, can specify the trust level of that VM as Trusted.
- The request will pass along all the way to OpenStack Nova scheduler.
- The scheduler will invoke a web-based remote attestation (OpenAttestation) service to decide the platforms' trustworthiness.
- Based on the results, the scheduler will schedule the Trusted VM to one of the trusted platforms.

In Figure 7, the web-based OpenAttestation service is a standalone open source project in BSD license^[7] (<https://github.com/OpenAttestation/>). It uses platform measurement credentials to complete the trust verification process and support compliance and audit activities. The implementation takes advantage of TCG (Trusted Computing Group) Infrastructure Work Group's Integrity Report Schema Specification. As a standalone SDK, OpenAttestation enables ISV software to remotely retrieve and verify target hosts' TPM PCRs and verify hosts' integrity, through exported Query API. This SDK is a critical open source building block upon which ISVs can build their Intel TXT-based security solution more easily.

“OpenAttestation enables ISV software to remotely retrieve and verify target hosts' TPM PCRs and verify hosts' integrity.”

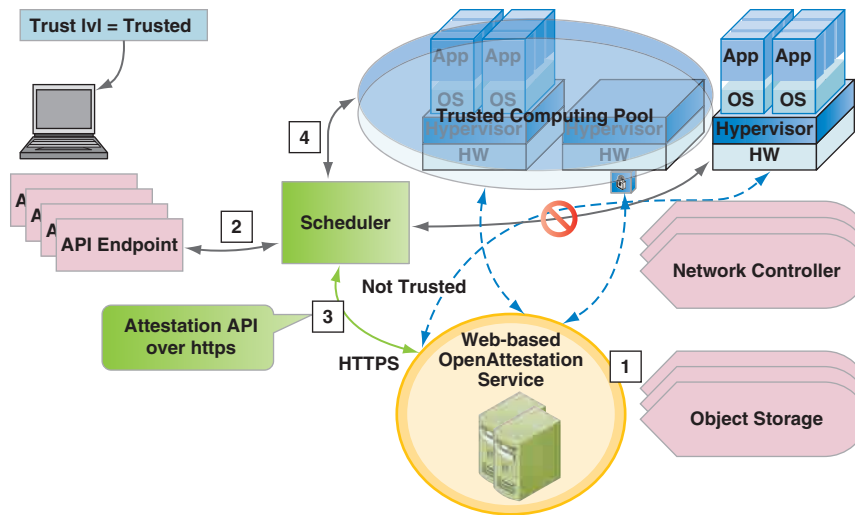


Figure 7: OpenStack trusted computing pool
(Source: Intel Corporation, 2012)

Conclusion

As IT moves into the era of “new IT,” cloud and related technologies are moving quickly. Many factors including cost constraints and consumerization of IT are driving these changes. Open source projects, including OpenStack, are leading the way to solutions for IT address these challenges. As one proof point, Intel IT has implemented an open source private cloud in a production environment based on OpenStack and other open source components. In addition to the challenge of integrating the various software elements into a solution, the Intel IT solution also shows the power of integration back into an existing enterprise IT organization.

Of the challenges faced by IT, power management and security are high on the list. Power is often ignored or misunderstood since the power bill is often buried in another part of the organization’s budget. Through the use of Intel Node Manager, key use cases such as meeting operating cost constraints, physical limitations of power and thermal, and poor visibility into the actual power and thermal environment can be overcome. In the scope of security challenges, providing support for compliance and audit are important considerations. The Intel Trusted Execution Technology (Intel TXT) is a powerful tool to provide hardware based root of trust. Combined with remote attestation, Intel TXT provides a strong basis, rooted in hardware, for compliance and audit scenarios.

OpenStack has very strong industry momentum and enjoys an active and innovative community. Open source technologies, including OpenStack are well positioned to lead the market for solutions to key IT challenges utilizing Intel technology.

“Power is often ignored or misunderstood since the power bill is often buried in another part of the organization’s budget.”

References

- [1] <http://wiki.openstack.org/>
- [2] <http://openstack.org/community/companies/>
- [3] <http://www.openstack.org/projects/essex/press-release/>
- [4] http://www.intelcloudbuilders.com/index.php?option=com_productsearch&view=displaysearch&lang=&filter_companyname=1322&filter_category=&filter_efficiency=&filter_region=&company_name=OpenStack&Itemid=181
- [5] <http://www.intel.com/content/www/us/en/data-center/data-center-management/node-manager-general.html>
- [6] <http://www.intel.com/content/www/us/en/trusted-execution-technology/trusted-execution-technology-security-paper.html?wapkw=intel+trusted+execution+technology>
- [7] <https://github.com/OpenAttestation/OpenAttestation>

Author Biographies

Billy Cox has been leading the cloud strategy efforts for Intel's Software and Services Group since 2007. He is also responsible for Intel's Cloud Builders program, which brings together cloud leaders to provide best practices and practical guidance on how to deploy, optimize, and support a cloud infrastructure. The program offers detailed reference architectures and best practices to build simplified, secure, and efficient cloud infrastructure. Previously, Billy was Director of Systems Engineering at HP for 14 years where he was responsible for the development of all infrastructure management tools used to manage the various server and storage platforms. With over 30 years of industry experience, Billy has led the design of compute, network, and storage solutions and actively participated in multiple standards efforts. His e-mail address is billy.cox at intel.com.

Jiangang Duan manages the Scalability Lab in Intel Asia-Pacific Research and Development Ltd. He has worked on enterprise solution tuning and optimization for more than ten years, including several generations of Intel processor performance evaluations and has worked jointly with local OEMs to complete industry benchmark publications. He now focuses on cloud computing and virtualization, and is responsible for development, deployment, and optimization of efficient cloud data centers with open source software (Xen/KVM/OpenStack). Before joining Intel, Jiangang got his bachelor's degree in 1999 and Master's degree in 2001 from the Electrical Engineering Department of Tsinghua University. His e-mail address is jiangang.duan at intel.com.

Fleming Feng is Chief Open Source Scientist in the Open Source Technology Center PRC team. He joined Intel in 1997 and in 2000 he started to work on open source software development from Carrie Grade Linux project; he then worked on various open source projects. His current role is mostly working as technical interface for PRC local engagement on government collaboration, the open source community, and business engagement based on open source solutions. His e-mail address is fleming.feng@intel.com.

Das Kamhout is a principal engineer in Intel IT responsible for the architecture, strategy, and execution of the Intel IT Cloud Journey. In his 15 years at Intel he has been responsible for everything from the clients that Intel IT uses, to the architecture that Intel's Design Grid runs upon. His e-mail address is das.kamhout@intel.com.

Susie Li is an engineering manager in the Open Source Technology Center in Shanghai, PRC. She joined Intel in 1999 and had been involved in IPMI, EFI, virtualization, and cloud projects. She currently manages the Cloud and Virtualization team in the Intel Open Source Technology Center. Her team is focusing on enabling Intel technologies in open source virtualization and cloud stack (Xen, KVM, and OpenStack). She received her Bachelor's and Master's degrees in Engineering from Shanghai Jiao Tong University, PRC. Her e-mail address is susie.li@intel.com.

Tonny Liu is an engineering manager from IT Flex China team. He and his team have been working on cloud computing, mobile development, and enterprise solutions for more than eight years. Tonny got his Bachelor's and Master's degrees from the Electrical Engineering department of Southwest Jiao Tong University in 1998 and 2002 respectively. He personally holds ICCP and PMP certifications. His e-mail address is tonny.h.liu@intel.com.

THE HADOOP STACK: NEW PARADIGM FOR BIG DATA STORAGE AND PROCESSING

Contributors

Jinquan Dai

Software and Services Group,
Intel Corporation

Jie Huang

Software and Services Group,
Intel Corporation

Shengsheng Huang

Software and Services Group,
Intel Corporation

Yan Liu

Software and Services Group,
Intel Corporation

Yuanhao Sun

Software and Services Group,
Intel Corporation

*“Big data is rich with promise, but
equally rife with challenges.”*

We are on the verge of the “industrial revolution of Big Data,” which represents the next frontier for innovation, competition, and productivity.^[1] Big data is rich with promise, but equally rife with challenges—it extends beyond traditional structured (or relational) data, including unstructured data of all types; it is not only large in size, but also growing faster than Moore’s law. In this article, we first present the new paradigm (in particular, the Hadoop stack) that is required for big data storage and processing. After that, we describe how to optimize the Hadoop deployment through proven methodologies and tools provided by Intel (such as HiBench and HiTune). Finally, we demonstrate the challenges and possible solutions for real-world big data applications using a case study of an intelligent transportation system (ITS) application.

Introduction

We are on the verge of the “industrial revolution of Big Data,” where a vast range of data sources (from web logs and click streams, to phone records and medical history, to sensors and surveillance cameras) is flooding the world with enormous volumes of data in a huge variety of different forms. Significant values are being extracted from this data deluge with extremely high velocity. Big data is already the heartbeat of Internet, social and mobile; more importantly, it is heading toward ubiquity as enterprises (telecommunications, governments, financial services, healthcare, and so on) have amassed terabytes and even petabytes of data that they are not yet prepared to process. And soon, all will be awash with even more data from ubiquitous devices and sensors as we enter the age of the Internet of Things. These big data trends represent the next frontier for innovation, competition, and productivity.

Big data is rich with promise, but equally rife with challenges. It extends beyond traditional structured (or relational) data, including unstructured data of all types (text, image, video, and more); it is not only large in size, but also growing faster than Moore’s law (more than doubling every two years).^[2] In this article, we first present the new paradigm (in particular, the Hadoop stack) that is required for big data storage and processing. After that, we describe how to optimize the Hadoop deployment (through proven methodologies and tools provided by Intel). Finally, we demonstrate the challenges and possible solutions for real-world big data applications using a case study.

New Paradigm for Big Data Analytics

Big data is powering the next industrial revolution. Pioneering Web companies (such as Google, Facebook, Amazon, and Taobao) have already been looking

at data in completely new ways to improve their business. It will become even more critical for corporations and governments to harvest values from the massive amount of untapped data in the future.

However, big data is also radically different from traditional data. In this section, we describe the new challenges brought by big data and the new paradigm for big data processing that addresses these challenges.

Big Data Is Different from Traditional Data

Some define big data as the “datasets whose size is beyond the ability of typical database software tools to capture, store, manage and analyze.”^[1] However, big data is not only massive in scale, but also diverse in nature.

“Big data is not only massive in scale, but also diverse in nature.”

- *Unstructured data:* Unlike traditional structured (or relational) data in enterprise databases or data warehouses, big data is mostly about unstructured data—coming from many different sources, in many different forms (such as text, image, audio, video, and sensor readings), and often with conflicting syntax and semantics.
- *Massive scale and growth:* Unstructured data is growing 10–50 times faster than structured data, and soon will represent 90 percent of all data.^[2] Consequently, big data is both large in size (10–100 times larger than traditional data warehouses^[3]) and growing exponentially (increasing by roughly 60 percent annually), faster than Moore’s law.^[2]
- *Scale-out framework:* The massive scale, exponential growth, and viable nature of big data necessitate a much more scalable and flexible data management and analytics framework. Consequently, emerging big data frameworks (such as Hadoop/MapReduce and NoSQL) have adopted a scale-out (rather than scale-up), shared-nothing architecture, with massively distributed software running on clusters of independent servers.
- *Real-time, predictive analytics:* There is significant value to be extracted from big data (for example, a 300 billion US dollar [USD] potential annual value to US healthcare^[11]). To realize this value, a new class of predictive analytics (with complex machine learning, statistic modeling, graph analysis, and so on) is needed to identify the future trends and patterns from within massive seas of data and in (near) real time as data is streaming in continuously.

The Hadoop Stack: A New Big Data Processing Paradigm

The massive scale, exponential growth, and variable nature of big data necessitate a new data processing paradigm. In particular, the Hadoop stack, as illustrated in Figure 1, has emerged as the de facto standard for big data storage and processing. In this section, we provide a brief overview of the core components in the Hadoop stack (namely, MapReduce^{[4][5]}, HDFS^{[5][6]}, Hive^{[7][8]}, Pig^{[9][10]}, and HBase^{[11][12]}).

“The massive scale, exponential growth, and variable nature of big data necessitate a new data processing paradigm.”

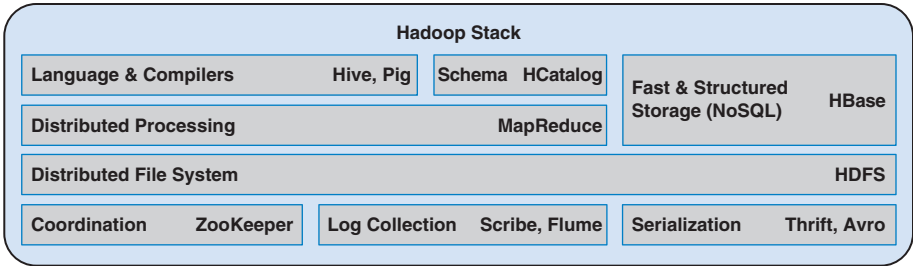


Figure 1: The Hadoop stack
(Source: Intel Corporation, 2012)

MapReduce: Distributed Data Processing Framework

At a high level, the MapReduce^[4] model dictates a two-stage group-by-aggregation dataflow graph to the users, as shown in Figure 2. In the first phase, a *Map* function is applied in parallel to each partition of the input data, performing the grouping operations. In the second phase, a *Reduce* function is applied in parallel to each group produced by the first phase, performing the final aggregation. In addition, a user-defined *Combiner* function can be optionally applied to perform the map-side “pre-aggregation” in the first phase, which helps decrease the amount of data transferred between the map and reduce phases.

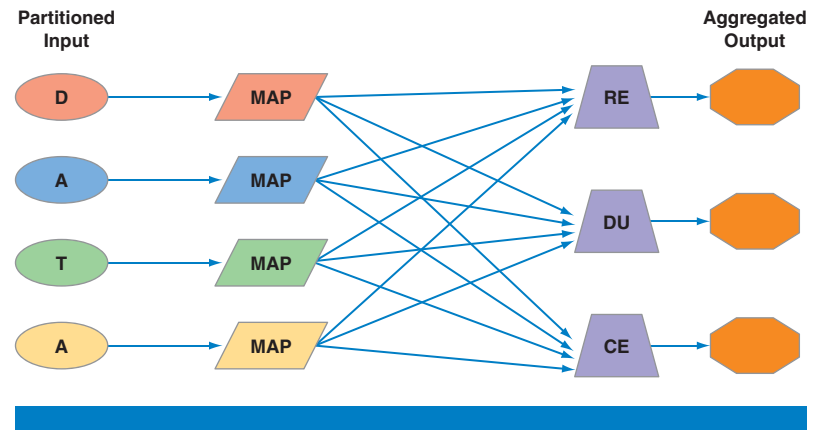


Figure 2: MapReduce dataflow model
(Source: Intel Corporation, 2012)

Hadoop is a popular open source implementation of MapReduce. The Hadoop framework is responsible for running a MapReduce program on the underlying cluster that may comprise thousands of nodes. In the Hadoop cluster, there is a single master (*JobTracker*) controlling a number of slaves (*TaskTrackers*). The user writes a Hadoop job to specify the Map and Reduce functions (in addition to other information such as the locations of the input and output), and submits the job to the JobTracker. The Hadoop framework divides a Hadoop job into a series of map or reduce tasks; the master is responsible for assigning

the tasks to run concurrently on the slaves, and rescheduling the tasks upon failures. Figure 3 illustrates the architecture of the Hadoop cluster.

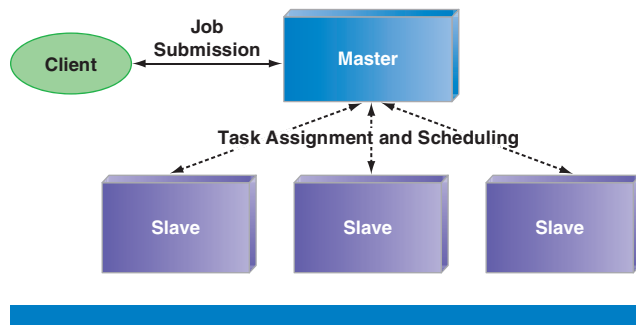


Figure 3: Architecture of the Hadoop cluster
(Source: Intel Corporation, 2012)

Hive and Pig: High-Level language for Distributed Data Processing

The Pig^[9] and Hive^[7] systems allow the users to perform ad-hoc analysis of big data on top of Hadoop, using dataflow-style scripts and SQL-like queries respectively. For instance, The following example shows the Pig program (an example in the original Pig paper^[9]) and Hive query for the same operation (that is, finding, for each sufficiently large category, the average *pagerank* of high-*pagerank* *urls* in that category). In these two systems, the high level query or script program is automatically compiled into a series of MapReduce jobs that are executed on the underlying Hadoop system.

Pig Script

```

good_urls = FILTER urls BY pagerank > 0.2;
groups = GROUP good_urls BY category;
big_groups = FILTER groups BY COUNT(good_urls)>1000000;
output = FOREACH big_groups GENERATE category, AVG(good_urls.pagerank);
  
```

Hive Query

```

SELECT category, AVG(pagerank)
FROM (SELECT category, pagerank, count(1) AS recordnum
      FROM urls WHERE pagerank > 0.2
      GROUP BY category) big_groups
WHERE big_groups.recordnum > 1000000
  
```

HDFS: Hadoop Distributed File System

The Hadoop distributed file system (HDFS) is a popular open source implementation of Google File System.^[6] An HDFS cluster consists of a single master (or *NameNode*) and multiple slaves (or *DataNodes*), and is accessed by multiple clients, as illustrated in Figure 4. Each file in HDFS is divided into large (64 MB by default) blocks and each block is replicated on multiple (3 by default) *DataNodes*.

The *NameNode* maintains all file system metadata (such as the namespace and replica locations), and the *DataNodes* store HDFS blocks in local file systems and

handle HDFS read/write requests. An HDFS client interacts with the NameNode for metadata operations (for example, open file or delete file) and replica locations, and it directly interacts with the appropriate DataNode for file read/write.

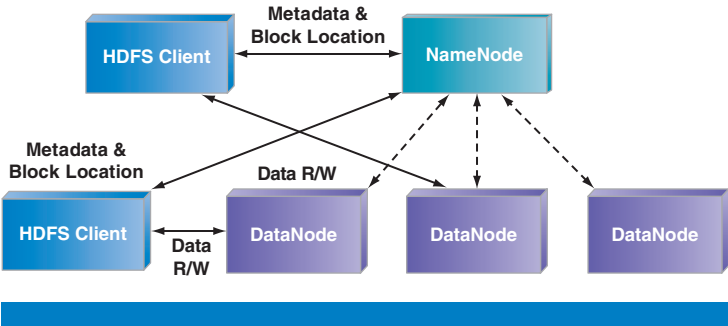


Figure 4: HDFS Architecture
(Source: Intel Corporation, 2012)

HBase: High Performance, Semi-Structured (NOSQL) Database

HBase, an open source implementation of Google’s Bigtable^[11], provides a sparse, distributed, column-oriented table store. In HBase, each value is indexed by the tuple (row, column, timestamp), as illustrated in Figure 5. The HBase API provides functions for looking up, writing, and deleting values using the specific row key (possibly with column names), and for iterating over data in several consecutive rows.

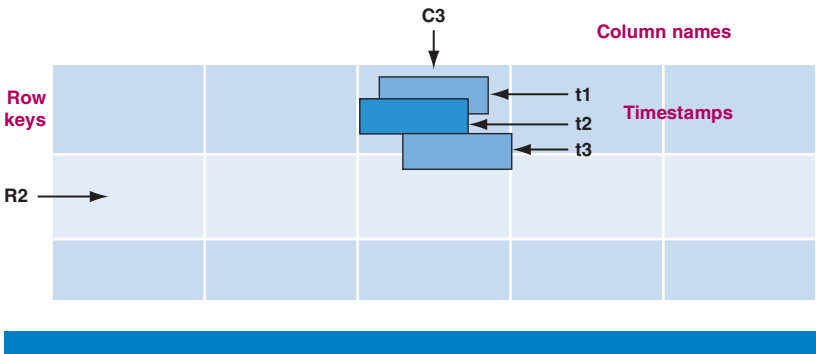


Figure 5: The HBase Data Model
(Source: Intel Corporation, 2012)

Physically, rows are ordered lexicographically and dynamically partitioned into row ranges (or *regions*). Each region is assigned to a single *RegionServer*, which handles the all data accesses requests to its regions. Mutations are first committed to the append-only log on HDFS and then write to the in-memory *memtable* buffer; the data in the region are stored in *SSTable*^[11] files (or *HFiles*) on HDFS, and a read operation is executed on a merged view of the sequence of HFiles and the memtable.

Optimizing Hadoop Deployments

As Hadoop-based big data systems grow in pervasiveness and scale, efficiently tuning and optimizations of Hadoop deployments remain a huge challenge for the users. For instance, within the Hadoop community, tuning Hadoop jobs is considered to be a very difficult problem and requires a lot of effort to understand Hadoop internals^[13]; in addition, the lack of tuning tools for Hadoop often forces users to resort to trial-and-error tuning.^[14] In this section, we present systematic approaches to optimizing Hadoop deployments, including representative workloads, dataflow-based analysis, and a performance analyzer for Hadoop.

“Tuning Hadoop jobs is considered to be a very difficult problem and requires a lot of effort to understand Hadoop internals.”

HiBench: A Representative Hadoop Benchmark Suite

To understand the characteristics of typical Hadoop workloads, we have constructed HiBench^[15], a representative and comprehensive benchmark suite for Hadoop, which consists of a set of Hadoop programs including both synthetic micro-benchmarks and real-world applications. Currently the HiBench suite contains ten workloads, classified into four categories, as shown in Table 1.

Category	Workload
Micro Benchmarks	Sort
	WordCount
	TeraSort
	EnhancedDFSIO
Web Search	Nutch Indexing
	Page Rank
Machine Learning	Bayesian Classification
	K-means Clustering
Analytical Query	Hive Join
	Hive Aggregation

Table 1: HiBench Workloads
(Source: Intel Corporation, 2012)

Micro Benchmarks

The *Sort*, *WordCount*, and *TeraSort*^[16] programs contained in the Hadoop distribution are three popular micro-benchmarks widely used in the community, and therefore are included in HiBench. Both the *Sort* and *WordCount* programs are representative of a large subset of real-world MapReduce jobs—one transforming data from one representation to another, and another extracting a small amount of interesting data from a large data set.

In HiBench, the input data of *Sort* and *WordCount* workloads are generated using the *RandomTextWriter* program contained in the Hadoop distribution. The *TeraSort* workload sorts 10 billion 100-byte records generated by the *TeraGen* program contained in the Hadoop distribution.

We have also extended the *DFSIO* program contained in the Hadoop distribution to evaluate the aggregated bandwidth delivered by HDFS. The

original DFSIO program only computes the average I/O rate and throughput of each map task, and it is not a straightforward process to properly sum up the I/O rate or throughput if some map tasks are delayed, retried or speculatively executed by the Hadoop framework. The Enhanced DFSIO workload included in HiBench computes the aggregated bandwidth by sampling the number of bytes read/written at fixed time intervals in each map task; during the reduce and post-processing stage, the samples of each map task are linear interpolated and resampled at a fixed plot rate, so as to compute the aggregated read/write throughput by all the map tasks.^[15]

Web Search

The *Nutch Indexing* and *Page Rank* workloads are included in HiBench, because they are representative of one of the most significant uses of MapReduce (that is, large-scale search indexing systems).

The Nutch Indexing workload is the indexing subsystem of Nutch^[17], a popular open-source (Apache) search engine; we have used the crawler subsystem in Nutch to crawl an in-house Wikipedia mirror and generated about 2.4 million Web pages as the input of this workload. The Page Rank workload is an open source implementation of the page-rank algorithm in Mahout^[18] (an open-source machine learning library built on top of Hadoop).

Machine Learning

The *Bayesian Classification* and *K-means Clustering* implementations contained in Mahout are included in HiBench, because they are representative of one of another important uses of MapReduce (that is, large-scale machine learning).

“The Bayesian Classification workload implements the trainer part of Naive Bayesian.”

The Bayesian Classification workload implements the trainer part of Naive Bayesian (a popular classification algorithm for knowledge discovery and data mining). The input of this benchmark is extracted from a subset of the Wikipedia dump. The Wikipedia dump file is first split using the built-in *WikipediaXmlSplitter* in Mahout, and then prepared into text samples using the built-in *WikipediaDatasetCreator* in Mahout. The text samples are finally distributed into several files as the input of the benchmark.

“We have developed a random data generator using statistic distributions to generate the workload input.”

The K-means Clustering workload implements K-means (a well-known clustering algorithm for knowledge discovery and data mining). Its input is a set of samples, and each sample is represented as a numerical d-dimensional vector. We have developed a random data generator using statistic distributions to generate the workload input.

Analytic Query

The *Hive Join* and *Hive Aggregation* queries in the *Hive performance benchmarks*^[19] are included in HiBench, because they are representative of another one of the most significant uses of MapReduce (that is, OLAP-style analytical queries).

Both Hive Join and Aggregation queries are adapted from the query examples in Pavlo et al.^[14] They are intended to model complex analytic queries over

structured (relational) tables—Hive Aggregation computes the sum of each group over a single read-only table, while Hive Join computes the both the average and sum for each group of records by joining two different tables.

Data Compression

Data compression is aggressively used in real-world Hadoop deployments, so as to minimize the space used to storing the data, and to reduce the disk and network I/O in running MapReduce jobs. Therefore, each workload in HiBench (except for Enhance DFSIO) can be configured to run with compression turned on or off. When compression is enabled, both the input and output of the workload will be compressed (employing a user-specified codec), so as to evaluate the Hadoop performance with intensive data compressions.

“Data compression is aggressively used in real-world Hadoop deployments, so as to minimize the space used to storing the data.”

Characterization of Hadoop Workloads Using a Dataflow Approach

In this section we present the Hadoop dataflow model, which provides a powerful framework for the characterizations of Hadoop workloads. It allows the users to understand the application runtime behaviors (such as task scheduling, resource utilizations, and system bottlenecks), and effectively conduct performance analysis and tuning for the Hadoop framework.

Hadoop Dataflow Model

The Hadoop framework is responsible for mapping the abstract MapReduce model (as described earlier) to the underlying cluster, running the input MapReduce program on thousands of nodes in a distributed fashion. Consequently, the Hadoop cluster often appears as a big black box to the users, abstracting away the messy details of data partitioning, task distribution, fault tolerance, and so on. Unfortunately, this abstraction makes it very difficult, if not impossible, for the users to understand the runtime behavior of Hadoop applications.

Based on the Hadoop framework, we have developed a dataflow model (as shown in Figure 6) for the characterization of Hadoop applications. In the Hadoop framework, the input data are first partitioned into *splits*, and then a distinct map task is launched to process each split. Inside each map task, the *map* stage applies the Map function to the input; the *spill* stage divides the intermediate output into several partitions, combines the intermediate output (if a Combiner function is specified), and finally stores the output into temporary files on the local file system.

“Based on the Hadoop framework, we have developed a dataflow model.”

A distinct reduce task is launched to process each partition of the map outputs. The reduce task consists of three stages (namely, the *shuffle*, *sort*, and *reduce* stages). In the shuffle stage, several parallel copy threads (or *copiers*) fetch the relevant partition of the outputs of all map tasks via HTTP (there is an HTTP server running on each node in the Hadoop cluster); in the meantime, two *merge* threads in the shuffle stage merge those map outputs into temporary files on the local file system. After the shuffle stage is done, the *sort* stage merges all the temporary files, and finally the *reduce* stage processes the partition (that is, applying the Reduce function) and generates the final results.

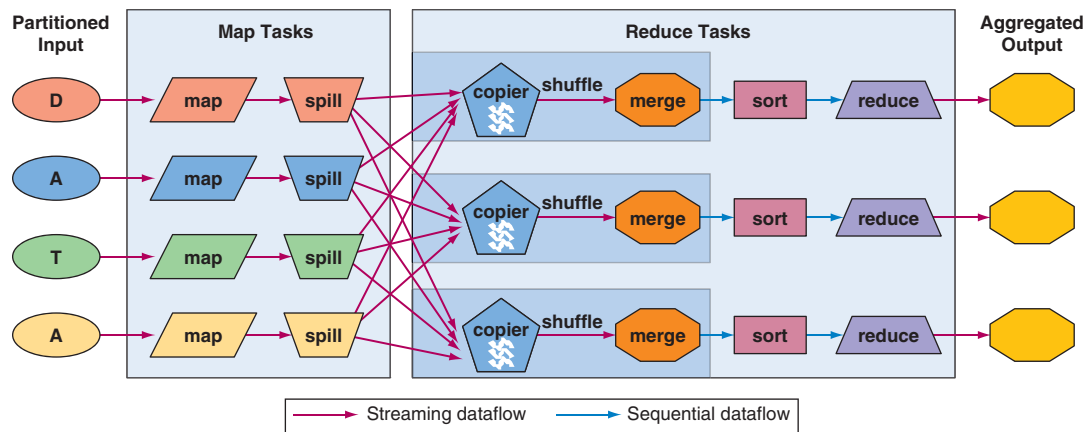


Figure 6: The Hadoop dataflow model
(Source: Intel Corporation, 2012)

Performance Analysis

Figure 7 shows the characteristics of Hadoop TeraSort using the dataflow-based approach. In these charts, the x-axis represents the time elapsed. The first chart in Figure 7 illustrates the timeline-based dataflow execution chart for the workload, where the “bootstrap” line represents the period before the map or reduce task is launched, the “idle” line represents the period after the map or reduce task is complete, the “map” line represents the period when the map tasks are running, and the “shuffle,” “sort,” and “reduce” lines represent the periods when the corresponding stages are running. The other charts in the figure show the timeline-based CPU, disk, and network utilizations (as sampled by the sysstat package every second) of the slaves respectively. We stack these charts together in one figure, so as to understand the system behaviors during different stages of the Hadoop applications.

“TeraSort is a standard benchmark that sorts 10 billion 100-byte records.”

As described earlier, TeraSort is a standard benchmark that sorts 10 billion 100-byte records. It needs to transform a huge amount of data from one representation to another, and therefore is I/O bound in nature. In order to minimize the disk and network I/O during shuffle, we have compressed the map outputs in the experiment, which greatly reduces the shuffle size. Consequently, TeraSort has very high CPU utilization and moderate disk I/O during the map tasks, and moderate CPU utilization and heavy disk I/O during the reduce stages, as shown in Figure 1. In addition, it has almost zero network utilization during the reduce stages, because the final results are not replicated (as required by the benchmark).

HiTune: Hadoop Performance Analyzer

We have built HiTune^[20], a dataflow-based Hadoop performance analyzer that allows users to understand the runtime behaviors of Hadoop applications for performance analysis.

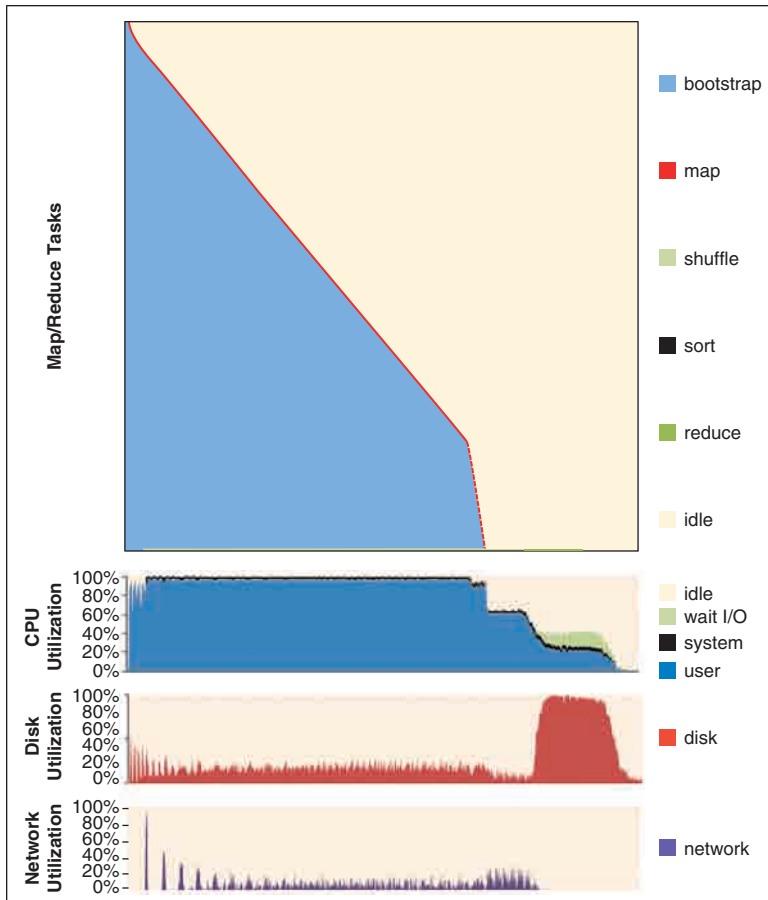


Figure 7: Dataflow-based performance analysis of TeraSort
(Source: Intel Corporation, 2012)

Challenges in Hadoop Performance Analysis

As described before, the Hadoop dataflow abstraction makes it very difficult, if not impossible, for users to efficiently provision and tune these massively distributed systems. Performance analysis for the Hadoop application is particularly challenging due to its unique properties.

- *Massively distributed systems:* Each Hadoop application is a complex distributed application, which may comprise tens of thousands of processes and threads running on thousands of machines. Understanding system behaviors in this context would require correlating concurrent performance activities (such as CPU cycles, retired instructions, and lock contentions) with each other across many programs and machines.
- *High level abstractions:* Hadoop allows users to work at an appropriately high level of abstraction, by hiding the messy details of parallelisms behind the dataflow model and dynamically instantiating the dataflow graph (including resource allocations, task scheduling, fault tolerance, and so on). Consequently, it is very difficult, if not impossible, for users to understand how the low level performance activities can be related to the high level abstraction (which they have used to develop and run their applications).

“Hadoop allows users to work at an appropriately high level of abstraction.”

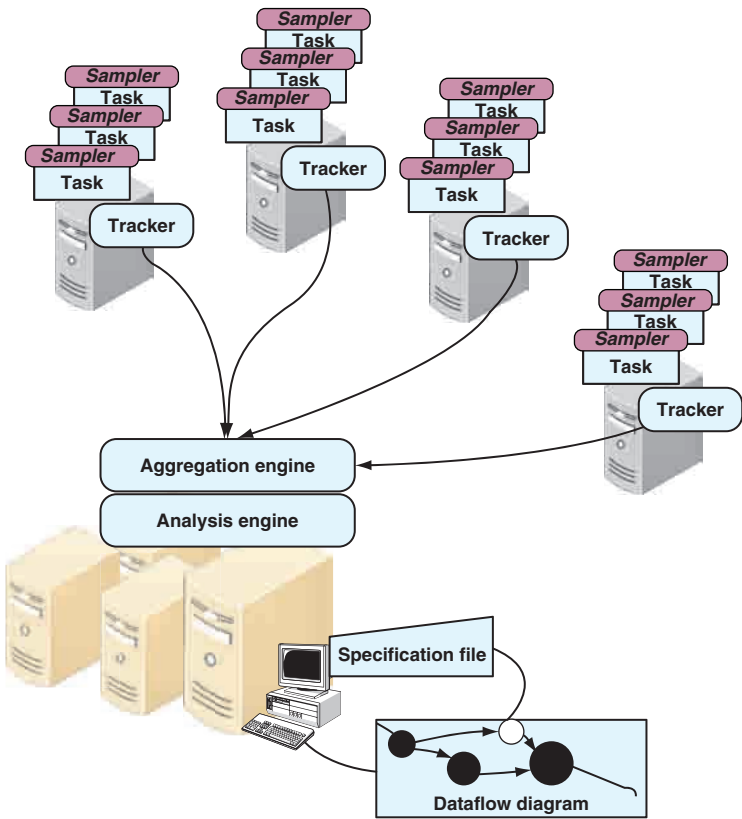


Figure 8: HiTune performance analysis framework
(Source: Intel Corporation, 2012)

HiTune Implementations

Based on the Hadoop dataflow model described earlier, we have built HiHune^[20], a Hadoop performance analyzer that allows users to understand the runtime behaviors of Hadoop applications so that they can make educated decisions regarding how to improve the efficiency of these massively distributed systems—just as traditional performance analyzers like gprof^[21] and Intel® VTune™^[22] allow users to do for a single execution of a single program.

Our approach relies on distributing instrumentations on each node in the Hadoop cluster and then aggregating all the instrumentation results for dataflow-based analysis. The performance analysis framework consists of three major components, namely the tracker, the aggregation engine, and the analysis engine, as illustrated in Figure 8.

Timestamp	Type	Target	Value
-----------	------	--------	-------

Figure 9: Format of the sampling record
(Source: Intel Corporation, 2012)

The tracker is a lightweight agent running on every node. Each tracker has several samplers, which inspect the runtime information of the programs and system running on the local node (either periodically or based on specific events), and sends the sampling records to the aggregation engine. Each sampling record is of the format shown in Figure 9.

- *Timestamp* is the sampling time for each record.

- *Type* specifies the type of the sampling record (such as CPU cycles, disk bandwidth, and log files).
- *Target* specifies the source of the sampling record. It contains the name of the local node, as well as other sampler-specific information (such as CPUID, network interface name, or log file name).
- *Value* contains the detailed sampling information of this record (such as CPU load, network bandwidth utilization, or a line/record in the log file).

The aggregation engine is responsible for collecting the sampling information from all the trackers in a distributed fashion and storing the sampling information in a separate monitoring cluster for analysis. Any distributed log collection tools (examples include Chukwa^{[23][24]}, Scribe^[25], and Flume^[26]) can be used as the aggregation engine. In addition, the analysis engine runs on the monitoring cluster, and is responsible for conducting the performance analysis and generating the analysis report, using the collected sampling information based on the Hadoop dataflow model.

Experience

HiTune has been used intensively inside Intel for Hadoop performance analysis and tuning. In this section, we share our experience on how we use HiTune to efficiently conduct performance analysis and tuning for Hadoop, demonstrating the benefits of dataflow-based analysis and the limitations of existing approaches, including system statistics, Hadoop logs and metrics, and traditional profiling.

Tuning Hadoop Framework

One performance issue we encountered was extremely low system utilization when sorting many small files (3200 500-KB-sized files) using Hadoop 0.20.1; system statistics collected by the cluster monitoring tools (such as Ganglia^[27])

“HiTune has been used intensively inside Intel for Hadoop performance analysis and tuning.”

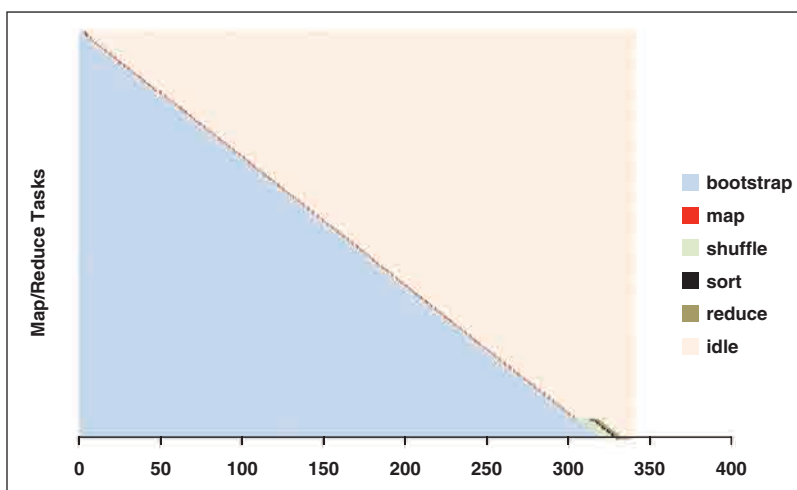


Figure 10: Dataflow execution for sorting many small files
(Source: Intel Corporation, 2012)

showed that the CPU, disk I/O, and network bandwidth utilizations were all below 5 percent. That is, there were no obvious bottlenecks or hotspots in our cluster; consequently, traditional tools like system monitors and program profilers failed to reveal the root cause.

To address this performance issue, we used HiTune to reconstruct the dataflow execution process of this Hadoop job, as illustrated in Figure 10. As is obvious in the dataflow execution, there are few parallelisms between the map tasks, or between the map tasks and reduce tasks in this job. Clearly, the task scheduler in Hadoop 0.20.1 (Fair Scheduler^[28] is used in our cluster) failed to launch all the tasks as soon as possible in this case. Once the problem was isolated, we quickly identified the root cause: by default, the Fair Scheduler in Hadoop 0.20.1 only assigns one task to a slave at each heartbeat (that is, the periodical keep-alive message between the master and slaves), and it schedules map tasks first whenever possible; in our job, each map task processed a small file and completed very fast (faster than the heartbeat interval), and consequently each slave ran the map tasks sequentially and the reduce tasks were scheduled after all the map tasks were done.

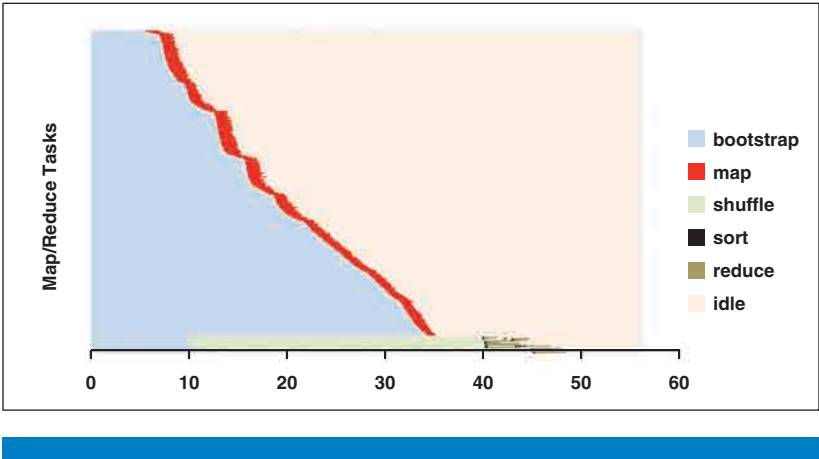


Figure 11: Sorting many small files with Fair Scheduler 2.0
(Source: Intel Corporation, 2012)

To fix this performance issue, we upgraded the cluster to Fair Scheduler 2.0^{[29][30]}, which by default schedules multiple tasks (including reduce tasks) in each heartbeat; consequently the job runs about six times faster (as shown in Figure 11) and the cluster utilization is greatly improved.

Analyzing Application Hotspots

In the previous section, we demonstrated that the high level dataflow execution process of a Hadoop job helps users to understand the dynamic task scheduling and assignment of the Hadoop framework. In this section, we show that the dataflow execution process helps users to identify the data shuffle gaps between map and reduce, and that relating the low level performance activities to the high level dataflow model allows users to conduct fine-grained, dataflow-based hotspot breakdown (so as to understand the hotspots of the massively distributed applications).

Figure 12 shows the runtime behavior of TeraSort. The dataflow execution process of TeraSort shows that there is a large gap (about 15 percent of the total job running time) between the end of map tasks and the end of shuffle phases. According to Hadoop dataflow model (see Figure 7), shuffle phases need to fetch the output from all the map tasks in the copier stages, and ideally should complete as soon as all the map tasks complete. Unfortunately, traditional tools or Hadoop logs fail to reveal the root cause of the large gap, because during that period, none of the CPU, disk I/O, and network bandwidth are bottlenecked; the “Shuffle Fetchers Busy Percent” metric reported by the Hadoop framework is always 100 percent, while increasing the number of copier threads does not improve the utilization or performance.

To address this issue, we used HiTune to conduct hotspot breakdown of the shuffle phases, which is possible because HiTune has associated all the low level sampling records with the high level dataflow execution of the Hadoop job. The dataflow-based hotspot breakdown (see Figure 13) shows

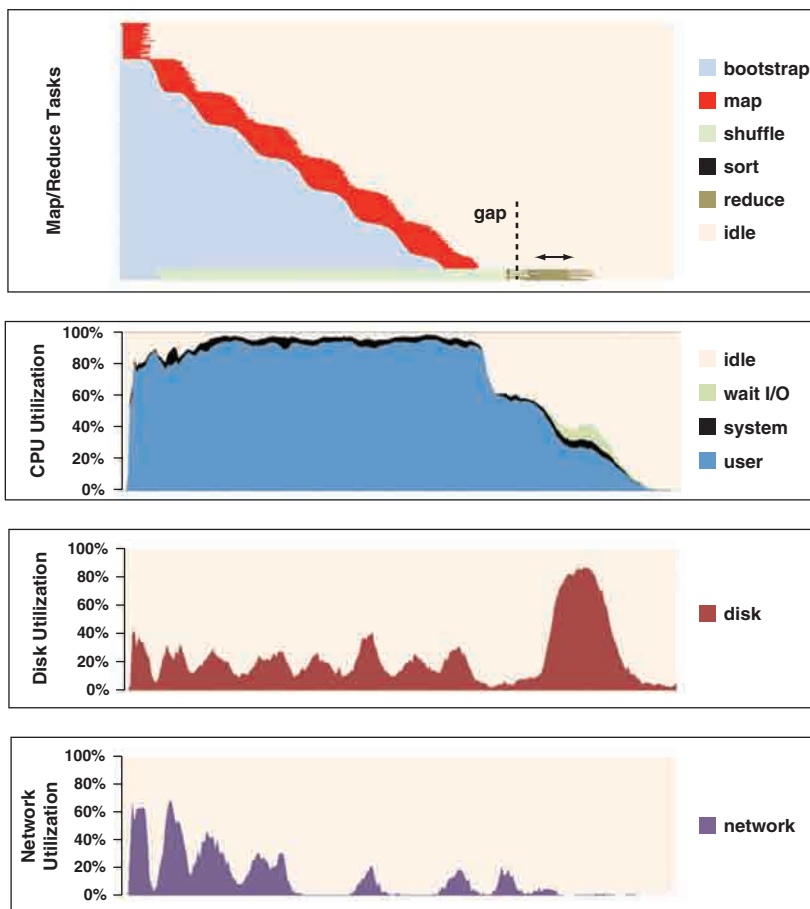


Figure 12: TeraSort (using default compression codec)
(Source: Intel Corporation, 2012)

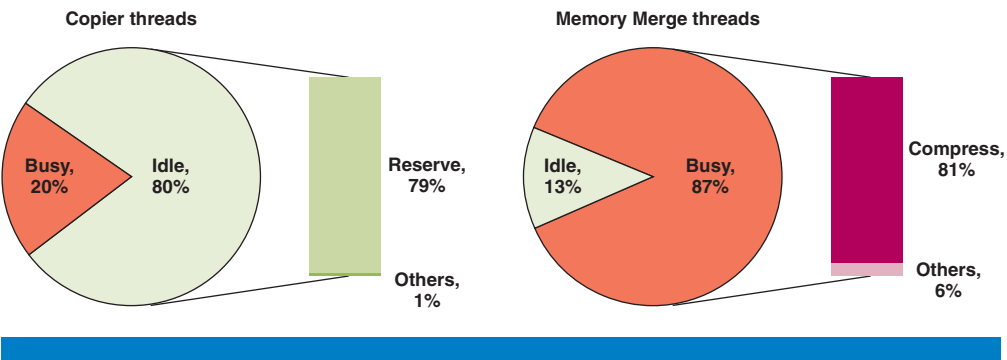


Figure 13: Copier and Memory Merge threads breakdown (using default compression codec)
(Source: Intel Corporation, 2012)



Figure 14: TeraSort (using LZO compression)
(Source: Intel Corporation, 2012)

“Most of the busy time of the memory merge thread is due to the compression.”

that, in the shuffle stages, the copier threads are actually idle 80 percent of the time, waiting (in the *ShuffleRamManager.reserve* method) for the occupied memory buffers to be freed by the memory merge threads. (The idle-versus-busy breakdown and the method hotspot are determined using the Java thread state and stack trace in the task execution sampling records respectively.) On the other hand, most of the busy time of the memory merge thread is due to the compression, which is the root cause of the large gap between map and shuffle. To fix this issue and reduce the compression hotspots, we changed the compression codec to LZO^[31], which improves the TeraSort performance by more than 2x and completely eliminates the gap (see Figure 14).

Big Data Solution Case Study

Having moved beyond its origins in large Web sites, Hadoop is now heading toward ubiquity for big data storage and processing, including telecom, smart city, financial service, and bioinformatics. In this section, we describe the challenges in Hadoop-based solutions for these general big data applications

and illustrate how these challenges can be addressed through a case study of intelligent transportation system (ITS) applications.

Challenges for Big Data Applications

Hadoop has been emphasizing on high scalability (that is, scaling out to more servers) via linear scalability and automatic fault tolerance, as opposed to high (per-server) efficiency such as low latency and high performance. Unfortunately, for many enterprise users, it is almost impossible to set up and maintain a cluster of thousands of nodes, which negatively impacts the total cost of ownership (TCO) of Hadoop solutions.

In addition, the MapReduce dataflow model assumes data parallelism in the input applications, a natural fit for most of the big data applications. On the other hand, for some application domains, such as computational fluid dynamics, matrix computation and graph analysis in weather forecast and oil detection, and iterative machine learning, extensions to the MapReduce model are required to support these applications more efficiently.

Third, some mission-critical enterprise applications, such as those in telecommunications and financial services, require extremely high availability, with complete eliminations of SPOFs (single point of failures) and sub-second failover. Much effort will be required to improve the high availability support in the Hadoop stack.

Finally, for many enterprise big data applications, advanced security support (including authentication, authorization, access control, data encryption, and secure data transfer) is needed. In particular, fine-grained access control support is required for all components in the Hadoop stack.

Case Study: Intelligent Transportation System

In this section, we present a case study of an intelligent transportation system (ITS), a key application in smart cities, which monitors the traffic in a city and provides real-time traffic information. In the city, tens of thousands of cameras and sensors are installed in the road to capture the traffic information, including vehicle speed, vehicle identifications, and images of the vehicles. The system stores this data in real time, and users and applications need to query the data in near real time, typically scanning records within a specific time range.

HBase is a good fit for this application, as it provides real-time data ingestion and query, and high throughput table scanning. However, an ITS application is usually a large-scale distributed system, where cameras and sensors are scattered around the city, and there are edge data centers that connect to these devices through dedicated networks. Existing big data solutions (including HBase) are not designed for this type of geographically distributed data centers and cannot properly handle the associated challenges, especially the slow and unreliable WAN (wide area network) connections among different data centers.

To address this challenge, we extend HBase so that a global table can be overplayed on top of multiple HBase systems across different data centers,

“For many enterprise users, it is almost impossible to set up and maintain a cluster of thousands of nodes.”

“Fine-grained access control support is required for all components in the Hadoop stack.”

“The massive scale, exponential growth and variable nature of big data necessitate a new processing paradigm.”

which provide low latency locality-aware data capturing, automatic failover across different data centers, and a location-independent global view of the application to query the data.

Conclusions

Big data is rich with promise, but equally rife with challenges. It extends beyond traditional structured (or relational) data, including unstructured data of all varieties (text, image, video, and more); it is not only large in size, but also growing faster than Moore’s law (more than doubling every two years). The massive scale, exponential growth and variable nature of big data necessitate a new processing paradigm.

The Hadoop stack has emerged as the de facto standard for big data storage and processing. It originates from big web sites and is now heading toward ubiquity for large-scale data processing, including telecom, smart city, financial service, and bioinformatics. As Hadoop-based big data solutions grow in pervasiveness and scale, efficiently tuning and optimizations of Hadoop deployments, and extending Hadoop to support general big data applications become critically important. Intel has provided many proven methodologies, tools, a solution stack, and reference architecture to address these challenges.

References

- [1] McKinsey Global Institute. “Big data: The next frontier for innovation, competition, and productivity,” May 2011.
- [2] IDC. “Extracting Value from Chaos,” June 2011.
- [3] Teradata. “Challenges of Handling Big Data,” July 2011.
- [4] J. Dean, S. Ghemawat. “MapReduce: Simplified Data Processing on Large Clusters.” The 6th Symposium on Operating Systems Design and Implementation, 2004.
- [5] Hadoop. <http://hadoop.apache.org/>
- [6] S. Ghemawat, H. Gobioff, and S. Leung. “The Google File System,” 19th ACM Symposium on Operating Systems Principles, October 2003.
- [7] A. Thusoo, R. Murthy, J. S. Sarma, Z. Shao, N. Jain, P. Chakka, S. Anthony, H. Liu, N. Zhang. “Hive - A Petabyte Scale Data Warehousing Using Hadoop.” The 26th IEEE International Conference on Data Engineering, 2010.
- [8] Hive. <http://hadoop.apache.org/hive/>
- [9] C. Olston, B. Reed, U. Srivastava, R. Kumar, A. Tomkins. “Pig latin: a not-so-foreign language for data processing.” The 34th ACM SIGMOD international conference on Management of data, 2008.

- [10] Pig. <http://hadoop.apache.org/pig/>
- [11] F. Chang, J. Dean, S. Ghemawat, W. Hsieh, D. Wallach, M. Burrows, T. Chandra, A. Fikes, R. Gruber. "Bigtable: A Distributed Storage System for Structured Data." 7th USENIX Symposium on Operating Systems Design and Implementation (OSDI'06), 2006.
- [12] HBase. <http://hadoop.apache.org/hbase/>
- [13] "Hadoop and HBase at RIPE NCC." <http://www.cloudera.com/blog/2010/11/hadoop-and-hbase-at-ripe-ncc/>
- [14] A. Pavlo, E. Paulson, A. Rasin, D. J. Abadi, D. J. DeWitt, S. Madden, M. Stonebraker. "A comparison of approaches to large-scale data analysis." The 35th SIGMOD international conference on Management of data, 2009.
- [15] S. Huang, J. Huang, J. Dai, T. Xie, B. Huang. "The HiBench Benchmark suite: Characterization of the MapReduce-Based Data Analysis." IEEE 26th International Conference on Data Engineering Workshops, 2010.
- [16] Sort Benchmark. <http://sortbenchmark.org/>
- [17] Nutch. <http://lucene.apache.org/nutch/>
- [18] Mahout. <http://lucene.apache.org/mahout/>
- [19] "A Benchmark for Hive, PIG and Hadoop," <http://issues.apache.org/jira/browse/HIVE-396>
- [20] J. Dai, J. Huang, S. Huang, B. Huang, Y. Liu. "HiTune: dataflow-based performance analysis for big data cloud," 2011 USENIX conference on USENIX annual technical conference (USENIX ATC'11), June 2011.
- [21] S. L. Graham, P. B. Kessler, M. K. McKusick. "Gprof: A call graph execution profiler." The 1982 ACM SIGPLAN Symposium on Compiler Construction, 1982.
- [22] Intel VTune Performance Analyzer. <http://software.intel.com/en-us/intel-vtune/>
- [23] A. Rabkin, R. H. Katz. "Chukwa: A system for reliable large-scale log collection," Technical Report UCB/EECS-2010-25, UC Berkeley, 2010.
- [24] Chukwa/ <http://incubator.apache.org/chukwa/>
- [25] Scribe. <http://github.com/facebook/scribe>
- [26] Flume. <http://incubator.apache.org/flume/>
- [27] Ganglia. <http://ganglia.sourceforge.net/>

- [28] A fair sharing job scheduler. <https://issues.apache.org/jira/browse/HADOOP-3746>
- [29] M. Zaharia, D. Borthakur, J. S. Sarma, K. Elmeleegy, S. Shenker, I. Stoica. "Job Scheduling for Multi-User MapReduce Clusters." Technical Report UCB/EECS-2009-55, UC Berkeley, 2009.
- [30] Hadoop Fair Scheduler Design Document. https://svn.apache.org/repos/asf/hadoop/mapreduce/trunk/src/contrib/fairscheduler/designdoc/fair_scheduler_design_doc.pdf
- [31] Hadoop LZO patch. <http://github.com/kevinweill/hadoop-lzo>

Authors' Biographies

Jinquan (Jason) Dai is a principal engineer in Intel's Software and Services Group, managing the engineering team for building and optimizing big data platforms (e.g., Hadoop). Before that, he was the lead architect and the engineering manager for building the first commercial auto-partitioning and parallelizing compiler for many-core many-thread processors (Intel Network Processor) in the industry. He received his Master's degree from the National University of Singapore, and his Bachelor's degree from Fudan University, both in computer science.

Jie (Grace) Huang is a senior software engineer in Intel's Software and Services Group, focusing on the tuning and optimization of big data platforms. She joined Intel in 2008, after getting her Master's degree in Image Processing and Pattern Recognitions from Shanghai Jiaotong University.

Shengsheng Huang is a senior software engineer in Intel's Software and Services Group, focusing on big data platforms, particularly the Hadoop stack. She got her Master's and Bachelor's degrees in Engineering from Zhejiang University in 2006 and 2004 respectively.

Yan Liu is a software engineer in Intel's Software and Services Group, focusing on Hadoop workloads and optimizations. She holds a Bachelor's degree in Software Engineering and a Master's degree in Computer Software and Theory.

Yuanhao Sun is a software architect in Intel's Software and Services Group, leading the efforts on big data and Hadoop related product development. Before that, Yuanhao was responsible for the research and development on XML acceleration technology. Yuanhao received his Bachelor's and Master's degrees from Nanjing University, both in computer science.

SOME KEY CLOUD SECURITY CONSIDERATIONS

Contributors

Manish Dave

Intel IT, Intel Corporation

Toby Kohlenberg

Information Risk and Security Group,
Intel Corporation

Stacy Purcell

Intel IT,
Intel Corporation

Alan Ross

Intel IT,
Intel Corporation

Jeff Sedayao

Intel IT,
Intel Corporation

As enterprises embark on their cloud computing journey, tactical as well as strategic elements that complement existing security capabilities must be considered. The enterprise must demonstrate legal and regulatory compliance while supporting application and data access via cloud service providers. The first consideration is federation across cloud service providers and disparate environments and frameworks. It is important to share identity and authorization information with partner entities in a reliable, secure, and scalable fashion. A second consideration is the protection of information, not just of an enterprise's intellectual property, but the private information of individuals as well. The shared nature of cloud computing environments demands that steps must be taken to protect privacy. A third consideration is the need to segregate computing environments into trust zones and to deliver the appropriate level of detective, preventative, and corrective controls to match the criticality of the information and applications in a given service level. The final consideration is security business intelligence to support operations, investigations, and forensics. Enterprises must ensure that the right information and predictive analytics are available to support the business. This article will walk through these four elements in order to provide tactical and strategic guidance on some of the extended elements of cloud computing security.

Introduction

Cloud security has been a hot topic for the past several years, garnering attention of enterprises as well as researchers and providers, who seek ways to ensure appropriate risk mitigation and information protection within hosted environments.

Intel IT is no different from other large enterprises in their desire to have flexible and scalable alternatives for computing. We have made excellent strides in our internal cloud computing environment and now look for the most appropriate ways to scale our enterprise environment externally. We are looking at platform, software, and infrastructure as a service models to find the most appropriate usages.

Security remains top of mind within the company and, as some of the people responsible for forward thinking in the security domain, we have spent some time working out the key capabilities needed to make a big move to external cloud service providers.

The purpose of this article is to share some of our current thinking and plans to help make cloud services more palatable from a security perspective. We have

"Security remains top of mind within the company."

a broader strategy and architecture that includes data protection, identity management, and application security, but thought that the four topics represented herein give a glimpse into some areas that we have focus in but others may not. Our purpose is not to give prescriptive guidance but to open up new areas of thought and exploration for others.

Let's look at four key areas of cloud security: federation, anonymization, segmented environments, and security business intelligence.

Federation Across Service Providers and Partners

As compute, storage, and network resources are used across autonomous domain boundaries, policy management frameworks need to be extended beyond single autonomous domains. The manual and cumbersome methods used today to create, negotiate, and manage policies across federated domains are not scaling as the demand for cross-domain resources usage becomes ubiquitous and the de facto standard for computing. Business trends are driving an increased demand for collaboration from anywhere, anytime not only within the enterprise network but externally also. Outsourcing of specific business processes is becoming common as opportunity of moving business flows towards external service providers and partners becomes more feasible from a cost and maturity perspective over time. A large enterprise like Intel federates with hundreds of partners, customers, and suppliers, and the number of new autonomous domains with which to federate increases every year.

The biggest roadblocks for enterprises to enable seamless collaboration with partners and adoption of maturing cloud-based service offerings are related to security and privacy. The enterprise security policy driven by business policies, intellectual property protection, regulatory compliance, and so on, is either not met by the partners or requires a manual, cumbersome process per instance of each partnership. To fill this gap, we see a need to have an automated, simplified way for partners to federate on security and privacy policies for establishment, negotiation, implementation, and audit. Some examples of such *policy domains* and the relationships required for enterprises are with partners, service providers (including cloud), offshore design centers, suppliers, or even organizations within the company (as it may apply to mergers and acquisitions as they happen over time). In this section, we explore how the policy negotiation, federation, agreement, and audits can occur for such a relationship between these policy domains.

Our Framework Design

Our approach has the following components and the high level framework comprised of the following elements:

- *Distributed PBMS across multiple domains.* As the resources used for an application or transaction are across multiple domains whose policy-based management systems (PMBS) are outside of control of a single policy

“Business trends are driving an increased demand for collaboration from anywhere, anytime.”

“Roadblocks for enterprises to enable seamless collaboration with partners and adoption of maturing cloud-based service offerings are related to security and privacy.”

“Even within an autonomous domain, it is not uncommon to have multiple PBMS due to organization boundaries, security requirements.”

“We propose the use of policy services to perform various common functions required to make the multi-domain policy enforcement, verification, and audit possible.”

management authority (PMA), we need to be able to deal with distributed PBMS. Note that even within an autonomous domain, it is not uncommon to have multiple PBMS due to organization boundaries, security requirements, and zones or structures of the infrastructure used.

- *Interfaces that securely expose PBMS functionality.* As we are dealing with cross-domain policy management, the management APIs for these distributed PBMS must be securely exposable across autonomous domains. While this may serve the purpose of some simple forms of federation (commonly used today by cloud providers), by itself, this does not resolve the key issues and gaps for PBMS in a true federated environment.
- *Trusted federation models.* We need to base the cross-domain PBMS trust models on existing and realistic inter-domain relationships. This may require a preexisting trust model or a dynamic setup and teardown of the trust relationship.
- *A services model.* We propose the use of *policy* services to perform various common functions required to make the multi-domain policy enforcement, verification, and audit possible. Services have been successfully used as the abstracted components of a system that can interact with the underlying systems and also provide the functionality needed across individual systems. By creating a common layer using these policy services, we can allow integration and interface with the existing distributed PBMS for management by multiple autonomous policy management authorities.

Some examples of such services are:

- *Policy agreement/negotiation services.* This category of services can be used by multiple trusted PMAs to work together on a policy agreement, negotiation, and other pre-enforcement aspects of creating the cross-domain policies. It requires that domains have a way of securely and programmatically exposing their policy so that applications looking for resources can determine what resources can be used.
- *Policy translation and normalization services.* As the PBMS may not follow a standard interface or policy language, this category of services will perform the function of translating and normalizing the “agreed upon” policies (between two or more PMAs).
- *Policy conflict resolution services.* This category of services can be used to resolve any pre-verification policy conflicts as a validation, as well as work in conjunction with other services such as the policy agreement and translation/normalization service.
- *Policy interpretation services.* This category of services will interpret the translated and normalized policies after going through the pre-verification, such as conflict resolution to interpret the policy agreements (between PMAs), so they can be implemented across the distributed PBMS.
- *Policy audit and verification service.* To provide governance, auditability, and post-verification for the implemented policies on the distributed PBMS.

There are existing federation implementations such as PlanetLab's implementation of GENI's slice-based facility architecture that offer APIs that allow for determining a domain's resource policies. Knowing a policy is only the first step, however, as our design proposes services for not just learning about policies but finding conflicts between domain policies and allowing for negotiations that could resolve those conflicts.

Implementation Considerations

Policy services must be implemented with the required level of assurance and integrity to make trusted interaction possible on top of untrusted layers (analogous to IPSEC implementation on top of unprotected IP Internet) to provide the sufficient level of security and privacy.

The policy services can be hosted by third-party organizations or in a "closed user group" relationship. These services must be available for any ad-hoc, dynamic relationship to be set up and can be potentially used even for complex multi-party relationships. The economic feasibility of such services can provide the incentives for such implementations on the commercial Internet similar to security services implemented today such as network/cloud-based firewall, intrusion detection, and antivirus services hosted in the Internet cloud.

In addition, an organization or enterprise may choose to implement some or all of these policy services within their domain and provide the extensibility for its private federation partners. The key benefit for this proposed service-oriented architecture is that services can be selected as required by the level of trust and relationship to federate the resources. Not all services must be used in a relationship and a few can be used optionally (such as, for example, the post-verification and audit service).

Some examples for security and privacy policies that can be implemented using this model are authentication, authorization, encryption, rights management, and access control. However, the framework can be easily extended to any policy across domains such as resource management, service levels, or even pay-as-you-go policies. For example, a policy requirement for a consumer of infrastructure as a service would be not to exceed certain levels of resource utilization, so as to maintain the maximum usage charges per hour or day.

"The policy services can be hosted by third-party organizations."

"Not all services must be used in a relationship and a few can be used optionally."

Applying the Federation Framework

Now that we have designed an intercompany policy federation we need to find places where our framework can be used. After talking to a number of security architects and other security personnel, we came up with a list of possible scenarios where policy can work. One example: Intel has a supplier called Supplier A who must periodically log into systems inside of Intel to perform maintenance or to debug some problems. Our current approach is to create special limited access accounts every time that a vendor needs to access equipment. An implementation using our framework would allow our

“An audit service and controls at the Intel federation gateway provides detective controls.”

suppliers to use their own identities, saving Intel the time and cost of creating accounts for each access. An audit service and controls at the Intel federation gateway provides detective controls in case the vendor's personnel attempt something inappropriate.

A similar use case might be in a co-development situation, where Intel is developing hardware and software with a partner. The partner needs access to some set of machines at Intel containing newly developed unannounced hardware in order to develop drivers on them. Rather than having to manage the identities and attributes of all those needing access, Intel could use the identities provided by the development partner.

We are currently talking to different architects whose business lines require intercompany connectivity for further access. Our most likely avenue to adoption would be in new intercompany connections, as the cost of transitioning to a new approach that is not yet fully engineered would be very high for existing connections.

Future Work

There is much work to be done with our federation policy framework. As mentioned above, a key step is getting our approach accepted by those implementing intercompany connections. Our approach needs to be broadly understood and accepted in order to justify the investment needed to get it into production.

The next set of work involves improving the current implementation. Our POC implementation is rudimentary and needs to be able to work with a host of different IDAM products, from active directory to LDAP servers. There needs to be an easier, graphical interface for configuring policies. We used SOAP for communicating security policy between federation gateways, but did not implement security features like message signing. In addition, other companies might want to implement other security policy protocols like SAML or XACML.

We have created a framework for policy federation that can enable Intel to be more agile in establishing intercompany collaboration and do so at lower setup and management cost. Our next steps are to find the best fit for our approach and to improve its implementation to enable a more agile and efficient Intel.

Anonymization

“Anonymization: The act or process of making anonymous, of hiding or disguising identity.”

Enterprises need to keep data about people and other enterprises but need to maintain confidentiality about particular parts. Not only are enterprises concerned with protecting intellectual property and other proprietary information, but there are also regulatory and legal considerations,

“Enterprises need to keep data about people and other enterprises but need to maintain confidentiality about particular parts.”

particularly concerning personally identifiable information. Enterprises like Intel would like to use public cloud computing, but they must make sure that their data is secure.

Anonymization is one potential answer to privacy concerns and for cloud computing security. By obscuring key pieces of customer data and other confidential data, privacy could be maintained while still having usable information for processing. Anonymized data could be put in the cloud (or elsewhere) and processed without having to worry if other people captured that data. Later, the results could be collected and matched back to the private data in a secure area.

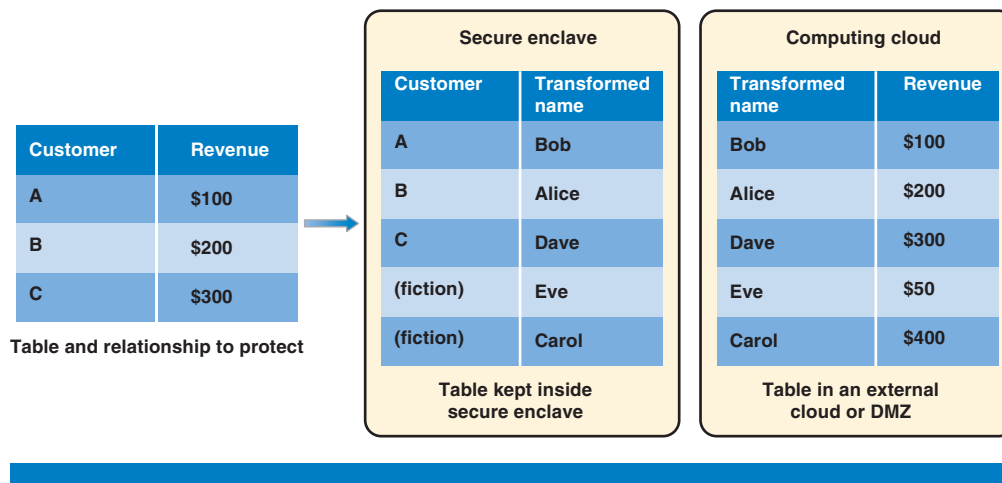


Figure 1: Using anonymization to do safe computing in the cloud
(Source: Intel June, 2012)

We show how such a process could work in Figure 1. We want to calculate total revenue but don't want to expose company names associated with that revenue. We want to make sure that the customer/revenue relationship is kept private even if the data in the cloud is completely compromised. Total revenue can be calculated in the cloud and corrected internally by subtracting off fictitious company amounts. Our approach even blocks some kinds of data mining attacks, as by adding fictitious data, we make it impossible to calculate things like the number of customers, companies with the most revenue, and companies with the least revenue.

Furthermore, as users traverse the compute continuum, devices increasingly implement GPS capabilities, and as location-oriented social media (such as Foursquare) becomes more common, location tracking becomes more of an issue. Location information can be very useful for providing customized and localized service, but the storage and mining of location data has privacy and possibly regulatory issues. Anonymization will become an emerging issue in that space. Anonymization can be a tricky process, and can have severe consequences if not done properly.

“As users traverse the compute continuum, devices increasingly implement GPS capabilities, and as location-oriented social media.”

The following sections go over commonly used anonymization techniques and planned next steps.

Anonymization Techniques

In our discussion of anonymization concepts, we talked about things like “obscuring data” for quasi-identifying attributes. Exactly how can we “obscure” data? This section discusses these obscuring techniques. We start by describing some example log data. That log data will be modified as an example of different anonymization techniques. For each technique, in addition to showing an example, we will describe the most appropriate uses for it. The section concludes with an anonymization example and a list of different techniques used to create it.

With hiding a value is replaced with a constant value (typically 0). Sometimes it is called “black marker.”

“Hiding is useful for suppressing sensitive attributes that may not be needed for processing.”

Hiding is useful for suppressing sensitive attributes that may not be needed for processing. If we needed to publish data for a phone directory, information like salary would not be needed.

A hash function maps each incoming value to a new (not necessarily unique) value. It is often used to map a large, variable amount of data into a number of a certain length.

Permutation maps each original value to a unique new value. This technique allows us to map back the new value back to the original value, provided we store the mappings. This property of permutation lets us do processing in the cloud using permutation-changed values, and then map the results to our original values in a secure place.

Shift adds a fixed offset to the numerical values.

Shift conceals data while letting us do computations in areas like the cloud. We can subtract out the effects of shifts in a secure area.

Enumeration maps each original value to a new value so that ordering is preserved. This allows us to perform studies on the data involving ordering.

Updated is not really a technique for anonymization, but it is a closely related. Updated data has checksums recalculated for the log file changes.

With *truncation*, a field is shortened, losing data at the end. This allows the ability to hide data but keep the information that the data is part of a group.

“Education about anonymization uses and tradeoffs (particularly with use in clouds) needs to be an ongoing process.”

Next Steps

It is clear that more work that needs to be done with anonymization. The potential pitfalls of anonymization do not seem to be well known, so education about anonymization uses and tradeoffs (particularly with use in clouds) needs to be an ongoing process. There are problems with available tools for doing anonymization. Available open source tools seem capable but are not well

documented, and some have been abandoned by their creators. Our POC used AES encryption to do anonymization. AES-NI instructions could be used to speed up the anonymization process, as a potential use case for AES-NI is secure use of public clouds.

Environment Segregation

The goal and theory of general cloud environments is to enable one large seamless, transparent computing space where system owners pay no attention to the infrastructure or underlying environment. While this works in situations where all the customers require a consistent level of security, most enterprises require varying levels of security depending on the sensitivity and importance of the data and computation being performed. In order to enable enterprises to take advantage of the public cloud environments there needs to be some implementation of segregated environments with distinct controls and security requirements that match the needs of the enterprise. This is also necessary to fully enable the concept of policy domains that we previously discussed.

In order to support the varying security requirements that an enterprise needs and that may be defined by different policies in a federated cloud we have implemented the concept of multiple *trust zones*. Each trust zone is an environment that is designed to meet the security requirements for a specific class of data and computing. A common cloud environment can be thought of as low trust. It has default controls to protect the infrastructure but no restrictions on the applications or virtual machines that run within the cloud. For the purpose of this discussion, we will focus on our work implementing a High Trust Zone (HTZ).

Key Risks

In developing the concept of the HTZ, we identified specific risks that must be addressed:

- *Security of the infrastructure management systems.* While the hypervisor frequently has the most focus from an attack perspective, we have found that the management systems are often more vulnerable and more useful for attackers to compromise.
- *Security of the hypervisor.* The hypervisor itself must be understood to be nothing more than another operating system. As such it must be secured and monitored for indications of compromise.
- *Security of the guest VMs.* Because no hypervisor is perfect, any vulnerability in a guest VM must be considered additive to the overall vulnerability of the environment. Compromise of one VM makes it easier to compromise the hypervisor and the rest of the cloud.
- *Security of the applications running in the guest VMs.* In the same vein as the security of the guest VMs, the applications running in the cloud must be considered as part of the attack surface for the overall environment.

“Most enterprises require varying levels of security depending on the sensitivity and importance of the data.”

“No hypervisor is perfect, any vulnerability in a guest VM must be considered additive to the overall vulnerability.”

“The HTZ uses 24 administrative controls.”

Key Controls

To reduce the consequences of these identified risks, the HTZ uses 24 administrative controls. We are implementing these controls in three phases, two of which are complete. The first phase used controls to isolate the virtualization management infrastructure from the servers being virtualized and to protect the accounts used to manage virtualization. The second phase established controls for extensive security monitoring, taking a holistic approach that included developing deep logging capabilities, and solutions for monitoring the management agents. For the third and final phase, we are adding complex network monitoring that includes a diverse mix of host and network intrusion detection capabilities. We have begun deploying our HTZ architecture and virtualization process in multiple data centers. Our HTZ solution:

- Takes advantage of virtualization and cloud computing to improve the agility and efficiency of our high security-sensitive applications
- Provides controls we can apply to our Internet-facing environments to further reduce their risk, as well as improve their agility and operating efficiency
- Prepares us to take advantage of public cloud services for internal and external facing applications in the future

Extending Trust Zones to Cloud Service Providers

As we move toward external cloud computing environments it will be important for us to communicate our key risks and controls to the cloud service providers to help ensure that they are meeting our basic requirements for hosting applications, services, and data in different trust levels. This is critical in terms of extending our enterprise cloud computing capabilities outside of Intel and should help us with overall security for cloud service providers. Our goal is that we can ultimately ensure that applications and information are protected independent of where they are being hosted.

“Our goal is that we can ultimately ensure that applications and information are protected independent of where they are being hosted.”

Federated Security Business Intelligence

Security Business Intelligence (SBI) is Intel IT's collection of services that is designed to capitalize on our detective controls: logging, alerts, threat intelligence, and so on. We have been working on SBI for the past decade and continue to grow our capabilities internally while looking to what will be needed externally in terms of information gathered from other sources (that is, cloud service providers, external threat intelligence services, and so on).

SBI Architecture

Our SBI architecture is based around four key business objectives: keep Intel legal, keep information available, keep information protected, and keep controls cost effective. These are drivers that the business has articulated from a threat management, investigations, and forensics perspective.

We also realized early on that it is a combination of people, processes, tools, and applications that would define an enterprise SBI service offering. With that in mind we determined the need to create business process applications, analytics applications, business intelligence capabilities along with an information infrastructure.

Foundationally we focus on layers to deliver the SBI architecture: the first layer is that of a common logging service, which facilitates the aggregation and normalization of multiple event sources and streams. Normalizing data is important because different event sources provide different event/log formats, structures, and time stamps. Aggregation is the ability to put all of these normalized event sources in one repository so they can be accessed and analyzed.

The second layer of the architecture is the correlation engine, which is used to write rules that allow us to correlate events across different sources in the common logging service. Rules are written that will allow us to detect new threats, correlate evidence of attacks or attempted intrusions, and help facilitate investigations. The correlation engine is the heart of the SBI architecture because without it we would end up looking through billions of events on a daily basis and trying to determine which ones were real, relevant, and meaningful. The correlation engine provides the capability to funnel down the raw event streams into actionable chunks.

The third layer of the architecture is the predictive analytics piece. This is where specific risk models are written and data is mined to seek evidence of these risk models in the environment. Predictive analytics are very useful for us to look for specific events or the evidence of activities against high value asset repositories where critical data is stored in the enterprise.

The instantiation of these architectural layers has been a multiyear process and continues to evolve with time. Each layer requires specific expertise on the part of analysts and engineers who work with security information. This foundation is very scalable within the enterprise and has helped us articulate and build a number of underlying services.

SBI Services

The underlying services that comprise the SBI solution each play a role in insuring the timely availability of current and older information that has been collected through our detective control sources.

Transaction data services are provided via an extract, transform, and load (ETL) process from one repository to the common logging service. This includes, but is not limited to, things like access logs from applications, firewalls, VPN services, and servers.

Context data services are another example of an ETL process that is used to provide relevant information related to employee data, IP address to location, time information as well as hardware asset information and software applications in use within the environment.

“Normalizing data is important because different event sources provide different event/log formats, structures, and time stamps.”

“Each layer requires specific expertise on the part of analysts and engineers who work with security information.”

“Analysis and mining services are process cubes that include measures, dimensions, and key process indicators.”

“Without an SBI system in place it is difficult to facilitate the risk management process.”

Stream data services are a third example of an ETL process but they are related to streams of log data that comes in near real time (that is, network intrusion sensor alerts) and then are put into the common logging service where custom business logic is used to filter, normalize, and aggregate these events.

Analysis and mining services are process cubes that include measures, dimensions, and key process indicators. These services are used to run model queries, to execute common data mining tasks, to fill reporting databases, and to capture metrics from all other service components.

Reporting services are the primary interface for customers and stakeholders to consume SBI information. These consist primarily of web reports, ad hoc queries through SQL, cube, and drilldown views along with web service endpoints. Reporting services also include dashboards and decision support tools along with data mining viewers.

Key Use Cases

We continue to explore new use cases internally and have monthly meetings where stakeholders can propose new use cases, but our SBI program was built on some primary use cases.

First, we monitor for at-risk behaviors: calculating the components of the “who, what, when” risk equation and triggering the follow-up processes for patterns that meet a certain threshold. The risk models associated with this use case are constantly evolving as we seek new ways to identify threats and intrusion attempts.

A second use case is the collection of incident data. Once we know that an incident has occurred, we need to be able to pull all of the relevant data based on an identity, asset, or resource. This is typically accomplished using our analysis and reporting services to examine events within our common logging framework.

Once we have collected any available incident-related data, the analysis process kicks off to assess the impact of an incident and to recommend any corrective or preventative actions that should be taken based on the type of incident.

We also have a use case for managing security configuration within the environment that is supported by our SBI services. This is the process of keeping hardware and software configuration in line with our security requirements and reporting when there have been excursions from policy.

Another key use case is the ability to use SBI systems to quantify risk of a threat or scenario, analyze the risk and potential impact, rank the risks seen in the environment, and mitigate risks or make recommendations on steps to mitigate and when we should accept risk in the environment. Without an SBI system in place it is difficult to facilitate the risk management process.

Federation of SBI

As we move to external cloud-based services it becomes critical that we are able to take information learned from external sources and include it with our

internal security business intelligence repositories. We are just embarking on this journey but know the types of information we will be seeking along with some key elements that will help make our ability to externalize applications and services possible.

The first step will be to have near real-time event streams of any log information coming from a service provider back to our common logging service. This is necessary for us to identify threats and intrusion attempts against our resources whether they are inside Intel, at a partner's site, or at a cloud service provider.

A second objective is to be able to aggregate and correlate real-time threat information with our SBI system so that we can execute the risk management process. We are seeking ways to facilitate the communication of threat management information to Intel from external providers.

A third objective in the future will be to tag information based on where it is coming from (outside Intel or inside) and be able to associate a confidence level in the information over time. This will allow us to make better risk management decisions and more proactively deal with incidents and investigations.

The combination of internal and external data will be a powerful tool for Intel going forward and through the maturity of anonymization techniques we see a time when we can share information across companies and entities to raise the overall bar of information security around the world.

Conclusions and Next Steps

The path to a new compute model can be long and complicated—this has been true with our journey to cloud computing. Intel IT has made great strides with our internal cloud computing environment and is now embarking on the process of extending to external service providers.

As we move down the external path there are several security considerations that need to be incorporated and we believe that tackling the four topics in this article will go a long way in enabling our computing needs for the future.

“The path to a new compute model can be long and complicated.”

Biographies

Manish Dave is a senior network engineer in Intel IT and has been with Intel for over 10 years. He is always looking to improve our network security and is very passionate about technical leadership. He can be reached at manish.dave@intel.com

Toby Kohlenberg is a senior information security specialist in Intel's Information Risk and Security Group. Toby loves to analyze systems, networks, and applications to identify weak or vulnerable points. He has been with Intel for over 15 years and is a regular public speaker on security topics. Toby can be reached at toby.kohlenberg@intel.com

Stacy Purcell is a senior security architect who has been with Intel for over two decades, focused on networking and security. He is very passionate about Security Business Intelligence and can be reached at stacy.p.purcell@intel.com

Alan Ross is a senior principal engineer who leads up Intel IT's security architecture and technology development team. Alan is passionate about new security models and ways to attack old problems. He has been with Intel for over a decade. Alan can be reached at alan.d.ross@intel.com

Jeff Sedayao is a researcher in Intel IT's Labs group, focused on cloud security and manageability. Jeff has been with Intel for over 20 years and has played a variety of roles in network and security areas. He often speaks publicly and continues to drive innovative thought in research and technology development. He can be reached at jeff.sedayao@intel.com

IMPLEMENTING CLOUD STORAGE METRICS TO IMPROVE IT EFFICIENCY AND CAPACITY MANAGEMENT

Contributor

Sudip Chahal

Intel IT
Intel Corporation

Krishnamurthy Anandarao

Intel IT
Intel Corporation

Chris Peters

Intel IT
Intel Corporation

Ajay Chandramouly

Intel IT
Intel Corporation

Shane Healy

Intel IT
Intel Corporation

Nigel Wayman

Intel IT
Intel Corporation

Steve Owen

Intel IT
Intel Corporation

Like many of our enterprise IT peers, we are being challenged with rapid growth in storage demand. In 2011 alone, we faced the need for a 53-percent increase in storage capacity to 38.2 PB from 2010 and have been averaging 35-percent year-over-year growth since 2007. The continued build-out of our office and enterprise private cloud could further increase demand and complexity of our storage environment. Clearly we could not increase costs linearly with demand. Through a variety of techniques described here, including implementation of capacity metrics, improving utilization, thin provisioning, and tiering, we have been able to support significant capacity and performance improvements while saving 9.2 Million US Dollars.

These techniques allow us to meet steep storage demand growth in a cost-efficient manner while not compromising on quality of service in our virtualized and multitenant computing environment.

Introduction

As we migrate to the private cloud, we seek to optimize the efficiency of the storage area network (SAN) environment by increasing storage capacity utilization. We have been further challenged by the simultaneous demand for storage capacity increasing unabated by 35 percent year over year since 2007. As a result, capacity purchases represent a significant and growing business cost. By increasing storage utilization, we can meet storage demand growth while not increasing costs linearly.

We are evaluating a wide array of technologies and approaches to improve SAN efficiency and utilization over time, including thin provisioning. Since many customers use only a small portion of their storage allocation, thin provisioning technology allows IT to over-allocate capacity and thereby increase utilization of storage resources. However, over-allocation also requires improved capacity and risk management so that we have adequate capacity to meet customer requirements.

We therefore recognized a need for enterprise-wide storage capacity metrics that accurately measure storage utilization and allocation across our SAN environment. Over the past year, we have developed and begun refining a storage metrics methodology that could span both our new private cloud and our traditional enterprise environment.

These metrics focus on three related areas: efficiency, capacity management, and risk management. We are building business intelligence and reporting

capabilities based on these metrics. Our storage capacity metrics are designed to:

- Establish a clear link between cost (the purchase of raw capacity) and value (the use of that capacity to store internal customer data).
- Reflect efficiency gains due to new technology adoption and storage efficiency approaches such as thin provisioning and data de-duplication.
- Establish operational thresholds, based on allocation and utilization levels, that alert us when we need to add or reallocate capacity before we jeopardize customer service levels.
- Remain uniform over time, providing us with a consistent view of efficiency across supplier product lines and technology generations.

We are using these metrics to analyze and compare the efficiency of our private cloud at multiple levels, from individual storage pools to a global view.

Custom reports provide an efficiency overview for senior IT managers, help data center managers plan capacity purchases, and enable operations engineers to respond to day-to-day capacity requests.

Potential future enhancements include customized risk management thresholds based on a predictive algorithm that can consider factors such as historical growth in customer storage requirements. We also anticipate being able to apply our metrics across our traditional office and enterprise network attached storage (NAS) environment.

As we scale our storage environment and increase utilization, performance health metrics will be essential to understand application and system requirements, identify potential bottlenecks, and assist in successful deployment. Accordingly, we are in the process of incorporating storage performance health into our capacity management and efficiency views.

Business Challenge

Intel IT is undertaking a major transition to an enterprise private cloud that will support our office and enterprise computing applications. This phased, multiyear initiative is designed to enable greater agility and efficiency.

We are building this multitenant environment on virtualized infrastructure as a service (IaaS). This infrastructure is based on clusters of Intel® Xeon® processor-based servers accessing shared pools of storage through storage area networks.

As we transition to the private cloud, SAN storage is becoming an increasingly important element of Intel's IT storage infrastructure. Currently, Intel has a total about 50 PB of global storage; Intel's office and enterprise environment accounts for about 13 PB of this. Within this office-enterprise environment, SAN storage is used to support our enterprise private cloud as well as mission-critical and other applications.

“Custom reports provide an efficiency overview for senior IT managers, help data center managers plan capacity purchases.”

“Infrastructure is based on clusters of Intel® Xeon® processor-based servers accessing shared pools of storage.”

“The cost of SAN storage is becoming a significant factor in overall IT costs.”

The cost of SAN storage is becoming a significant factor in overall IT costs because SAN storage is a more expensive technology than locally attached storage while a larger proportion of our storage is becoming SAN based. Key factors in SAN capacity growth include server virtualization and optimizing SAN capacity utilization.

Server Virtualization

To build the foundation for our private cloud, we began accelerating the pace of server virtualization in late 2009. This drives incremental demand for SAN storage, because data that was previously stored on physical servers' local drives is now migrated to the shared SAN environment. We are currently 68 percent virtualized as measured by OS instance and are on-track to be 75 percent virtualized by the end of this year.

We are experiencing a rapid growth in storage demand with a 35 percent annual growth rate in raw data requirements due to trends such as increasing use of video and other graphical information, legal retention requirements, larger databases for operational and business intelligence use, and backup needs.

The Value of Optimizing SAN Capacity Utilization

With our increasing use of SAN storage, we saw the need for strategies and tools that would enable us to measure and optimize efficiency by maximizing SAN capacity utilization. Increasing utilization can reduce cost by helping to curb growth in the amount of capacity we need to purchase.

“Due to the large size of the Intel IT environment, even small efficiency improvements can result in significant financial savings.”

Due to the large size of the Intel IT environment, even small efficiency improvements can result in significant financial savings because capacity purchases are the largest factor in Intel IT storage costs. Other costs such as operating costs, power and cooling, SAN switches, and cabling represent a much smaller percentage of overall total cost of ownership (TCO).

The Need for Enterprise-wide Metrics

To manage capacity and improve efficiency, we need to be able to accurately measure capacity utilization across our SAN environment. To date, this has been challenging for several reasons:

- Storage suppliers often use different terminology and methodologies for measuring capacity utilization. It is difficult to combine information from multiple suppliers into a single view that reflects Intel IT business requirements.
- Existing metrics do not accurately reflect efficiency improvements achieved using new technologies such as thin provisioning.
- Due to the complexity of the environment and the lack of a standard method for measuring storage utilization, we have found inconsistencies in the way we measure efficiency across different groups and teams within Intel IT.

We decided to implement a single set of metrics that provides standard measures of utilization across our diverse and constantly changing technology base.

Strategy Focus Areas

These enterprise-wide metrics must account for three interrelated focus areas within our storage management strategy: efficiency, capacity management, and risk management. The relationship between these three areas and their importance to different groups within Intel IT are described below and shown in Figure 1.

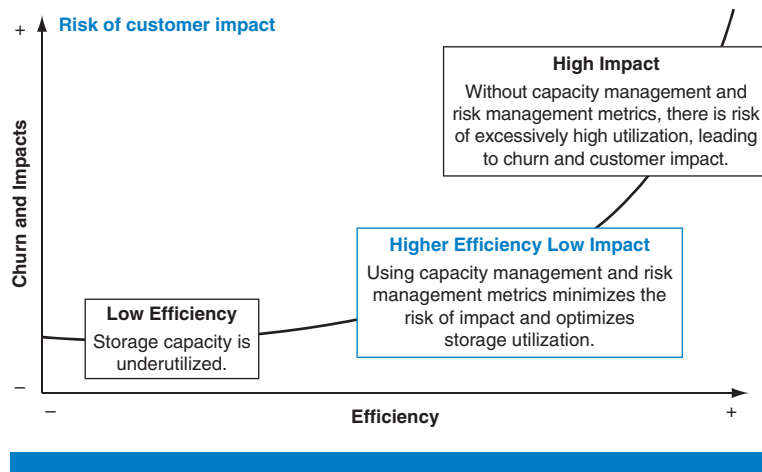


Figure 1: Metrics—efficiency, capacity, and risk management

(Source: Intel Corporation, 2012)

Efficiency

Clearly, our goal is to achieve high storage capacity utilization without impacting our internal customers or increasing IT operational costs. Storage efficiency metrics provide visibility into how efficiently our storage capacity is being used to store customer data. Efficiency metrics are particularly interesting to Intel IT executives and senior managers, who view storage efficiency as a key IT indicator.

Capacity Management

As we increase efficiency and utilization, capacity management becomes increasingly important. We need to be able to analyze utilization growth rates and accurately anticipate and determine when we need to purchase more capacity. This enables us to make purchases in the most timely and cost-effective way.

Capacity management metrics are used by a variety of people within Intel IT. These include the data center managers, capacity planners, and operations engineers who are responsible for making sure there is enough SAN capacity to meet demand.

“Storage efficiency metrics provide visibility into how efficiently our storage capacity is being used to store customer data.”

“Running out of capacity can result in an unacceptable disruption to business applications.”

Risk Management

As we drive storage efficiency and utilization to higher levels, there is an increasing risk of customer impact. In an extreme case, running out of capacity can result in an unacceptable disruption to business applications. Very high storage utilization can also impact application performance.

There is also the potential for increased IT operational costs. To avoid customer impact, Intel IT operations engineers may devote a considerable amount of time to manually migrating data from heavily utilized to less-utilized pools.

These potential issues mean that as we improve efficiency, we also need to develop metrics that enable us to manage risk and have capacity available when needed. These include capacity utilization and allocation thresholds that can be used to trigger alerts before utilization reaches unacceptably high levels.

Risk management metrics are of primary interest to Intel IT storage operations engineers, who are responsible for making capacity available to support day-to-day business needs.

Improving Storage Efficiency with Thin Provisioning

We are applying a number of approaches to reduce cost and curb storage growth by utilizing capacity more efficiently. These include thin provisioning, reclaiming unused storage, and sharing storage infrastructure across multiple projects.

Thin provisioning is a key technology, especially within our private cloud. The concept is analogous to the way that airlines overbook flights. As shown in Figure 2, we can oversubscribe, or over-allocate, capacity to provide customers with adequate headroom, based on the assumption that they typically ask for more capacity than they need. The effect of oversubscription is to increase capacity utilization and enable us to avoid purchasing more capacity than customers actually need.

Because we are oversubscribing available capacity, our storage capacity metrics must be able to measure the efficiency benefits achieved with thin provisioning and help us manage capacity and risk.

We are widely deploying thin provisioning to increase efficiency across our private cloud storage environment.

We are also exploring the use of data de-duplication, which recognizes when multiple instances of the same data are stored by different customers. The technology consolidates the multiple data sets into a single copy, freeing space that can then be allocated for other use.

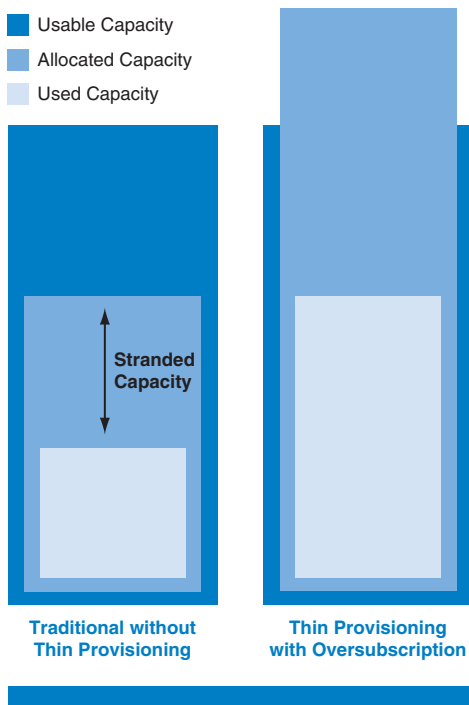


Figure 2: Thin provisioning enables oversubscription
(Source: Intel Corporation, 2012)

Storage Metrics

At the core of this framework is a single set of capacity utilization metrics that we can apply across our entire SAN environment to support efficiency, capacity management, and risk management. In the future, we also anticipate being able to apply this approach across our traditional office and enterprise network attached storage (NAS) environment.

We designed our metrics so that we can aggregate data to create different views that enable us to analyze storage use and efficiency across the entire storage environment. We can analyze the metrics data by data center, storage tier, and storage frame. In the future, our goal is to show by individual customers.

Efficiency Metrics

We defined three core metrics that measure the storage efficiency of different aspects of the environment. We believe that these metrics represent a minimum set necessary to measure the efficiency of storage capacity use and thereby help contain cost. The metrics are orthogonal, or independent; to maximize efficiency throughout the environment, we need to optimize all three:

- *Slot Utilization.* The ratio of storage frame slots that are populated with drives compared to the total available storage frame slots.
- *Overall Storage Efficiency.* The ratio of customer stored data compared to the raw storage capacity.
- *Low-Cost Storage Percentage.* The ratio of customer data stored on our low-cost storage tier compared to the total customer data stored.

Slot Utilization Percentage

Frames are expensive and occupy costly data center capacity. Therefore, efficient utilization of frame capacity is an important factor in minimizing overall storage cost.

Our measure of frame capacity utilization is the Slot Utilization Percentage metric. Each frame includes a manufacturer-defined maximum number of slots. Intel IT storage operations groups may restrict the number of these slots that are available for use to provide good performance for Intel workloads. As shown in Figure 3, we define Slot Utilization Percentage as the percentage of these available slots that are populated with drives.

“The metrics are orthogonal, or independent; to maximize efficiency throughout the environment.”

“Efficient utilization of frame capacity is an important factor in minimizing overall storage cost.”

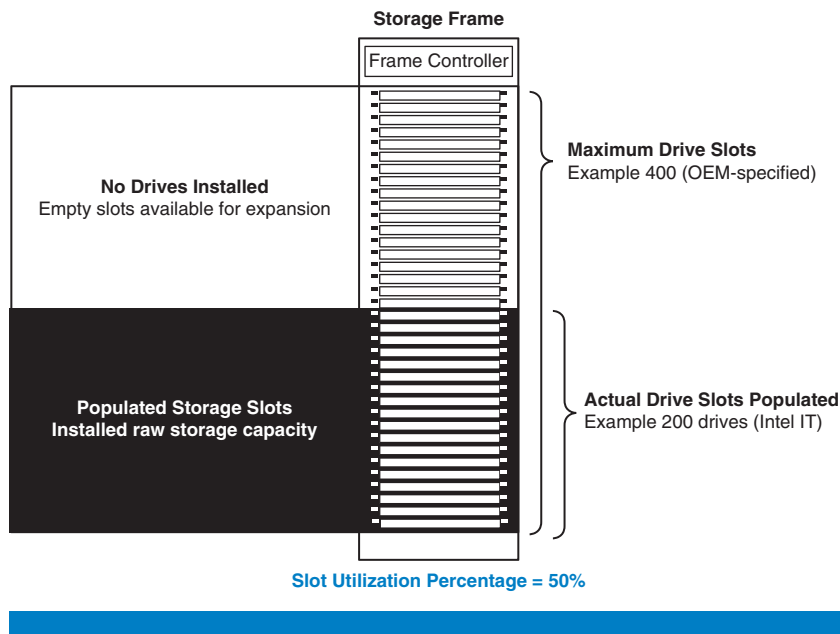


Figure 3: Slot utilization
(Source: Intel Corporation, 2012)

Overall Storage Efficiency

The goal of our storage strategy is to store data required by our internal customers in a cost-effective manner. Therefore, we defined a metric that measures how efficiently storage capacity is utilized for storing customer data. This metric, Overall Storage Efficiency, is defined as the ratio of stored customer data (Used Capacity) to raw storage capacity (Raw Capacity), as shown in Figure 4. Improving Overall Storage Efficiency is a primary focus of our storage strategy.

“Measuring Used Capacity presents some complexity for IT organizations because of data duplication.”

Measuring Used Capacity presents some complexity for IT organizations because of data duplication (multiple instances of business data stored by one or more customers) and inefficient management of orphaned data (stored data that is no longer accessed by customers). The challenge is that the customer’s view of the amount of data stored—the Used Capacity—may differ from

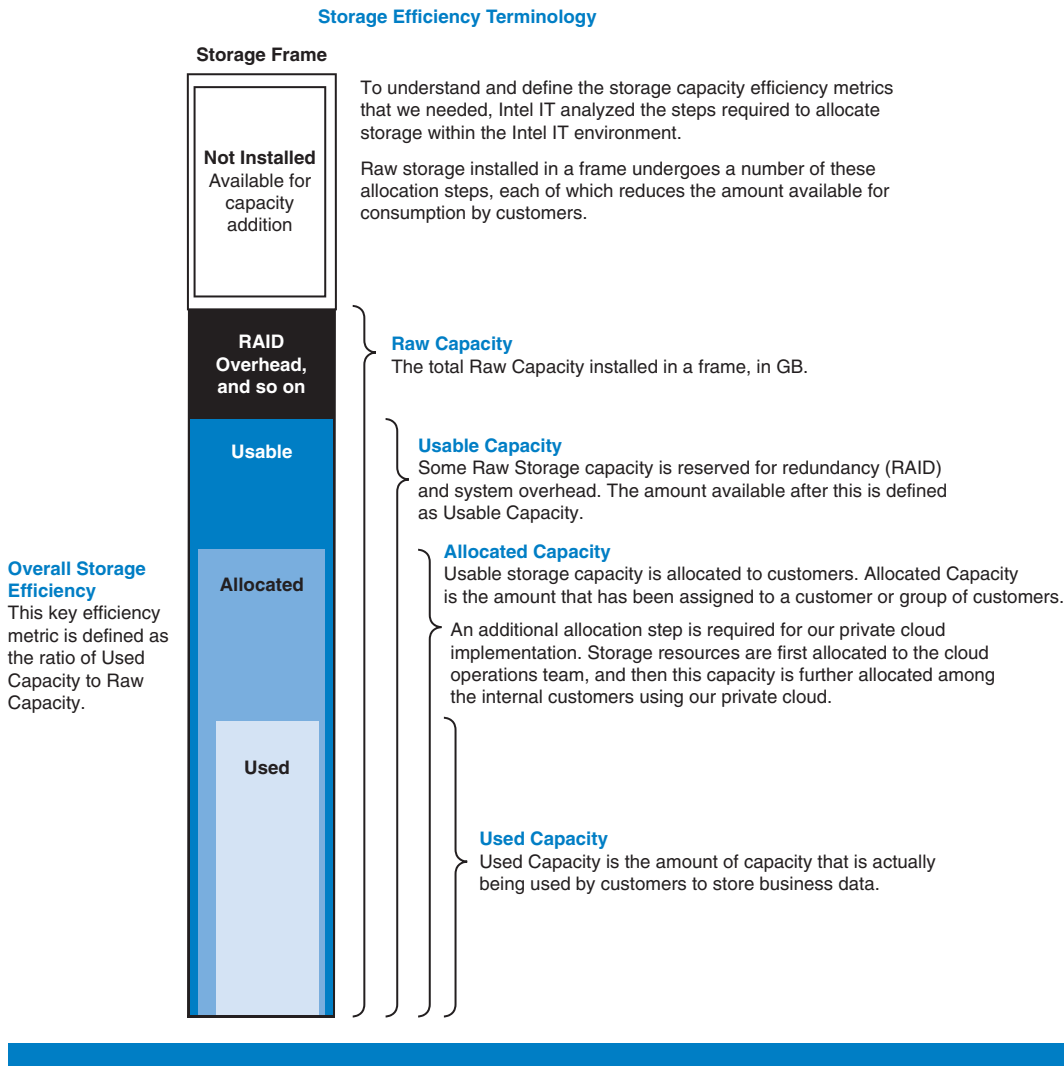


Figure 4: Storage efficiency metrics
(Source: Intel Corporation, 2012)

the storage frame view. As we evaluate data de-duplication technology, we are initially focused only on the storage frame view of Used Capacity as a reasonable approximation for customer stored data.

Low-Cost Storage Percentage

This represents the percentage of customer data that is stored on the lowest-cost tier, which is currently our M2 tier. Intel IT is planning storage tier initiatives with the goal of moving less critical business data to lower cost tiers while continuing to support customer needs.

“The metrics measure allocation and utilization relative to Usable Capacity, not Raw Capacity.”

Capacity Management and Risk Management Metrics

Some storage efficiency techniques, such as thin provisioning, increase utilization but also increase the risk of having insufficient capacity to meet growing and changing business needs. If we oversubscribe capacity (allocate more than is actually available), it becomes even more important to carefully monitor how much of this capacity is being used, how quickly usage is growing, how much more capacity we can safely allocate, and when we need to add or rebalance capacity.

Our key capacity management and risk management metrics are designed to support these requirements. The metrics measure allocation and utilization within each storage pool, but can also be aggregated to provide summary views at the frame and data center level. The metrics measure allocation and utilization relative to Usable Capacity, not Raw Capacity, since this is what matters to the IT operations team and our customers. The metrics are shown in Figure 5 and described below.

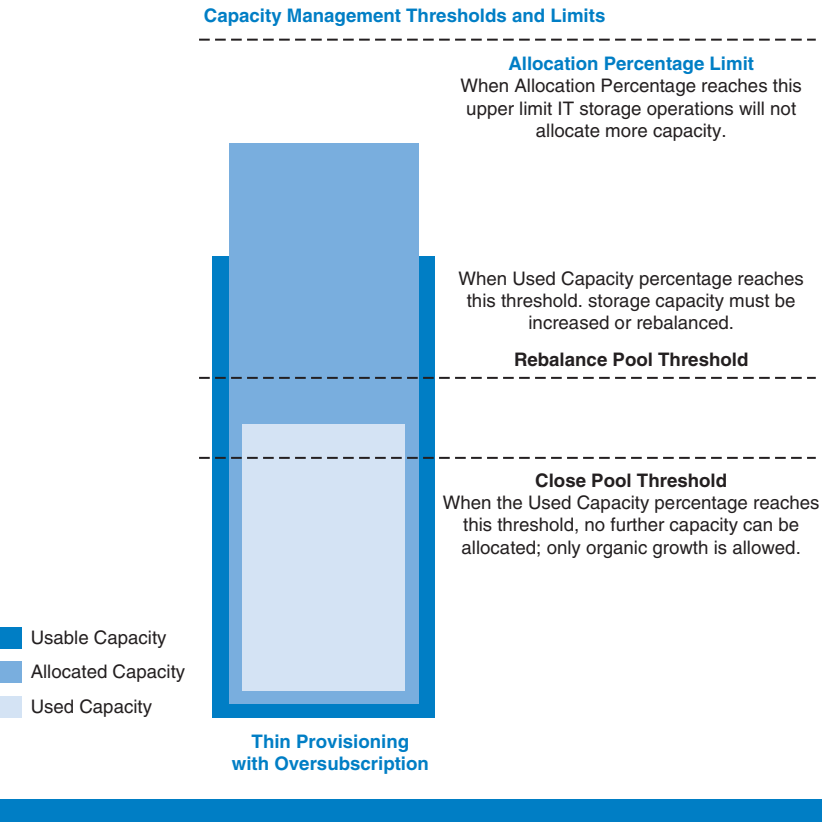


Figure 5: Capacity and risk management metrics
(Source: Intel Corporation, 2012)

To minimize risk, we need to be alerted when allocation or utilization levels reach predefined thresholds. We have established several initial policy-based thresholds, as described below. We expect to refine these policies as we gain more experience in production. For example, we anticipate being able to incorporate better forecasting of customer utilization behavior and trends in storage demand growth, enabling us to include “time to full” projections. We also anticipate including performance health monitoring.

“To minimize risk, we need to be alerted when allocation or utilization levels reach predefined thresholds.”

Used Percentage

This is the percentage of Usable Capacity that is used to store customer data. It is defined as the ratio of Used Capacity to Usable Capacity. By definition, this metric can never exceed 100 percent. We currently define two operational thresholds to manage risk as storage pools reach high utilization levels:

- *Close pool threshold.* When a pool’s Used Percentage reaches 50 percent, the pool is considered closed. No further capacity can be allocated from this pool; only organic growth of existing customer storage workloads is allowed.
- *Rebalance pool threshold.* When the Used Percentage for a pool reaches 70 percent, storage capacity must be reassigned to other pools to avoid the risk of customer impact.

Allocation Percentage

This is the percentage of Usable Capacity that has been allocated to customers. With thin provisioning, this metric can exceed 100 percent. We have established an allocation limit of 150 percent. Once a pool’s Allocation Percentage reaches this level, IT operations will not allocate more capacity, regardless of how much of the capacity is actually being used by customers.

Customer Utilization of Allocated Capacity

This represents how much specific customers have used the storage capacity that has been allocated to them. Ideally, all customers would consume 100 percent of their allocated capacity. However, most customers use only a small percentage of their allocation. Higher levels of storage pool oversubscription are possible in pools with lower customer utilization. We often refer to this metric simply as Customer Utilization.

“Most customers use only a small percentage of their allocation.”

Allocation Headroom

Ultimately, the operations team needs to know whether they can allocate more storage from a pool in response to customer demand and, if so, how much storage can be allocated. The Allocation Headroom metric is designed to enable IT storage operations engineers to quickly and easily determine this.

Allocation Headroom is defined as the additional storage capacity, in gigabytes, that can be allocated from a pool. It is calculated based on several factors, including Allocated Percentage, Used Percentage relative to the thresholds described above, and Customer Utilization of Allocated Capacity. We define policies that alert operations engineers when a pool’s Allocation Headroom has decreased to a predetermined low level.

“Our goal is to proactively analyze and manage SAN storage across our office and enterprise private cloud.”

Analyzing Efficiency: Business Intelligence and Reporting

We have begun building business intelligence and reporting capabilities based on our metrics. Our goal is to proactively analyze and manage SAN storage across our office and enterprise private cloud.

We created a storage resource management (SRM) tool that automatically gathers storage capacity-related data across our SAN environment. Using the SRM tool, we can pull data from individual storage pools or aggregate data by frame or by data center. This allows managers to compare efficiency across different data centers, frames, and pools.

As of May 2012, we have established regular on-demand reporting across 130 pools in 16 data centers, representing over 3 PB of raw storage within our private cloud.

We use the information gathered by the SRM tool to create customized reports for different purposes and users within Intel IT.

For example, executives may receive a high-level overview of average efficiency and other summary data. Engineers responsible for day-to-day operations need reports that are at a more granular level and that alert them to potential problems such as over-allocated storage pools.

Some example reports are shown below. These show actual Intel IT operational data; we have changed some identifying details to protect Intel proprietary information. We are applying thin provisioning to increase efficiency within these data centers.

Executive Storage Efficiency Report

This report, shown in Table 1, includes the Overall Storage Efficiency metric and also provides a capacity overview for Intel IT executives and senior managers at the data center level.

Data Center	Raw Capacity (GB)	Used Capacity (GB)	Overall Efficiency %
DC1	114,174	13,999	12.26%
DC2	40,260	13,328	33.10%

Table 1: Executive Report
(Source: Intel Corporation, 2012)

Data Center Manager Report

These managers need to know how much usable capacity exists, how that capacity is being allocated and used, and how much headroom is available in each storage tier.

This information is critical for planning purchases of more storage capacity based on established IT operations policies and customer service-level agreements (SLAs).

While this report provides a useful overview at data center and storage tier levels, data center capacity managers typically also need information that helps them manage capacity at the level of individual frames.

This frame-level report tells data center managers that they need to take action by working with the operations team to rebalance storage utilization among the existing frames or by acquiring more storage capacity. If they identify that more storage capacity is required, they can examine slot availability to determine whether there are available slots in existing frames or whether new frames need to be purchased.

Pool Status Report

Operations engineers responsible for day-to-day storage management need more granular information at the storage pool level. This enables them to respond to requests for more storage by determining which pools have enough capacity.

Conclusion and Next Steps

Storage capacity utilization metrics enable us to improve the efficiency of our private cloud by providing a framework we can use to analyze and manage SAN storage use across our diverse IT environment. Metrics also provide information required by Intel IT for capacity management and risk management.

We find that the reports and metrics help our storage operations engineers and storage capacity managers work together to maximize storage use efficiency. The reports help them identify the best course of action. For example, when a pool's utilization or allocation reach high levels, they can determine whether they need to purchase new capacity or can simply rebalance storage across existing pools, frames, or data centers.

We have established on-demand reporting across 130 pools in 16 data centers, representing more than 3 PB of raw storage within our private cloud. We continue to increase the reporting coverage across our private cloud, and, in the future, we also anticipate being able to apply this approach across our traditional office and enterprise NAS environment.

We continue to improve our metrics. For example, we have developed an adaptive algorithm that can create customized allocation headroom estimates for each pool based on factors such as the rate at which a customer's capacity utilization is increasing.

As we scale our storage environment and increase utilization, performance health metrics will be essential to understand application and system requirements, identify potential bottlenecks, and assist in successful deployment. Accordingly, we are in the process of incorporating storage performance health into our capacity management and efficiency views. These

“Data center capacity managers typically also need information that helps them manage capacity at the level of individual frames.”

“Storage capacity utilization metrics enable us to improve the efficiency of our private cloud by providing a framework we can use to analyze and manage SAN storage.”

performance metrics include concepts such as response time, queue length, and storage processor utilization.

While we roll-out thin provisioning, we are also exploring the use of other efficiency technologies such as data de-duplication. Our metrics will enable us to measure the efficiency improvements resulting from the implementation of these technologies.

Author Biographies

Sudip Chahal is a principal engineer and compute and storage architect in Intel IT Enterprise Architecture group where he and his team work on IT architectures, strategies, and roadmaps for the enterprise compute and storage services. Sudip attended the Indian Institute of Technology, Kanpur and the University of Wisconsin, Madison earning a B. Tech. EE and MS CS respectively.

Chris Peters has more than 21 years' experience in business and technical fields ranging from information technology, to product marketing, manufacturing, supply chain, nuclear power, and consumer products. Chris is currently Director of Industry Engagement with Intel's Information Technology Team, where he and his team work closely with senior IT peers worldwide to share IT best practices with the goal of delivering better business value from IT innovation and investment. Chris attended the University of Rochester and University of Connecticut earning a BS Physics and MBA respectively.

Ajay Chandramouli has over 13 years of experience in the technology industry with over 10 years of experience at Intel Corporation. Ajay has held a variety of IT, software and hardware engineering positions while at Intel and the Lawrence Livermore National Laboratory. Ajay's current role is to share Intel IT's Cloud and Data Center best practices with his senior IT peers across the industry. Ajay holds both an MBA and MSE from UC Davis.

Steve Owen has held a wide variety of positions in IT Operations since joining Intel Corporation in 2000. Steve's current role is in the Storage Operations team designing and developing automation tools.

Krish Anandarao is currently Storage and Data Restoration Service Manager within IT Operations. He held a wide variety of positions at Intel IT since 2001, including Storage Capacity Management and Database Administration previous to his current role.

Shane Healy is Storage and Data Restoration Service Configuration Manager and Architect within IT Operations. He is based out of the Intel Ireland site.

Nigel Wayman has 22 years' experience in designing and administering IT infrastructure systems. Nigel is currently an IT technology specialist providing customer support for the SAN systems based on service-level agreements. He supports internal employees, and/or external customers/suppliers globally.

Nigel documents and troubleshoots SAN related problems for resolution and specializes in SAN performance issues. Nigel collaborates with other groups and organizations to provide a service for infrastructure changes, foster partnerships with other levels of support, elevate data quality concerns, share best known methods and information, communicate known system problems, and keep abreast of upcoming developments.

DISTRIBUTED CLOUD WITH EDGE SERVER

Contributors

Albert L Chen

Software and Services Group,
Intel Corporation

Haifeng Gong

Intel Architecture Group,
Intel Corporation

Faye Briggs A

Intel Architecture Group,
Intel Corporation

Jackson He

Software and Services Group,
Intel Corporation

“The usage models of the edge server pose unique challenges to both edge server hardware and software infrastructure.”

The distributed hierarchical cloud is an emerging cloud architecture that provides cloud services to users and processes client data at locations close to them. It allows users to achieve a better user experience, reduces network bandwidth, provides lower latency, and achieves a lower total cost of ownership.

The edge server proposed in this article is a critical component of the distributed cloud. Deployed at the “edge”—the network location closest to the client—the edge server becomes “glue” between the distributed cloud and the organization deploying a cloud-based application. The edge server, often in an appliance form factor, integrates cache, security, storage, and the capability to host local applications to deliver cloud applications seamlessly, effectively, and with high performance and low latency.

The usage models of the edge server pose unique challenges to both edge server hardware and software infrastructure. From the hardware perspective, it calls for an appliance-like form factor and user experience design while still retaining scalable compute and storage capabilities. From the software perspective, it induces using provisioning and service infrastructure to deploy edge service and using virtualization technology to provide a separated and secure environment for different edge services. Integrated hardware/software co-design ensures the edge server can meet the requirements from different market segments ranging from SMB (small to medium business), education, and community, by employing consistent software and hardware infrastructure.

Finally, a case study of deploying edge server in an education cloud for Hongkou MOE, Shanghai PRC, is presented.

Introduction

Enterprise, government, and education institutions increasingly deploy cloud computing for their mission-critical applications in an effort to lower TCO and improve the computing efficiency.

Cloud computing refers to applications and services that run on a distributed network using virtualized resources and accessed by common Internet protocols and networking standards. It is distinguished by the notion that resources are virtual and limitless and that details of the physical systems on which software runs are abstracted from the user.

The NIST (The US National Institute of Standards and Technology) definition for the four deployment models for cloud computing is as follows:^[1]

- *Public cloud.* The public cloud infrastructure is available for public use alternatively for a large industry group and is owned by an organization selling cloud services.
- *Private cloud.* The private cloud infrastructure is operated for the exclusive use of an organization. The cloud may be managed by that organization or a third party. Private clouds may be either on- or off-premises.
- *Hybrid cloud.* A hybrid cloud combines multiple clouds (private, community or public) where those clouds retain their unique identities, but are bound together as a unit. A hybrid cloud may offer standardized or proprietary access to data and applications, as well as application portability.
- *Community cloud.* A community cloud is one where the cloud has been organized to serve a common function or purpose.

Three service types have been universally accepted:

- **Infrastructure as a Service: IaaS** provides virtual machines, virtual storage, virtual infrastructure and other hardware assets as resources that clients can provision. The IaaS service provider manages the entire infrastructure, while the client is responsible for all other aspects of the deployment. This can include the operating system, applications, and user interactions with the system.
- **Platform as a Service: PaaS** provides virtual machines, operating systems, applications, services development frameworks, transactions, and control structures. The client can deploy its applications on the cloud infrastructure or use applications that were programmed using languages and tools that are supported by the PaaS service provider. The service provider manages the cloud infrastructure, the operating systems, and the enabling software. The client is responsible for installing and managing the application that it is deploying.
- **Software as a Service: SaaS** is a complete operating environment with applications, management, and the user interface.

In this transformation to cloud computing, no matter which deployment model and service type the organization chooses, they face a set of challenges that must be overcome.

- **Content delivery.** With increased use of computing equipment ranging from PCs to tablets and exponential growth of cloud applications, latency and bandwidth are becoming the two biggest challenges to the organization, where people need to access cloud-based services and applications on a regular basis. A traditional content delivery network (CDN) solves latency and bandwidth issue by deploying caches closest to the place where the request initiates. In most of cases, a CDN grid sits in an ISP's data center. However, for organizations to deploy cloud services, the cache needs to be deployed within the organization.
- **Local application hosting.** Even though the organization plans to deploy a cloud-based application, there are still some services that it prefers to host

“A hybrid cloud may offer standardized or proprietary access to data and applications, as well as application portability.”

“For organizations to deploy cloud services, the cache needs to be deployed within the organization.”

“Concern about data security is very common in the small and medium business in emerging economies.”

locally or from within the organization. There are two reasons for this: first, some applications are more suitable to be deployed from within an organization rather than from a cloud. Those applications typically involve large and frequent data storage and transfer. For example, large data sharing and collaboration might be better deployed within an organization to take advantage of a faster local area network while avoid burdening limited broadband bandwidth. Second, some data generated by the organization is regarded as sensitive and mission critical and it is therefore preferable not to store this data outside the organization. This concern about data security is very common in the small and medium business in emerging economies, where trusted brands of cloud hosting services are not yet established.

- TCO. For an organization to address these challenges, several additional services need to be deployed on the edge. Those edge services typically include cache, security, and the capability for data storage and local service. This actually counteracts the effort of moving all services to cloud to make the organization server-less, more efficient, and cost effective.

Those challenges lead to the deployment of distributed cloud and edge server in favor of traditional centralized data center model.

Distributed Cloud

In traditional cloud architecture, the cloud services are provided by the servers centrally deployed in the data centers, and the users obtain these services through an Internet connection. There are latency and reliability challenges associated with this architecture because of the limited network bandwidth. And there are also scalability challenges because of the inflexibility of cloud infrastructure upgrades. The distributed hierarchical cloud is an approach to address these challenges.

The concept of the distributed hierarchical cloud is illustrated in Figure 1. The clients that are physically distributed in different locations obtain the cloud services through the edge servers closest to them, and the clients' data is processed by the edge servers closest to them too. The edge servers that are physically distributed in the various edge locations and the servers deployed in the data centers are all centrally managed and are logically one coherent cloud infrastructure.

The distributed hierarchical cloud architecture can be applied to many usage models. For example:

- Business: in companies with remote offices or stores, edge servers are deployed in the distributed locations and are networked to the central data center. The distributed hierarchical architecture ensures the client/terminal devices (personal computers, POS machines, video surveillance, and so on) are securely connected and managed. The data is safely and efficiently transferred and stored across the sites.
- Education: in K12 schools, students and teachers use the services and contents from the cloud for the classroom interactions, homework,

“The distributed hierarchical cloud architecture can be applied to many usage models.”

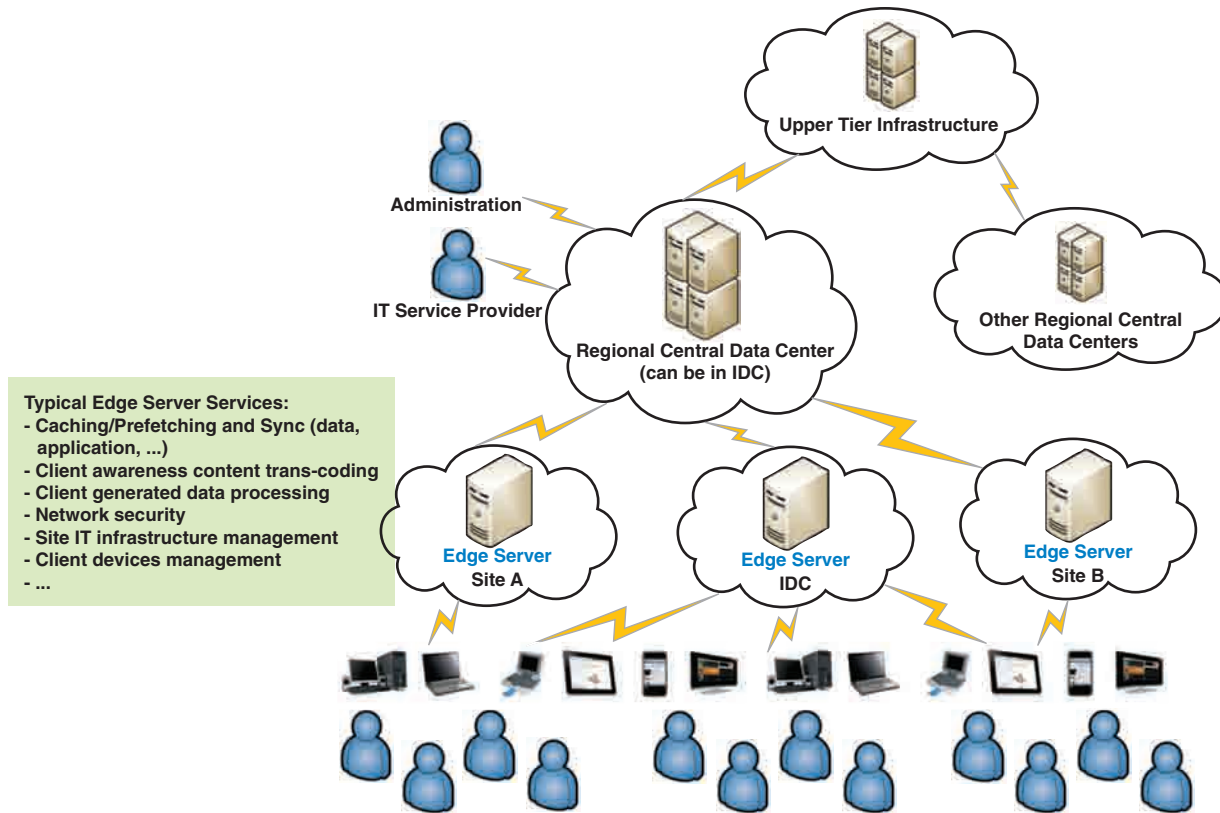


Figure 1: Distributed hierarchical cloud architecture
(Source: Intel Corporation, 2012)

examinations, and so on. The edge servers deployed in each school can cache the data and applications from the cloud and thus reduce the network traffic and improve the latency. The edge server can also host the other edge services like network security and school office automation system.

- **Community:** the edge server can be strategically deployed in the dense population areas or remote rural areas. The edge servers can offload the burden of data center and Internet infrastructure by doing caching and pre-fetching; they can even further improve the network efficiency and improve the user experience by providing the services like client awareness trans-coding and client data analysis.

Edge Server in the Distributed Cloud

To support the various types of applications and different scale of users and workloads, the edge server should easily scale out in terms of processors, memory, storage and network.

Most importantly, because of the nature of the edge server deployment locations (resident communities, rural villages, schools, hospitals, and remote offices), which are not located in data centers, it is desirable that the edge server can provide an appliance-like user experience. In other words, it should be able

“Normal server products or even special-purpose server products may be used as edge servers.”

to work over a wide range of environmental conditions (power, temperature, and air quality), be remotely managed without requiring on-site support, and provide the feature set that is sensitive to both capital and operational expenses.

Normal server products or even special-purpose server products may be used as edge servers. However, many of them are not optimized and may even have inherent limitations for the edge usage. For example, the traditional pedestal server products and rack-mount or blades are normally designed for the data center environment. Their architecture and feature sets are not targeting the usage model of a distributed hierarchical cloud. Some special-purpose devices such as network firewall appliances can also provide some compute and storage capabilities for the edge applications. But the performance, flexibility, compatibility, and scalability are all constrained by the nature of those special-purpose devices.

In order to meet the distributed cloud requirements and overcome the limitations of the traditional server products, a scalable appliance-like edge server is needed.

Edge Server System Design

The basic idea of the scalable appliance-like edge server system is a modular all-in-one system with optimizations for scalability, environmental flexibility, system manageability, and lower total cost of ownership.^[3]

Figure 2 illustrates the main building blocks of the scalable appliance-like edge server system. Figure 3 illustrates a concept system.

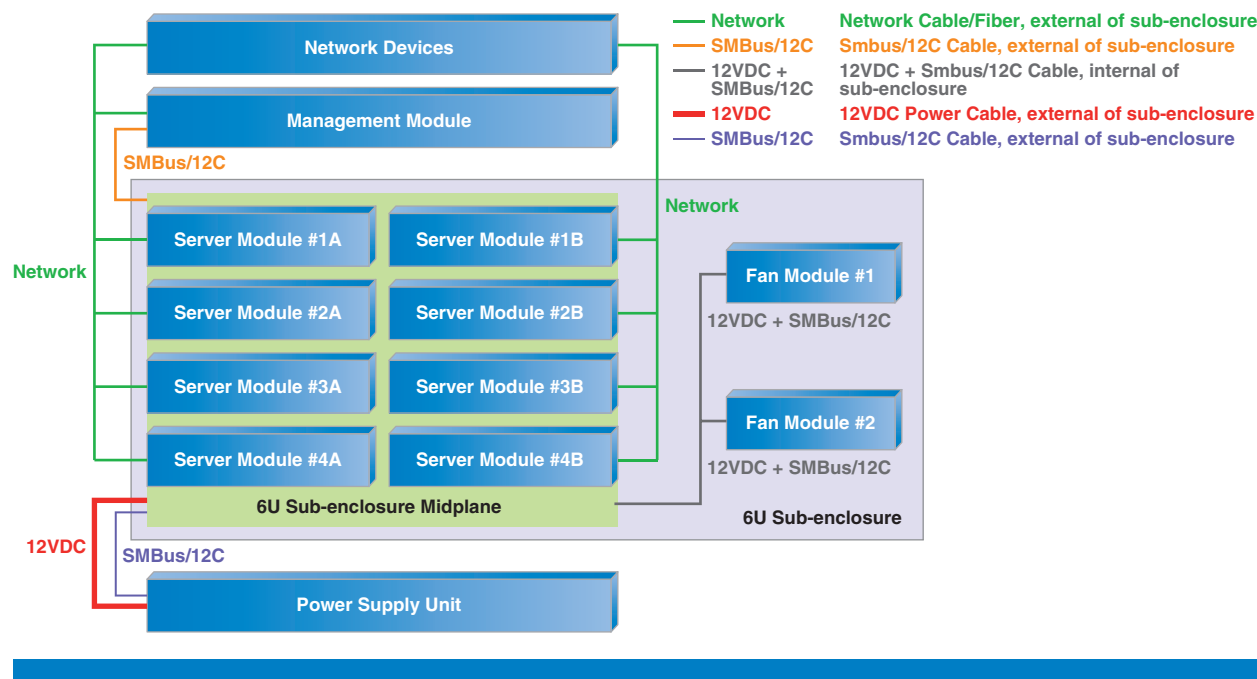


Figure 2: Scalable appliance-like edge server system
(Source: Intel Corporation, 2012)

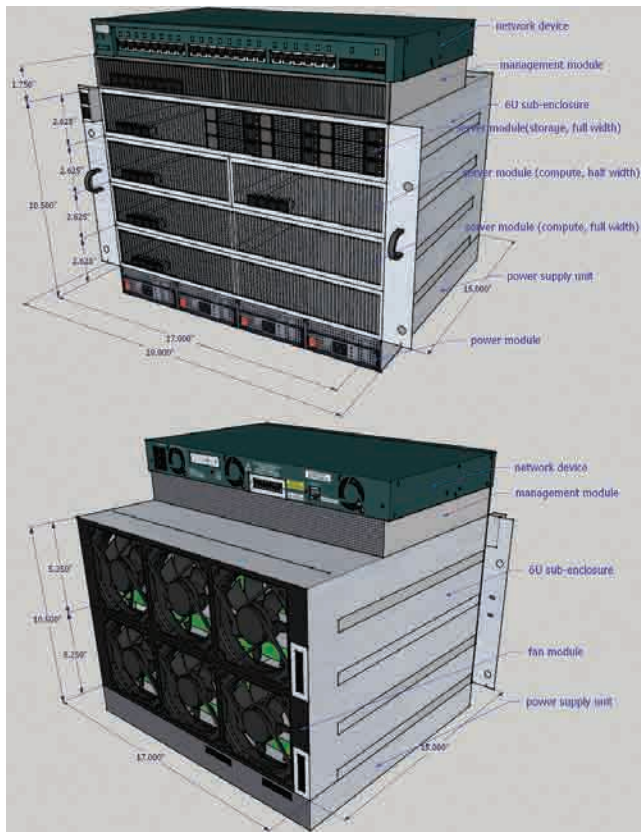


Figure 3: Edge server system concept
(Source: Intel Corporation, 2012)

The innovative design features include:

- Modular system with access to the server modules and the I/O ports from the front side. In contrast to the traditional servers, which have a control panel at the front side and the I/O ports at the rear side, the all-front-access allows this edge server to be installed in a constrained space.
- The structure, dimensions, and layout of the server modules as well as the system components are optimized for cooling efficiency. Keeping good air flow is one of the priorities of this edge server system design.
- The server modules have the flexibility to accommodate different motherboards with different types and numbers of processors and memory.
- The compute type server modules are designed with no local hard drive and are “state-less,” they use the SAN or iSCSI storage provided by the storage type server modules. This ensures good serviceability of the edge server system.
- All server modules are interconnected with network cables/fibers through the I/O ports on the front side to the network switches. The network

“Keeping good air flow is one of the priorities of this edge server system design.”

“All server modules use a shared system power and cooling solution.”

switches are the standard off-the-shelf products; the users have the flexibility for cost savings.

- The power supply and the fans are separate from the server modules. All server modules use a shared system power and cooling solution. The shared power and fans save the BOM cost of the server system without compromising the redundancy. Besides, the shared fans also improve system power efficiency because the larger size fans are more efficient than the smaller ones in normal server products.
- The system has the flexibility to use a variety of power supplies with different input voltages and overall capacity. The system also has the option to use the power supplies with a backup battery input.
- A management module monitors and manages all components and resources in the system, which provides the interfaces for remote management.
- With a compact form factor, the system can be installed in the standard EIA-310 19-inch (482.6mm) rack, or just placed on the desk or ground. The footprint of the concept system in Figure 3 is 17 inches (W) x 15 inches (D). It is about half the size of normal server products (typically 17 inches x 28 inches).
- The system configuration can start from one server module and easily scale out to many compute type and storage type server modules. One system chassis in Figure 3 can be configured with a maximum of four full-width or eight half-width server modules, and the system can further scale out to multiple system chassis (up to six if installed in a standard 42U rack). All server modules can be remotely managed through the management module.

Edge Server Software Infrastructure

The design of software infrastructure for the edge server takes into account the following considerations:

“The appliance form factor leads to the expectation of low maintenance needs even after stretched service time.”

- Reliability and robustness. The edge server is a converged appliance. It hosts multiple services ranging from cache to collaboration tools. Meanwhile, the appliance form factor leads to the expectation of low maintenance needs even after stretched service time. As a result, the edge server software infrastructure needs to be reliable and robust, capable of hosting edge services in a highly secure, recoverable, and segregated environment.
- Scalability and extensibility. The edge server, though in a form of a closed appliance, needs to be customized and extended to meet various needs. For example, the administrator of an organization needs to install new local services to meet evolving organization needs. Sometimes, the edge server even needs to be repurposed for a different use case. As such, a provisioning infrastructure needs to be in place to support customization and extension of the edge server.

- Easy to maintain and manage. The edge server typically is deployed outside of the data center and within the organization that lacks sufficient IT skill and infrastructure. The software infrastructure of the edge server must enable the edge server to be managed and serviced remotely and in an automated fashion.
- Easy for third party development. The edge server is deployed for different usage cases and needs to host various services based on different requirements. It is critical for the edge server to abstract its core capability to the services and applications so that those services and applications can take advantage of them seamlessly. Also, as a critical requirement to scale the ecosystem, a software development kit (SDK) must be provided to enable the developer to quickly develop and integrate their services with the edge server.

As a result, the edge server software infrastructure includes the following components, as illustrated in Figure 4.^[6]

- Edge service as a VM (virtual machine) appliance. Edge service can be packaged in the form of a VM appliance. There are two categories of edge services hosted in the edge server: generic services and domain-specific services. The generic services such as caching/pre-fetching, network security, and management console are common for any usages. The domain-specific services are the ones specifically for certain usages such as classroom applications for an education cloud or hospital applications for a health cloud. In the form of a VM appliance, an edge service retains the maximum degree of security and manageability and is easier to deploy.
- Edge service infrastructure. It is hosted in a remote data center by service providers or an OEM. It works with an agent residing on the edge server to: a) deploy additional edge services purchased by end users, b) upgrade existing pre-built edge services and to recover damaged ones, c) manage overall security of the edge server including authorization, authentication, and encryption, d) keep in track of a profile of the edge server, e) perform critical manageability tasks including backup on regular basis, f) provide an edge service store. It is hosted in a remote data center by service providers or OEM. This service store hosts a range of edge services that can be provisioned from cloud to edge server sitting near the organization. An edge service store is implemented as web service and could be part of a more general application store like Intel's AppUp™.
- Intel Edge Server API. Intel Edge Server API abstracts the core capability of the Intel edge server including manageability, storage, virtualization, and an application/service framework. It enables quick, consistent, and robust service development and deployment.
- Intel Edge Server SDK. Built on top of Intel Edge Server API, Intel Edge Server SDK is a development tool for any developers who want to develop and deploy both generic edge services and domain-specific edge services for Intel edge server.

“It is critical for the edge server to abstract its core capability to the services and applications.”

“An edge service store is implemented as web service.”

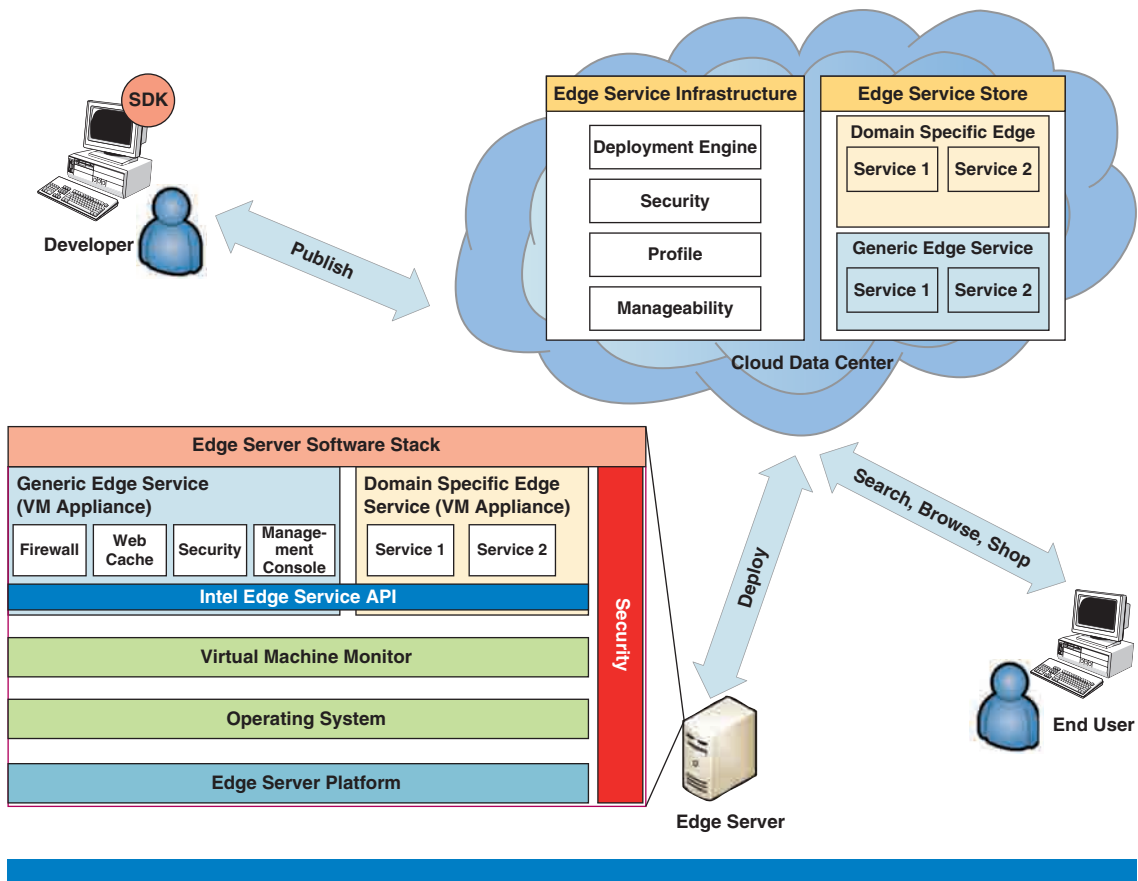


Figure 4: Edge server software infrastructure
(Source: Intel Corporation, 2012)

A Case Study: The Edge Server in an Education Cloud

The following case study describes a case study of an edge server deployment for an education cloud.^[4]

Background

The government of Hongkou district, one of 15 districts of Shanghai, China, plans to roll out education cloud for all its public K12 schools. Intel, as one of its strategic industry partners, is responsible for delivering an architecture blueprint for Hongkou Ministry of Education, who is in charge of the entire education cloud project. The goal of Intel as its trusted advisor is to ensure that the education cloud is deployed smoothly with the right technology and product from Intel and the industry.

Following is a list of facts about the education cloud to be deployed by the government of Hongkou district.

- Hongkou district comprises about 100 K12 schools, each, on average, having around 1000 students and teachers. Current broadband connection for each school is 10 Mb.

- Hongkou district plans to deploy one education-purpose-built computing client for each student and teacher. Those education-purpose-built clients, dubbed e-schoolbag, integrate various applications and services and serve as a portal for students and teachers to access education content. E-schoolbag is preferred in the form of a laptop and tablet.
- All e-schoolbags are connected via Wi-Fi* within the school.

An education application is hosted in the cloud. The application initially includes e-textbook and education resource sharing and will extend to education community, education office online, and education data analytics in the future.

The Challenge

The challenges of architecting an education cloud for the Hongkou district are fourfold:

- **Bandwidth.** According to usage model within the K12 schools, each student is required to download about 50 MB of educational resources within five minutes between classes. Given the fact that there are around 1000 students in school and each download consumes 0.5 Mbps bandwidth, assuming 10 percent concurrence (which is fairly conservative as each break between classes is almost on the same time), bandwidth required for school is $(1000 \times 10 \text{ percent} \times 0.5 \text{ Mbps}) = 50 \text{ Mbps}$. Correspondingly, the data center's bandwidth required is $(100(\text{total number of schools}) \times 1000 \times 10 \text{ percent} \times 0.5 \text{ Mbps}) = 5 \text{ Gbps}$. This bandwidth requirement is unacceptable.
- **Concurrency.** Even though 10 percent concurrency assumed and bandwidth is not the problem, given each web server is expected to handle 200 http requests at the same time, the number of web servers alone in the data center is $((10,000 \times 10 \text{ percent})/200) = 50$. Again, this is too expensive and inefficient to deploy.
- **Security.** Because most of customers for this education cloud are students, both parents and administration of schools are concerned with the possibility of students accessing inappropriate information when surfing the Internet with e-schoolbag within the school.
- **Local application hosting.** Even though the education cloud brings down operation cost and improves overall application delivery efficiency, there are some education applications better deployed locally rather than from the cloud. In case of Hongkou, a suite of teacher collaboration tools needs to be deployed to facilitate teacher's context switch between classes. This involves frequent large data storage and transfer and is therefore not suitable to be deployed from the cloud for performance considerations.

Solution

To address these challenges, the blueprint of the Hongkou Education cloud proposed by Intel calls for the deployment of the edge server as illustrated in Figure 5.

“E-schoolbag is preferred in the form of a laptop and tablet.”

“There are some education applications better deployed locally rather than from the cloud.”

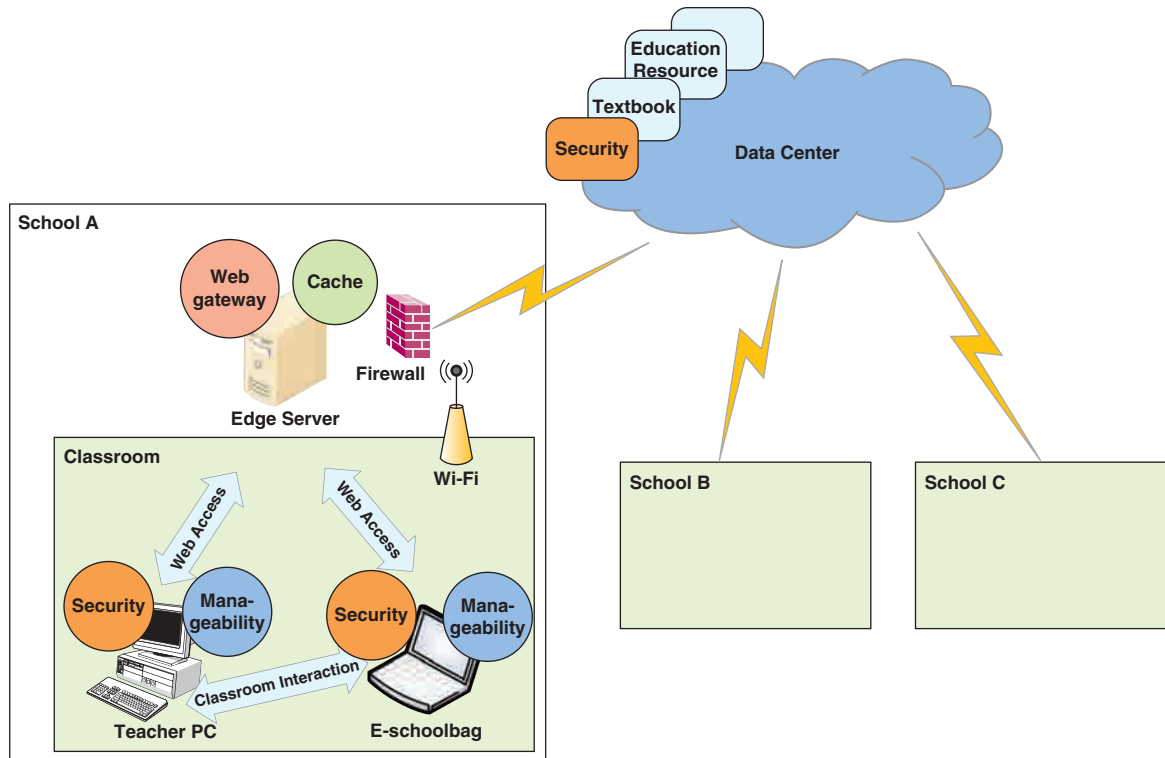


Figure 5: Edge server in Hongkou education cloud
(Source: Intel Corporation, 2012)

This edge server employs the following components:

- **Cache.** The edge server employs an intelligent cache to cache the frequently accessed content. The education content requested by students tends to be the same. In the case of Hongkou, students in the same class actually request the same content with the size of about 50 MB. As a result, the cache hit rate is very high and predictable. Cache deployment in this case is very effective. Furthermore, there are some cache policies that can be deployed to optimize the cache usage. For example, because all education contents are generated by the teacher one day before, a simple policy can be written for the cache to pre-fetch all education contents in a certain timeframe like early morning, avoiding network rush hour in the day. Overall, deployment of the cache greatly mitigates both the bandwidth and concurrency issues.
- **Web gateway.** Because the edge server is forwarded all web traffic from the school, it is a great opportunity to deploy a web gateway to secure all Internet access from the students. The Web gateway integrates various security technology like anti-malware, anti-virus, and SSL scanning, scanning all Internet content and protocols including web content, streaming media, and user-generated content to detect potential threats based on prior knowledge. In the case of Hongkou, McAfee's Web Gateway is proposed along with the edge server.

- Local application hosting. The edge server in the case of Hongkou is also proposed to host local applications including classroom collaboration tools. The primary function of the edge server is the cache, which is I/O bound, not CPU bound. As such, spare CPU bandwidth in a server, typically an Intel® Xeon®-based platform, can be used to host local applications. Co-location of local applications reduces the server footprint and greatly lowers the manageability overhead for the school, which lacks relevant IT skill and infrastructure.

All those functions are consolidated into one edge server instance with each in a separated virtual machine environment to ensure the security and reliability, as illustrated in Figure 6.

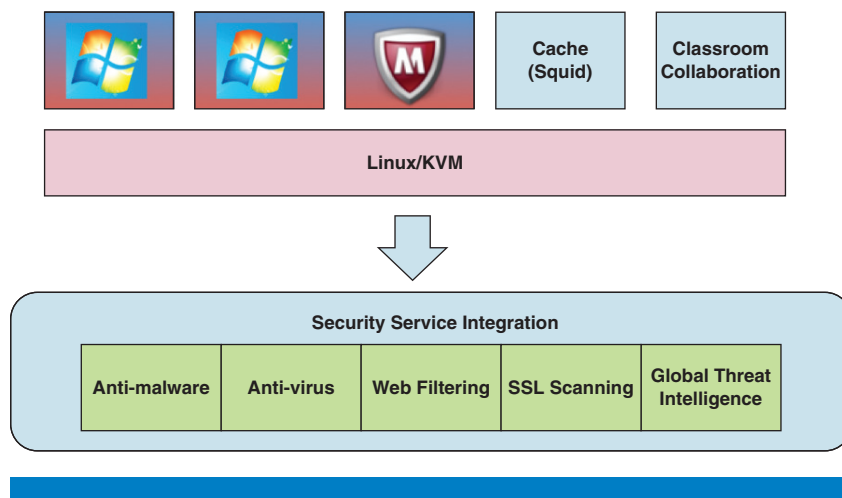


Figure 6: Edge server function consolidation
(Source: Intel Corporation, 2012)

References

- [1] Mell, Peter and Grance, Timothy 2011. “The NIST Definition of Cloud Computing.” National Institute of Standards and Technology, Special Publication 800–145.
- [2] Intel Corporation 2010. Intel’s Vision of the Ongoing Shift to Cloud Computing.
- [3] Gong, Haifeng and Pronozuk, Tony 2012. A Scalable and Appliance-like Edge Server Architecture. Intel APAC Strategic Mini Conference 2012.
- [4] Chen, Albert, Haifeng Gong, Jackson He, et al. Education Cloud Blueprint for Shanghai Hongkou Education Bureau.
- [5] Hofmann, Markus and Leland R. Beaumont, Content Networking Architecture, Protocol and Practice.
- [6] Chen Albert, Gong Haifeng, He Jackson, et al. Scalable edge server Software Infrastructure.

Author Biographies

Albert Chen is a software architect of Intel APAC R&D Ltd., has been with Intel for 13 years, and has led many system software innovations. Currently Albert works on architecting an education cloud for the Chinese Ministry of Education. Email: albert.l.chen@intel.com.

Haifeng (Bill) Gong is server board design manager of Intel APAC R&D Ltd., has been with Intel for 8 years, and has participated in the development of four generations of Intel server products. Currently Bill is also working on the server solutions for the data center and cloud. Email: haifeng.gong@intel.com.

Fayé Briggs, Ph.D. is an Intel Fellow and director of Scalable Server Architecture for the Intel Architecture Group at Intel Corporation, with 15 years at Intel. Faye is responsible for ensuring that Intel's multi-core and many-core-based server architectures achieve world-class performance. Email: faye.a.briggs@intel.com.

Jackson He, Ph.D. is General Manager of Intel APAC R&D Ltd., has been with Intel for 17 years, and has led many technical evangelizations. In the past a few years, Dr. He has been driving cloud technology innovations in both the US and China. Email: Jackson.he@intel.com.

A JOURNEY TOWARD A COMPUTE CONTINUUM: CLIENT-AWARE CLOUD SERVICES FOR SMART CLIENTS

Contributors

Jin, Jennifer

Intel Architect System Integration

Ding, Jonathan (JunYong)

Intel Software and Service Group

Xu, Samuel (Shu)

Intel Software and Service Group

“Client-awareness and HTML5 are critical architecture building blocks.”

One of the primary objectives of mobile Internet services is to deliver a compelling user experience. The key is to deliver the services that best fit the devices that serve the end users and make the users have a consistent continuum of experiences across these devices. As part of the much grander vision of a compute continuum, client-awareness and HTML5 are critical architecture building blocks. This article focuses on the architectural essence of the compute continuum and how client-aware cloud services for smart devices fit into the grand vision. Since HTML5 is a key technology for cross-platform content delivery from PC to tablet, TV, and smart phone, this article also focuses on key technology challenges of cloud-awareness using HTML5 technology.

Introduction

The *compute continuum* is Intel's vision of cloud computing to enable the seamless computing to cross different device form factors, for both enterprises and consumers. In this article, we introduce the emergence of client-aware cloud to fulfill the vision of the compute continuum, the technical challenges of the client-aware cloud, and Intel's initiative to explore and solve those challenges. Combining with the HTML5 initiative from W3C, we also introduce how client-aware technology will be shaped in near future to present the seamless Compute Continuum.

The Road to the Compute Continuum—Architecture Tenants

We constantly understand how people are currently using connected devices and looking for ways to make connected devices work better, and work better together.

The Cross-Device Continuum Experience

The compute continuum is an Intel-wide initiative aimed at delivering a range of seamless, personalized, and protected cross-device experiences for end users using their preferred devices. Intel is laying the groundwork in 2012 for bringing exciting and new ways for users to enjoy connected device experiences.

There are two different experiences on the compute continuum: enterprise IT and consumers.

From the enterprise IT perspective, by taking advantage of a combination of technology trends and emerging computer models, such as ubiquitous

Internet connectivity, virtualization, and cloud computing, there is an opportunity to proactively address changing user requirements and redefine the way IT provides service. This represents the next major change in the way that employees will use technologies. Key aspects include the following:

- Users have access to corporate information and IT services securely from anywhere on any device at any time, no matter whether it is personal or corporate-owned
- Multiple personal and corporate devices work together seamlessly
- Corporate information and service are delivered across these devices while the information continues to be protected

For consumers, the compute continuum as one of the smart technologies and services is becoming the industry trend. Consumers are looking for these to blend into their everyday work, lives, and activities, while maintaining privacy and security. Users can experience the compute continuum during their time of entertainment, work, or learning, by enabling simple, secure interactions between varieties of devices across a broad spectrum of such usages. These interactions enable the users' devices to collaborate, to overcome the limitations of a single computing platform, and to enjoy a familiar and consistent experience.

There is no singular usage that compute continuum delivers, but it should include but not be limited to:

- *Device Pairing and Composition.* Every device can seamlessly connect to other devices and share resources, securely.
- *Context Awareness.* Every device is aware of the user and is able to do actions on his/her behalf.
- *My Media, Anywhere.* Every device gives easy means to purchase, access, and enjoy premium content.
- *Identity and Security.* Every device understands and protects the user's identity;
- *Natural UIs.* Every device provides intuitive modes of interaction.

Intelligent Content Delivery

Currently, most users have more than one end-point device, such as a PC, tablet, smart phone, or smart TV. How to share, sync, or pass control among these devices is becoming more and more important. The essentials of intelligent content delivery include:

- Allowing access to content from another device for sharing/viewing
- Controlling an application on another device
- Consuming an application from another device

Furthermore, a smart enough mechanism may process the data before it is delivered. For instance, based on its knowledge of the client, the cloud could

“For consumers, the compute continuum as one of the smart technologies and services is becoming the industry trend.”

“A smart enough mechanism may process the data before it is delivered.”

pick up the suitable quality of the multimedia, such as a video streaming, and only transfer the high definition clips to the device that has a sound network connection and also the capability to decode it.

Client-Aware Device Management

With IT device management evolving from a device-centric model to a user-centric one, we experienced unmanaged clients and now we have all kinds of management software to manage the clients, servers, and mobility end-points. It makes the compute environment more secure and more efficient; however, these solutions are “one size fits all” and don’t support users bringing their own devices. Therefore we believe a centrally managed virtualization era will come, as shown in Figure 1. It drives to centralized administration and will support both client-hosted and server-hosted virtualization, and will deliver a secure, effective, and seamless environment for computing, collaboration and communication.

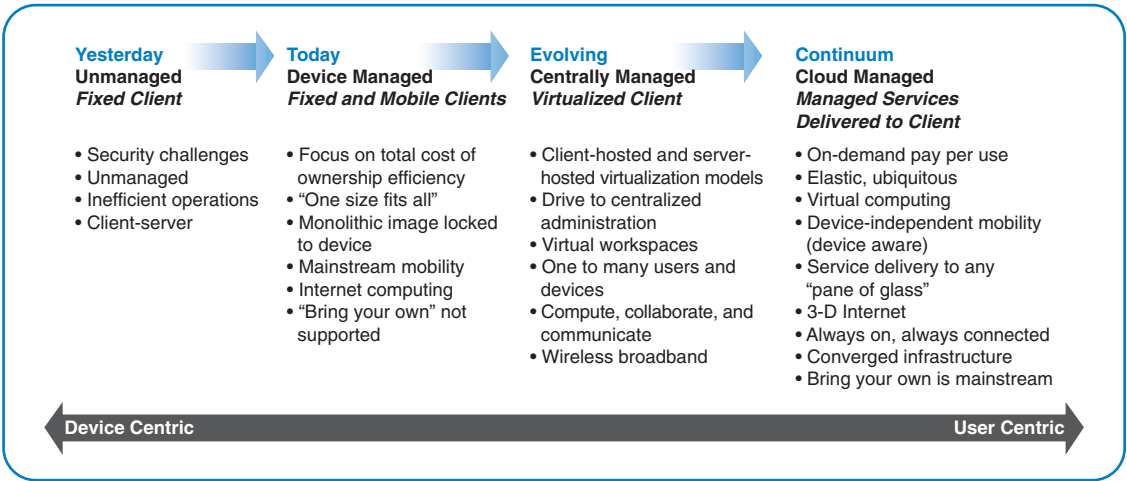


Figure 1: Intel IT path to the compute continuum⁽¹⁾
(Source: Intel Corporation, Future_Enterprise_Computing-Preparing_Compute_Continuum)

Eventually, we should have an elastic and on-demand IT environment that supports all kinds of devices and various computing models, and it will be client-aware. That will be the Compute Continuum!

Why a Client-Aware Cloud Is Important

A client-aware cloud is an intelligent cloud infrastructure that can deliver cloud-based applications taking into account the capabilities of the device at hand.

Today’s users are increasingly likely to utilize multiple devices, including smart phones, tablets, and PCs, to access information. They embrace new

applications and devices in their daily life and expect those same capabilities to be available at work. Yet when it comes to the ability to access, display, manipulate, or secure data, some devices are clearly more capable than others.

The reality is that different devices may have different capabilities, such as different screen size, graphics capacity, and security level. Unfortunately, most Internet services are dumbed down. They probably can recognize the screen size but may be unable to determine and then take advantage of other local capabilities.

End users and cloud providers each have unique needs. Rather than relying solely on the data center to drive cloud capabilities, Intel believes that there's a better and more balanced approach. By taking advantage of the client-aware cloud, service providers have greater flexibility to optimize application delivery. End users can also benefit from improved experience.

“Unfortunately, most Internet services are dumbed down.”

User Benefits

The user benefits of a client-aware cloud are:

- *Responsiveness*: for cloud-based applications, response time depends on multiple factors including network, performance, application performance, and cloud infrastructure performance. Enabling client-aware or even client-side execution for cloud-based applications can help improve end-user experience, in particular, for applications that are compute-intensive or bandwidth-constrained.
- *Productivity*: client-aware cloud services may allow users to access data either online or offline. Furthermore, once reconnected, applications can synchronize the change since last connection.

IT Benefits

The IT benefits of a client-aware cloud are:

- *Application delivery*: the ability to execute a given application or portions of it on the end point provides IT with an additional option to improve application delivery.
- *Flexible architecture*: with profiles exposed to cloud-aware services, deployment becomes easier because of the reduced dependence on the architecture. For example, organizations that have complex, sensitive, or network-constrained applications may find it hard to take advantage of the cloud because of the need for such things as a high bandwidth network connection, a secure communication tunnel, or dedicated client resources. With a client-aware cloud, there may not be so much dependency on the architecture, as the intelligent cloud can detect a client's capability and then determine how the application should run accordingly, such as providing a lower resolution picture over a poor network connection or providing a sensitive application with additional authentication and authorization in an untrusted network.

“With a client-aware cloud, there may not be so much dependency on the architecture.”

“To make the cloud-aware clients’ performance, context, and capabilities exposed to the cloud may bring some challenges.”

Technical Challenges

Intel defines client awareness^[2] in terms of:

- *Compute performance*, which comprises the local performance capabilities in the device, mainly the processor and graphics
- *The device context*, which refers to the ability of the cloud to detect the end-user’s compute environment (that is, network connectivity and bandwidth, power source and battery level, and so on)
- *Security capabilities*, which is being aware of client security features and determining what applications can be delivered

To make the cloud-aware clients’ performance, context, and capabilities exposed to the cloud may bring some challenges:

- What’s the programing environment inside the client and how can the cloud discovery its information?
- How does the cloud know the client’s information?
- How to make sure the client security (protecting virtualized applications, managing systems with full disk encryption) and how to make sure processor-intensive encryption and decryption does not impact responsiveness too much.

As the industry leading provider of silicon building blocks that go into virtually every element of cloud, including client, server, storage and networking, Intel is investing in developing application programming interfaces (APIs) that allow developers to enable cloud-based applications to detect the capabilities of the local client device. Since the Web is the major container of clients of the cloud, Intel has unveiled the Intel Web APIs Connector to enable Web applications to detect endpoint information via JavaScript inside the browser, which contains the following APIs^[3]:

- Network bandwidth API
- Processor performance API
- Battery life API

The Intel® Core™-based vPro™ PC has hardware-based manageability and security, which can comply with the security requirements to run critical applications locally. Table 1 shows a high level overview of the capabilities^[2].

Based on these Intel Web APIs and the Intel Core-based vPro PC, there are two great pilot projects happening in the industry^[4]:

- Gproxy’s Client Device Score (CLIDES) can rank a client device’s capabilities including CPU, CPU load, connection type, bandwidth, and screen resolution. The score can assist the client web application to optimize the user experience
- NetSuite Ecommerce Cloud Platform offers enhanced web UI optimization, shopping cart optimization, preserving user-state and improving Web analytics via Intel Web API.

Challenges	Intel Solution
Theft of remote devices and data exposes organization to risk	Laptops with Intel® Anti-Theft Technology can disable themselves if lost or stolen.
Protecting virtualized applications running on the end point	Intel® Trusted Execution Technology (Intel® TXT) supports hardware-assisted attestation and trust at the end point.
Processor-intensive encryption and decryption impacts responsiveness	Intel® AES New Instructions (Intel® AES-NI) can encrypt and decrypt data up to 3.5 times faster.
Managing systems with Full Disk Encryption can be difficult	Intel® Core™ vPro™ processor enables the ability to unlock and manage an encrypted PC without compromising security or remote management

Table 1: Technical challenges and Intel Solutions

(Source: Intel Corporation, 2012)

These pilot projects are excellent examples of how to address the technical challenges mentioned previously. These challenges can definitely be solved by innovations based on existing techniques:

- *Client programing environment:* Web environment and JavaScript is a proven way to establish the client programing environment
- *How the cloud knows the client's information:* Http (Ajax) is a convenient way to bridge cloud and client
- *Client security:* The CPU API can determine whether the client processor is an Intel® Core™ vPro™ processor, which enables more processing to be done locally, and speeds up encryption and decryption via the Intel AES instruction set.

Another non-technical challenge is how the main Web browsers should implement similar APIs to benefit enterprise IT and the consumer broadly. Since the World Wide Web Consortium (W3C) is the main international standards organization for the Web, we'd like to introduce W3C's standardization roadmap to the client-aware cloud, combining with HTML5 evolution.

The Role of HTML5 for the Client-Aware Cloud

Web technology has gone through numerous evolutions since the early beginnings of the 1980s. Its Episode 5 starts at 2007 when the HTML5 specification was adopted as the starting point of the new HTML working group of the W3C.

“HTML5 is not a single technology, but a collection of improvements to HTML.”

The emergence of HTML5 opens tremendous opportunities to the Web, which is no longer limited to only make up the content delivered from server, but can also access and fully utilize the client’s capabilities inside web runtime.

HTML5 is not a single technology, but a collection of improvements to HTML and keeps expanding and migrating. Currently HTML5 already covers canvas, video/audio playback, offline capacity, drag-and-drop, cross-document messaging, history management, web storage, geolocation, Web SQL, index DB, file and file system, web socket, and so on. W3C is actively driving and standardizing more specifications to expand broader device capacities. For example, W3C WebRTC (Web-based Real Time Communication) API^[5] is a natural expansion from HTML5 video/audio playback to a bidirectional video/audio chat. Another greater example comes from W3C Device APIs Working Group’s DAP API serial^[6], which covers battery status, contacts, media capture, network information, sensor, vibration, web intents, calendar, and permissions APIs.

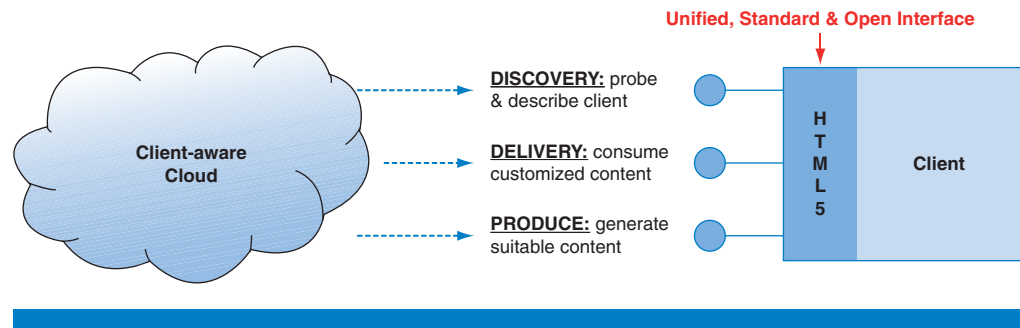


Figure 2: Client-Aware cloud and HTML5 client
(Source: Intel Corporation, 2012)

Therefore, as illustrated by Figure 2, HTML5 would play a significant role in achieving a compute continuum experience by exposing the client context information to client-aware cloud services, consuming the customized contents delivered from remote peers, and generating suitable contents as needed to end users or cloud. In other words, HTML5 is the essential technology to enable smart clients to address the challenges raised to client side due to client-aware services.

According to W3C’s announcement^[7] that the major standards of HTML5 will be ready in July 2014, W3C believes that the Open Web Platform including HTML5 and various APIs will be more mature and have greater interoperability in the near future. Industries are turning to the Web as the platform of choice for integrating diverse devices, services, and business models. There will be an explosion of HTML5 feature adoption in upcoming years. HTML5 definitely is on the road to playing a more and more important role for client-aware cloud for the following reasons:

- HTML5 provides standardized client status description
- HTML5 provides standardized client’s web runtime capacity description

“There will be an explosion of HTML5 feature adoption in upcoming years.”

- HTML5 provides standardized network information description
- HTML5 provides standardized service and application description

HTML5 for Device, Service, and Application Discovery

W3C Web and TV Interest Group^[8] is the pioneer W3C group that was formed in 2010 with one of its missions being to address and standardize the discovery gap and API for audio-visual content.

W3C Web Intents Task Force^[9] is a joint task force of the W3C Device APIs Working Group and the W3C Web Applications Working Group^[10] that focuses on service discovery and lightweight RPC mechanism for web applications. First web intents edit draft was published in 2012 and has enormous industry impact to drive the browser vendors' implementations.

HTML5 for Customized Content Delivery

HTML5 makes web-based bidirectional content delivery more realistic.

HTML5 Client as a Content Consumer

A Web-based customized content delivery capacity and usage model is greatly enlarged by HTML5 in various ways:

- The W3C Network Information API^[11] enables the adaptive content delivery model according to different network environment.
- The W3C Geolocation API^[12] can identify the location of client, which makes location-based content delivery a reality.
- The W3C Battery Status API^[13] exposes the battery status of a client. Based on the client's power information, content delivery and consumption can be more power efficient.
- HTML5 Web Sockets^[14] and Web-based Peer-to-Peer connection API take bidirectional communication between client and cloud to the next level.
- HTML 5 video and audio^[15] gracefully enables multimedia content consumption. With the help of canvas, more usage models can be implemented in a much richer way than ever before.
- HTML5's offline capacity, the Indexed Database API^[16] significantly enriches the local cache capability of web applications, which is key to a flexible and efficient content delivery to reduce the dependency on a network connection.
- The Web Audio API^[17] makes it possible to offload more server-side audio processing to the client side.
- The WebGL API^[18] makes it possible to interpret 3D semantic content and render it inside an HTML5 client.

The HTML5 Client as a Content Producer

Benefitting from HTML5, the Web-based client has also evolved to be a powerful content producer.

“HTML5 makes web-based bidirectional content delivery more realistic.”

“The W3C Battery Status API^[13] exposes the battery status of a client.”

- The WebRTC API^[5] utilizes the client's microphone and camera to capture audio and video streams, sends out captured media stream, and presents the received stream at the same time.
- W3C GetUserMedia^[19] and HTML media capture API^[20] standardizes the HTML5 client's audio and video capture capacity with pre- and post-media-content-processing capacity.

Conclusion

In this paper, we first introduced Intel's vision of the compute continuum for the enterprise and consumer. We then discussed why the client-aware cloud is important for the compute continuum and the technical challenges ahead. We illustrated Intel's exploration and initiatives to help address these roadblocks. Finally, we briefly described W3C's HTML5 and WebAPI efforts, which as the future client technology, would deliver the compute continuum by naturally collaborating with client-aware cloud.

References

- [1] Intel White Paper. "The Future of Enterprise Computing: Preparing for the Compute Continuum". May, 2011
- [2] Intel White Paper. "Benefits of a Client-aware Cloud". Intel Cloud 2015 Vision
- [3] Andy Idsinga. "Intel CPU Web API Documentation and Examples." Intel Software Network 2010.
- [4] Intel/NetSuite/Gproxy White Paper. "Intel Cloud Builders Guide: Cloud Design and Deployment on Intel Platforms - Balanced Compute Model with NetSuite and Gproxy Design". 2011
- [5] WebRTC Specification 1.0: Real-time Communication between Browsers. W3C Editor's Draft 28 April 2012
- [6] W3C Device APIs Working Group Roadmap <http://www.w3.org/2009/dap>
- [7] W3C Confirms May 2011 for HTML5 Last Call, Targets 2014 for HTML5 Standard <http://www.w3.org/2011/02/htmlwg-pr.html.en>
- [8] W3C Web and TV Interest Group <http://www.w3.org/2011/webtv>
- [9] W3C Web Intents Task Force <http://www.w3.org/wiki/WebIntents>
- [10] W3C Web Applications Working Group <http://www.w3.org/2008/webapps>
- [11] The Network Information API, W3C Working Draft 07 June 2011 <http://www.w3.org/TR/netinfo-api>

- [12] Geolocation API Specification, W3C Candidate Recommendation 07 September 2010 <http://www.w3.org/TR/geolocation-API>
- [13] Battery Status API, W3C Working Draft 29 November 2011 <http://www.w3.org/TR/battery-status>
- [14] The WebSocket API, W3C Candidate Recommendation 08 December 2011 <http://www.w3.org/TR/websockets>
- [15] W3C HTML5 Working Draft 29 March 2012 <http://www.w3.org/TR/html5>
- [16] Indexed Database API, W3C Working Draft 06 December 2011 <http://www.w3.org/TR/IndexedDB>
- [17] Web Audio API, W3C Working Draft 15 March 2012 <http://www.w3.org/TR/webaudio>
- [18] Specification, Editor's Draft 01 May 2012 [http://www.khronos.org/registry/webgl/specs/latest WebGL](http://www.khronos.org/registry/webgl/specs/latest/WebGL)
- [19] W3C getUserMedia Editor's Draft 28 April 2012 <http://dev.w3.org/2011/webrtc/editor/getusermedia.html>
- [20] HTML Media Capture, W3C Working Draft 14 April 2011 <http://www.w3.org/TR/html-media-capture>

Author Biographies

Jennifer Jin is the project owner for the Intel DCSG/CIG/IASI Compute Continuum PRC. She set up the Compute Continuum Engineering Experience Lab in Folsom in 2011. She supports the product group on ecosystem solution assessments to give the architects input on Intel's Compute Continuum project design.

Jonathan Ding (Junyong, jonathan.ding@intel.com) is a software engineer in the Software and Services Group at Intel. His expertise covers browser, Web runtime, HTML5, and related service frameworks.

Samuel Xu (Shu) is a staff software engineer at the System Optimization Technology Center in the Intel Software and Service Group. Samuel focuses on HTML5 optimization and WebAPI development. He holds an MS degree in Computer Science from the Chinese Academy of Sciences. His email is samuel.xu@intel.com.

GRAPHICS ACCELERATION FOR HTML5 AND JAVASCRIPT ENGINE JIT OPTIMIZATION FOR MOBILE DEVICES

Contributors

Jonathan Ding

Software and Services Group,
Intel Corporation

Yuqiang Xian

Software and Services Group,
Intel Corporation

Yongnian Le

Software and Services Group,
Intel Corporation

Kangyuan Shu

Software and Services Group,
Intel Corporation

Haili Zhang

Software and Services Group,
Intel Corporation

Jason Zhu

Software and Services Group,
Intel Corporation

HTML5^[1] is considered to be the future of the Web and is expected to deliver a user experience that is comparable to or better than that found only in native applications in the past. This does raise significantly higher demand of performance to sustain the much heavier web applications that manipulate more and richer contents than ever before. Consequently, the optimization to the web runtime is extremely important to achieve the success of the platforms, and in particular, the mobile devices, because their hardware capabilities are less powerful than PCs.

This article first introduces the challenges caused by HTML5 in terms of performance. As the rendering engine and JavaScript engine are two key fundamental building blocks of web runtime, it then discusses our solution of graphics acceleration and just-in-time (JIT) optimization applied to the Intel® Atom™^[2] platform to dramatically improve the performance of these two components respectively. The article evaluates the impact of our solutions on typical HTML5 and JavaScript benchmarks, and the vision of future work is included as well.

Introduction

HTML5, the open and compelling web technology, is considered to be the future of the Web, and consequently, more and more client applications are created via HTML5, connected with the cloud, and deployed throughout the Web. Gartner^[3] estimated that half of the mobile applications will be web applications by 2015 and HTML5 is one of the key driving forces. A recent and typical example is the famous Angry Birds game, which also has an HTML5 version^[4] that delivers a user experience consistent with previous native versions of the software.

However, the popularization of HTML5 raises significant demand of performance on client devices to sustain the smooth user experience.

Web applications today have much greater computation requirements than traditional Web surfing and navigating pages. They usually have heavier and more complex logic. The number of lines of JavaScript code in the Gmail client reached 443,000 in 2010, which was almost 50 times the size of 6 years ago.^[5] Furthermore, thanks to the proliferation of mobile devices with cameras and sensors attached, HTML5 web applications often produce and consume dramatically richer content, such as images and high definition video clips. Finally, a compelling user experience requires immediate responses to user

actions like touch gestures. It means that the web runtime should be capable to process such high priority interactive tasks in a timely manner.

In addition, there are more challenges of performance for web applications, compared with native or traditional managed binaries, which are composed with languages like C/C++, Java, or .NET. HTML5 promises cross-platform implementation; however, this is approached through an abstraction layer of JavaScript, HTML, and CSS. Such abstraction does bring more overhead than native solution. For example, even though JavaScript performance has improved by a factor of 100 in the past decade^[6], it is slower than C/C++ in most cases by at least an order of magnitude. Meanwhile, HTML5 is still an emerging technology. Therefore, unlike the mature Java or .NET runtime, which have already been heavily optimized over decades, the software stack of HTML5 is still young and has much room for improvement, in both functionality and performance.

With complex web runtime applications, the rendering engine and JavaScript engine usually play the key roles and dominate the performance of most of the HTML5 applications. The rendering engine is responsible for drawing the contents and presenting them to the underlying window framework so that operation system can show them on the display. The JavaScript engine executes the application's logic, which is composed with JavaScript. Our analysis shows that on Android^[7] phones, the sum of both components usually occupies more than 60 percent of the CPU execution cycles for typical HTML5 games or general rich page navigations.

Therefore, a lot of existing optimization effort targets these two components. For example, modern desktop browsers such as Chrome^[8] use graphics acceleration to improve the efficiency of the rendering engine, in particular, the HTML5 canvas 2D. JIT technology and other advanced compiler technologies are applied to JavaScript engine. Crankshaft^[9] of the V8^[10] JavaScript engine improves the score of v8bench^[11] by more than 50 percent according to Google measurements.

Nevertheless, on mobile devices, such as Android- or Linux-based phones and tablets, the progress is much slower. The Android browser still uses pure CPU for the rendering of HTML5 canvas 2D, even on the latest Android 4.0. WebKit^[12] is one of the most popular web engines. Advanced JIT technology, so-called DFG JIT, is enabled inside JSC (JavaScriptCore), the default JavaScript engine that comes with WebKit, whereas it is available on X64 architecture but doesn't work on IA32.

Our work in the article changes this situation on mobile devices. The rest of the article is organized as follows: "Graphics Acceleration for HTML5 Canvas 2D in Android 4.0" elaborates our solution of hardware-accelerated HTML5 Canvas 2D in Android 4.0 Intel Atom based devices. "DFG JIT on IA32" introduces the details of enabling of DFG JIT on 32-bit Intel architecture platforms. "Conclusion and Future Work" shares our view of future work on both areas.

"HTML5 is still an emerging technology."

"On mobile devices, such as Android- or Linux-based phones and tablets, the progress is much slower."

“HTML5 Canvas 2D^[13] is one of the most important and mature features in HTML5.”

Graphics Acceleration for HTML5 Canvas 2D in Android 4.0

HTML5 Canvas 2D^[13] is one of the most important and mature features in HTML5 and is widely supported by modern mobile devices. HTML5 Canvas 2D allows application developers to invoke a broad range of 2D operations through APIs^[14] defined in JavaScript on a canvas defined by the <canvas> tag. Such operations include manipulating images, drawing vector graphics like lines, rendering text, and so on. All of them are essential to a considerable number of typical usage scenarios such as games, photo albums, and visual editing tools.

HTML5 Canvas 2D and Benchmarks

For most games, taking the HTML5 version of the famous Angry Birds as an example, image-related HTML5 Canvas 2D APIs are particularly important, because image operations are typically heavily used in such scenarios and consequently dominate the user experience.

As a result, a few public benchmarks have been created to simulate these use cases and quantitatively measure their performance. FishIETank^[15] from Microsoft is one of the most popular of benchmarks. It is widely referenced as a key performance indicator for smart phones and tablets by many third-party publications and sites. The original FishIETank is sensitive to various canvas sizes and also has some run-to-run variance because a random number is used in implementation. We made some slight corresponding modifications to ensure consistent results between multiple runs, with a fixed canvas size at 700x480 for better “apple to apple” comparisons. Hereafter, without specific declaration, the FishIETank discussed in rest of this article refers to the version we modified rather than the original.

“GUIMark3^[16] is another widely used benchmark in the industry.”

GUIMark3^[16] is another widely used benchmark in the industry. It contains two image-operation-focused test cases that are similar as FishIETank. In addition, there is also another test case that stresses the vector operations without touching images, like drawing circles.

Somewhat different from these two benchmarks is the Canvas_Perf^[17] benchmark, which consists of a few API-level small test cases. It has a fairly broad coverage of HTML5 Canvas APIs. It evaluates the performance of this set of APIs, rather than one or two typical APIs invoked by FishIETank or GUIMark3.

HTML5 Canvas 2D Implementation in Android 4.0

Although the concept of HTML5 Canvas 2D is straightforward to understand, due to the complexity of web runtime, the underlying implementation actually involves a lot of building blocks and the execution flow usually crosses multiple processes and threads. As a good example, Figure 1 illustrates the high level implementation of the HTML5 Canvas 2D in the stock browser of Android 4.0.

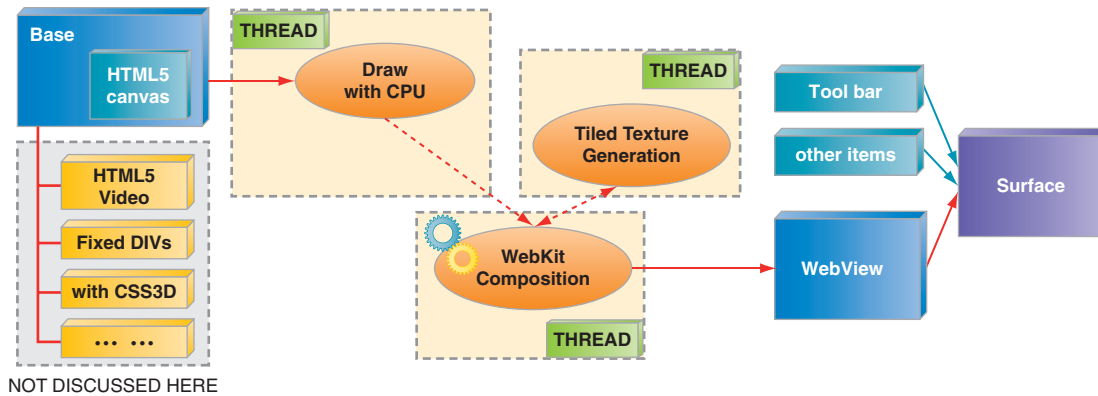


Figure 1: Default implementation of HTML5 Canvas 2D in the browser of Android 4.0
(Source: Intel Corporation, 2012)

In the stock Android browser, there are three different worker threads: the WebViewCore thread, the WebViewMain thread, and the TextureGenerator thread. Each thread serves different purposes.

- The WebViewCore thread generates the contents by utilizing the 2D operations supplied by Skia^[18], the 2D graphics engine of Android. Currently, Skia in Android uses the CPU backend for such operations, which means that related calculations of HTML5 Canvas 2D are conducted in the CPU.
- The WebViewMain thread dispatches UI events and merges the generated contents from multiple layers into one single image for the system to display. The latter process is also known as *composition*, and mostly offloaded to GPU for better UI response since Android 4.0.
- As the CPU-generated contents have to be composited by the GPU, the TextureGenerator thread is introduced to convert the image from the bitmap located in CPU memory to the texture that is available for the GPU. As such conversion is time-consuming, the image is split into pieces called *tiles* to reduce the overhead, and only the tile with updates would be converted as necessary.

At least two drawbacks come with this default implementation based on our analysis. Firstly, the CPU is much less efficient than the GPU at generating the contents with typical image operations such as scaling and rotation. Secondly, the overhead of data exchanging between CPU and GPU in texture generation is too significant to ignore. In our tests on an Intel Atom based mobile platform, more than 20 percent CPU utilization is consumed by memory copies associated with this.

“The overhead of data exchanging between CPU and GPU in texture generation is too significant to ignore.”

Graphics Acceleration of HTML5 Canvas 2D

Graphics acceleration of HTML5 Canvas 2D is a sound approach to address the above issues. By drawing contents with the GPU instead, it would boost the performance of content generation and meanwhile eliminate a large

portion of memory copies because data are already in the GPU. The improved implementation is illustrated in Figure 2. Inspired by Chromium and the implementation of HTML5 Video, we separate out the HTML5 Canvas 2D from other basic web contents and treat it as a standalone layer like HTML5 Video. With this change we are able to substitute the Skia CPU backend with a GPU backend specifically for this canvas layer, thus the GPU can draw the contents in a more efficient and direct way without any further need of texture generation.

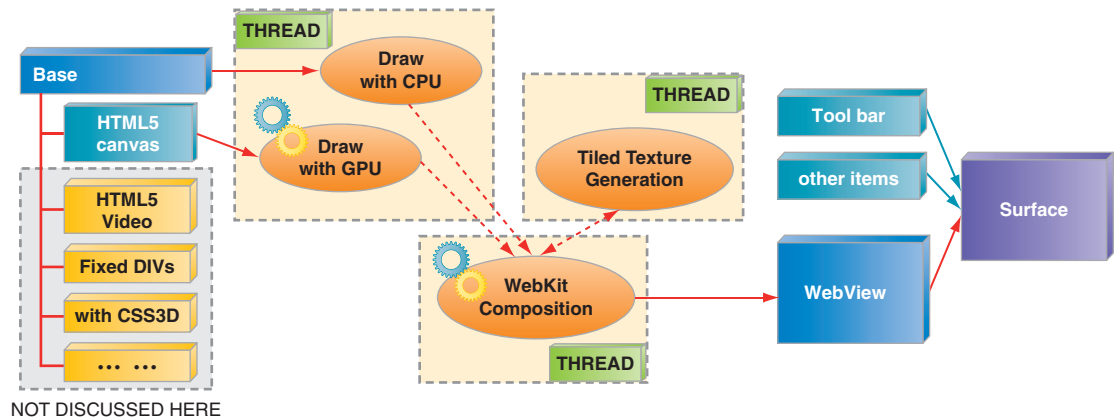


Figure 2: Optimization by using hardware to accelerate the HTML5 Canvas 2D
(Source: Intel Corporation, 2012)

“As expected, this implementation brings significant performance improvement in image-heavy HTML5 Canvas 2D benchmarks.”

As expected, this implementation brings significant performance improvement in image-heavy HTML5 Canvas 2D benchmarks. On the Intel Atom platform, the FPS (frames per second) of the modified FishIETank is increased to as much as three times that of the original solution.

However, it is noticeable that at most 70 percent performance regression is observed on some APIs in the Canvas_Perf benchmark. The analysis shows that GPU acceleration is not always the fastest approach in every HTML5 Canvas 2D API. The Skia CPU backend is better than the GPU backend in certain cases:

- Non-image operations. Vector operations involving multiple vertices can be quite time consuming for the GPU if there is no well-designed 2D graphics unit to support it.
- Certain APIs like `getImageData()`, which need access image data by CPU. This would cause remarkable performance regression if it were GPU accelerated, because in most cases, the GPU has to first be synchronized with the CPU and then must copy data to CPU memory.

This is actually one of the reasons why we need to separate out the HTML5 Canvas 2D rather than apply GPU acceleration to entire web contents.

In order to enjoy the benefit of graphics acceleration for image operations without paying the penalty for some inefficient scenarios, we designed and

implemented a mechanism to dynamically switch between the CPU and GPU path with certain heuristics.

As illustrated in Figure 3, execution of each frame would generate some hints such as the performance indicators and a list of HTML5 Canvas 2D APIs touched. The next frame would choose the suitable path through either GPU or CPU, based on these hints and predefined rules. The first frame always goes through the CPU path because there are no hints available at the very beginning.

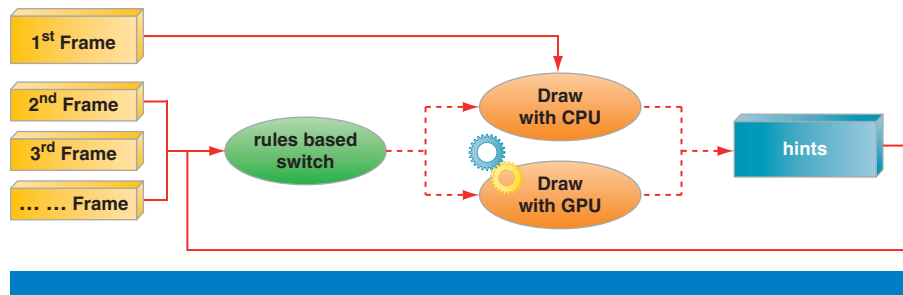


Figure 3: Dynamic GPU and CPU switch for HTML5 Canvas 2D
(Source: Intel Corporation, 2012)

Current rules are quite simple and conservative, and aimed to provide an easy solution without regression compared with the default implementation of Android 4.0. The frame would use the GPU path if all of the following requirements were satisfied:

- System does use GPU to composite;
- Graphics context has been initialized;
- At least one image operation is invoked;
- No very GPU-inefficient API is invoked, for example: no `getImageData()` or some (not all) vector APIs;
- After switches to GPU from CPU path, it never falls back to CPU path any more.

We regard this implementation based on current rules as fallback solution, because the last rule determines that the GPU path could only be switched at most once. Actually, this is very conservative and leaves much room to improve in the future. Even if the execution falls back to CPU for some reason, it could again be suitable to resume GPU acceleration for later frames. A set of smarter and more aggressive rules are under design and are part of our future work.

Impact of Graphics Acceleration and CPU Fallback

Figure 4 and Figure 5 illustrate the performance impact on an Intel Atom based device with the Canvas_Perf benchmark. For the pure GPU solution, Figure 4 indicates up to 70 percent regression in API hline and vline although it is 5 times faster on image-related operations. With the fallback mechanism applied, all of the regressions are resolved while outstanding boost remains for image operation, as illustrated in Figure 5.

It's worth noting that in Figure 5, the CPU fallback path is still faster than the original solution in the stock Android browser, which is also implemented by using a Skia CPU backend. This is because we always separated out the HTML5 Canvas 2D as a standalone layer, which would eliminate several memory copies and some other overhead, even after fallback to CPU.

Figure 6 and Figure 7 show the performance impact over FishIETank and GUIMark. Up to 3 times higher FPS and 50 percent reduction of CPU utilization can be observed at the same time.

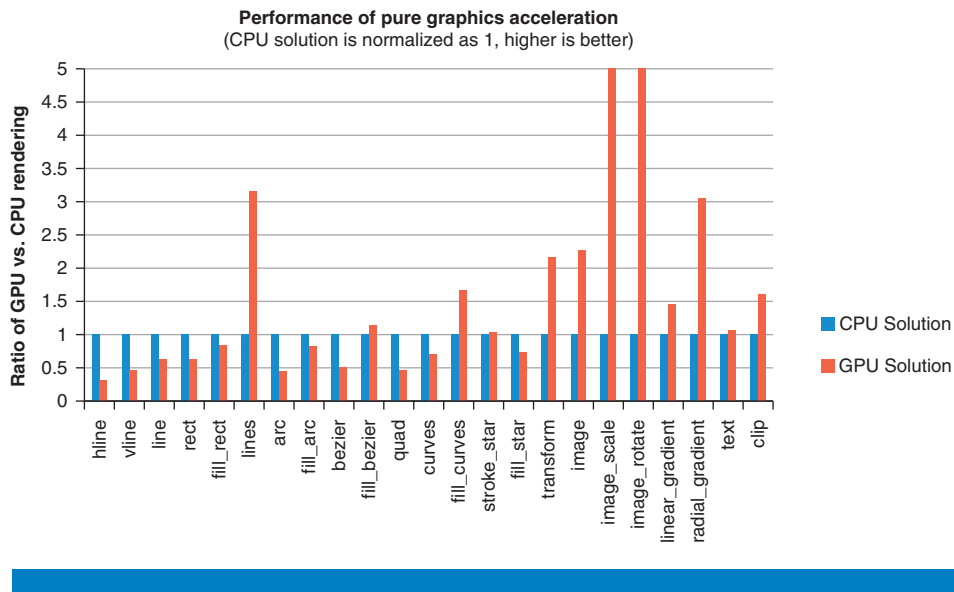


Figure 4: Performance of pure graphics acceleration
(Source: Intel Corporation, 2012)

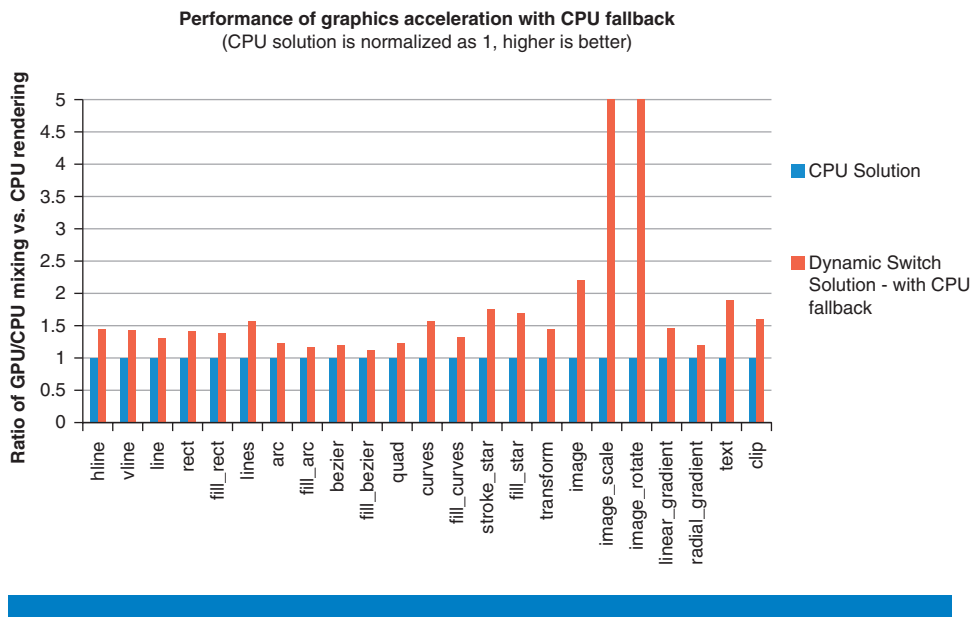


Figure 5: Performance of graphics acceleration with fallback
(Source: Intel Corporation, 2012)

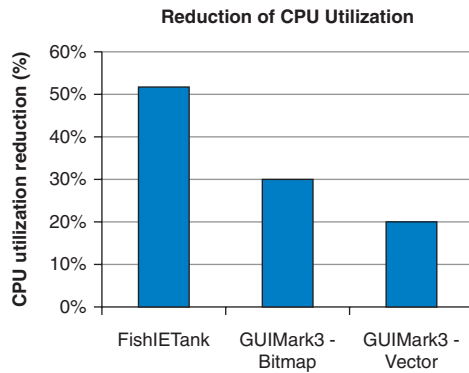


Figure 6: Reduction of CPU utilization
(Source: Intel Corporation, 2012)

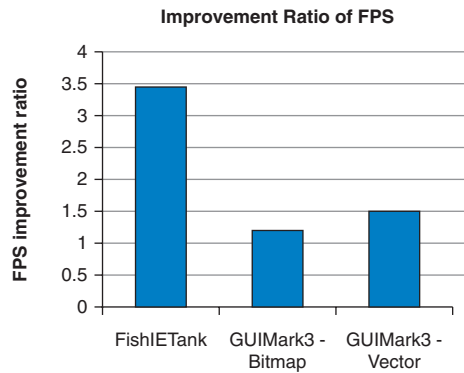


Figure 7: FPS improvement
(Source: Intel Corporation, 2012)

DFG JIT on IA32

JIT is a well-proven technology to dramatically improve the runtime performance of language engines. JavaScript, as a dynamic language, however, has more concerns when applying advanced JIT optimizations. In particular, its types can be changed during the execution and thus are not statically determined. Consequently, most of the advanced optimizations based on accurate type information cannot be applied here. As a result, a lot of special handling becomes necessary, including but not limited to efficient object property accessing, type profiling and speculations, and on-stack-replacement (OSR) mechanism to support hotspot optimizations and deoptimizations due to speculation failures, and so on.

“JIT is a well-proven technology to dramatically improve the runtime performance of language engines.”

JIT technology in JavaScriptCore

There are two types of JITs inside JSC (JavaScriptCore), the default JavaScript engine of WebKit.

- **Baseline JIT:** The baseline JIT is optimized for compilation speed, which generates native code from the JSC bytecode without any heavy optimizations except for the classic inline caching optimization for dynamic languages. As many existing Web sites don't have much frequently executed JavaScript code, this lightweight JIT is a good fit for fast page loading.
- **DFG JIT:** The DFG JIT is a highly optimized JIT for better code generation with the tradeoff of compilation speed. It does type speculations based on the type profiling feedback from the baseline JIT, generates SSA-like DFG IR (Intermediate Representation) from the JSC bytecode, performs optimizations including type inference, local CSE (Common Sub-expression Elimination), local register allocation, and so on, and generates optimized native code from the IR.

JSC uses a tiered model to dynamically switch between these two JITs. The JavaScript code is first compiled to generic native code with the baseline JIT before execution. During the generic code execution, a profiler collects the

hotspot information as well as the type information. When a method or a loop is considered to be hot enough, the DFG JIT is triggered to recompile the method and generate optimized code based on type speculations. Possible speculation failures are possible. In such cases, deoptimization happens and the engine falls back to the generic execution. The JIT infrastructure in JSC is described in Figure 8.

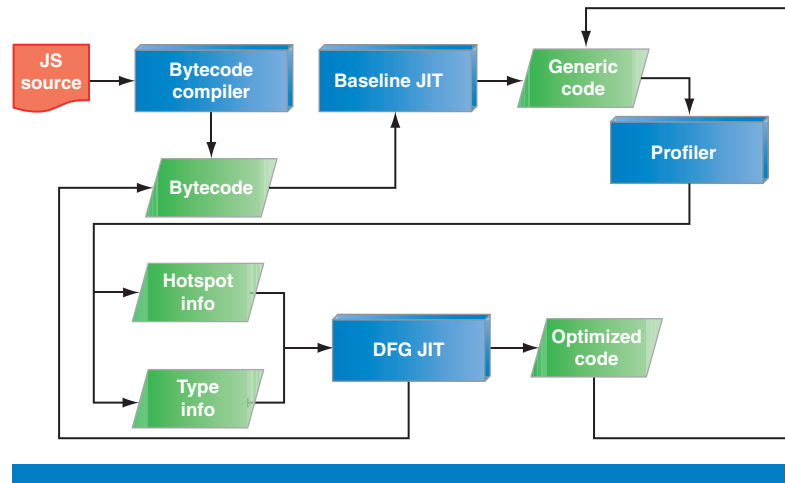


Figure 8: JSC JIT infrastructure
(Source: Intel Corporation, 2012)

It is worth noting that since March 2012, JSC has become a triple-tier VM (virtual machine) on Mac OS X and iOS, which introduces another tier called LLInt (Low Level Interpreter) below the baseline JIT.

“DFG JIT is really sound, especially for JavaScript heavy web applications.”

JavaScriptCore on Intel® Atom™ with 32-Bit Linux

DFG JIT is really sound, especially for JavaScript heavy web applications, yet when it first came out it only provided the 64-bit version, so the 32-bit mobile systems could not benefit from it. For instance, on the Kraken JavaScript benchmark suite^[20], which is constituted of some complex test cases for audio and image processing extracted from rich web applications, JSC without DFG JIT performs 2.5X slower than Google V8 on the Intel Atom platform with Linux. Therefore, we need to do more optimizations to process those programs more efficiently and the DFG JIT is a good fit for this purpose. Furthermore, with the tiered compilation model, the baseline JIT could remain simple enough for minimal compilation overhead. All of these facts tell us we need to bring the DFG JIT technology to more platforms for a broad usage, and our first target is IA32 Linux.

Applying DFG JIT to IA32

Usually in JavaScript engines, some specific data format is defined to represent different JavaScript values including pointers to the objects, integers, doubles, Booleans, undefined, and null values. We call the original values *unboxed* while

the specially formatted ones are referred to as *boxed* JSValues. JSC has two different data formats for 64-bit and 32-bit platforms respectively. Although both formats use a 64-bit length data to represent a JSValue, they have different notations for each bit.

As illustrated in Figure 9, on 32-bit platforms, the higher 32 bits of a JSValue is used as a tag to denote the type, and the lower 32 bits are the payload—the actual value, except for the doubles, which are represented with the total 64 bits. This makes use of unused NaN space for values other than doubles.

The data format for 64-bit platforms also utilizes unused NaN space, while it takes the higher 16 bits to denote the type, as illustrated in Figure 10.

Considering this background, the major challenge of enabling DFG JIT on 32-bit platforms is to ensure that DFG JIT is able to recognize a totally different data format. It's unfortunate that we need to write the same logic twice for those operations on boxed JSValues, one for 64-bit, which is already there, and the other for 32-bit to be added by us. On the other hand, it's fortunate that the DFG JIT is a type-speculative JIT and many operations can be performed on unboxed values, which can be shared between 64-bit and 32-bit.

Compared with x64, one major problem of x86 is the shortage of registers. There are only eight GPRs (general purpose registers) available on x86. In JSC, three of them are already reserved for special purposes so only five are left. In order to address the register pressure problem on x86, we need to have more values to be speculated and represented with unboxed 32-bit values. For example, the Boolean speculation for 32-bit is different from that for 64-bit, which uses a single 32-bit GPR to hold the unboxed Boolean value instead of two GPRs for a boxed JSBoolean. This design choice is made not only for performance, but also to save a register.

We have JS values and sometimes we need to speculate that a JS value is a specific type so that we can generate more efficient code. This inevitably involves value boxing and unboxing. Now considering the simple case where the value is in registers, since the JS values are 64-bit long values, we need to use two general purpose registers to represent the JS value on 32-bit platforms—one for the payload and the other for the tag. Unboxing a JS integer, Boolean, or pointer is cheap by simply adopting the register holding the payload. Similarly, boxing an integer, Boolean, or pointer just needs to fill the tag register with the correct data. What makes things more complex, however, is how we do conversion between JS doubles and unboxed doubles. Bear in mind that the JS double is in two general purpose registers and the unboxed double is in one floating point register. To do the conversion, one straightforward approach is to exchange the data through memory, while it results in very bad performance. On the other hand, we can exploit SSE2 packed data support to perform efficient conversion if the unboxed double is

```
Integer {   FFFF:FFFF:IIII:IIII
Boolean {   FFFF:FFFE:BBBB:BBBB
Pointer {   FFFF:FFFB:PPPP:PPPP
           / FFFF:FFF8:****:****
Double {    \ 0000:0000:****:****
```

Figure 9: JSC data format for 32-bit platforms

(Source: Intel Corporation, 2012)

```
Pointer {    0000:PPPP:PPPP:PPPP
           / 0001:****:****:****
Double {    \ FFFE:****:****:****
Integer {    FFFF:0000:IIII:IIII
```

Figure 10: JSC data format for 64-bit platforms

(Source: Intel Corporation, 2012)

actually in a XMM register. For example, the double boxing and unboxing can be performed using the below sequence respectively.

```
movd xmm0, eax
psrlq 32, xmm0
movd xmm0, edx
```

Code 1: Boxing a double value
Source: Intel Corporation, 2012

```
movd eax, xmm0
movd edx, xmm1
psllq 32, xmm1
por xmm1, xmm0
```

Code 2: Unboxing a JS double
Source: Intel Corporation, 2012

In this example, we suppose *eax* holds the payload of the JS double and *edx* holds the tag of it. Furthermore, *xmm0* holds the unboxed double value and *xmm1* is for temporary usage purpose. The example shows that the conversions between JS double and unboxed double no longer need to involve memory access, and as a proof, we get a 77-percent performance boost for the Kraken benchmark on IA32 by this approach. Over the long term, we may further improve double conversions if we know a JS value is a JS double at compile time and directly represent it with a FPR (floating pointer register) instead of two GPRs, though it may introduce additional complexities to the code generation logic in DFG.

Besides normal DFG JIT code generation, the different data format also impacts the deoptimization caused by speculation failures in DFG JIT. The generic code to be switched to always assumes that boxed JS values are in memory. However in fact, the DFG JIT code can produce unboxed values in both memory and registers. Consequently, when falling back to the generic code, we have to perform necessary data boxing and data transfers between memory and registers.

“When falling back to the generic code, we have to perform necessary data boxing and data transfers between memory and registers.”

The calling conventions for x86 differ from those of x64, and in fact there are different conventions for different operating systems on x86. So another thing we need to do is to teach the DFG JIT the different calling conventions, which is necessary as we have some runtime helper functions invoked by the JIT code. We now support the cdecl calling convention for x86 in DFG, different from the fastcall support in the baseline JIT. The helper function call interfaces are also redesigned with the help from the community to support more different calling conventions easily.

Impact of DFG JIT on IA32

Figure 11 clearly reveals the performance improvement due to the enabling of DFG JIT. Nearly 2X improvement on IA32 for Kraken benchmark is observed on the November 2011 code base when we eventually finished our

enabling. Furthermore, our further collaboration with the community on the optimizations over the DFG JIT results in more performance improvement from November 2011 through March 2012.

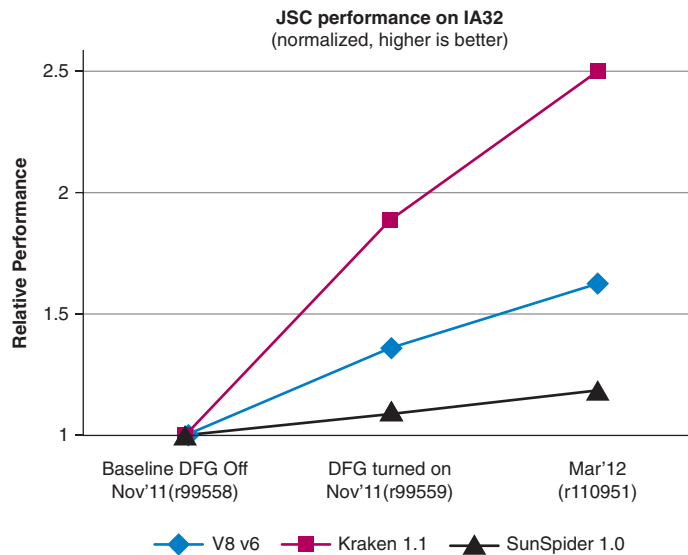


Figure 11: JSC performance on IA32
(Source: Intel Corporation, 2012)

Conclusion and Future Work

HTML5 is an inevitable trend of future web applications. It brings massive opportunities while it results in novel challenges in performance as well. Therefore, optimizations of web runtime become essential in the journey towards embracing HTML5. This article shares two important technologies, graphics acceleration and JIT, on the rendering engine and JavaScript engine respectively, with prominent performance gain achieved on Intel Atom based platforms.

Looking forward, several further improvements are pipelined as our future work:

- As mentioned in the section “Graphics Acceleration of HTML5 Canvas 2D,” a more aggressive set of rules to switch between CPU and GPU path is under design. We believe that the new design would maximize the opportunity to utilize graphics acceleration and bring the performance of HTML5 Canvas 2D to next level.
- We also have a few ideas to improve the Skia GPU backend. This would mitigate some inefficient implementation of certain vector-related APIs on the GPU.
- Regarding JSC JavaScript engine, its IR of DFG JIT is still quite simple and lacks of some advanced features. We are working together

“Optimizations of web runtime become essential in the journey towards embracing HTML5.”

with community engineers to improve this. With a more powerful implementation, adding a lot of typical JIT optimizations over JSC is expected to be much easier.

As graphics acceleration and advanced JIT are compelling technologies with high potential; we are exploring the feasibility of applying them to the implementation of other emerging web technologies such as CSS3 and WebGL, too.

Complete References

- [1] HTML5 specification (draft 25), World Wide Web Consortium (W3C), May 2011. <http://www.w3.org/TR/2011/WD-html5-20110525/>
- [2] Processor family for mobile segments, Intel. <http://www.intel.com/content/www/us/en/processors/atom/atom-processor.html>
- [3] Gartner, 2011. <http://www.gartner.com/it/page.jsp?id=1826214>
- [4] HTML5 Angry Birds, Rovio Entertainments Ltd. <http://chrome.angrybirds.com/>
- [5] Google to use HTML5 in Gmail. http://www.computerworld.com/s/article/9178558/Google_to_use_HTML5_in_Gmail?taxonomyId=11&pageNumber=2
- [6] Brad Neuberg, Google. Introduction to HTML5. <http://googlecode.blogspot.com/2009/09/video-introduction-to-html-5.html>
- [7] Mobile operating system, Google. <http://www.android.com/>
- [8] Open source project for browser and mobile OS, Google. <http://www.chromium.org/>
- [9] A New Crankshaft for V8, Google. <http://blog.chromium.org/2010/12/new-crankshaft-for-v8.html>
- [10] V8 JavaScript Engine. <http://code.google.com/p/v8/>
- [11] V8 Benchmark Suite Version 6. <http://v8.googlecode.com/svn/data/benchmarks/v6/run.html>
- [12] The WebKit Open Source Project. <http://www.webkit.org/>
- [13] HTML Canvas 2D Context, World Wide Web Consortium (W3C). <http://www.w3.org/TR/2dcontext/>
- [14] JavaScript Web API standards, World Wide Web Consortium (W3C). <http://www.w3.org/standards/webdesign/script>
- [15] FishIETank workload, Microsoft. <http://ie.microsoft.com/testdrive/Performance/FishIETank/Default.html>

- [16] Browser graphics benchmark, Sean Christmann. <http://www.craftymind.com/guimark3/>
- [17] Browser graphics benchmark, Hatena. http://flashcanvas.net/examples/dl.dropbox.com/u/1865210/mindcat/canvas_perf.html
- [18] Open project as 2D graphics engine, Google. <http://code.google.com/p/skia/>
- [19] SunSpider JavaScript Benchmark. <http://www.webkit.org/perf/sunspider/sunspider.html>
- [20] Kraken JavaScript Benchmark. <http://kraken-mirror.googlecode.com/svn/trunk/kraken/hosted/index.html>

Author Biographies

Jonathan Ding (jonathan.ding@intel.com) is a software engineer in the Software and Services Group of Intel. His expertise covers browser, web runtime, HTML5, and related service frameworks.

Yuqiang Xian (yuqiang.xian@intel.com) is a software engineer in the Software and Services Group of Intel. His major interests include compilers and virtual machines.

Yongnian Le (yongnian.le@intel.com) is a software engineer in the Software and Services Group of Intel. He currently focuses on browser-related analysis and optimizations, in particular on Android mobile platforms.

Kangyuan Shu (kangyuan.shu@intel.com) is a software engineer in the Software and Services Group of Intel. His interests include Android and graphic subsystems.

Haili Zhang (haili.zhang@intel.com) is a software engineer in the Software and Services Group of Intel. His expertise covers HTML5 and open web platform related application frameworks and tools.

Jason Zhu (jason.zhu@intel.com) is a software engineer in the Software and Services Group of Intel. His interests include advanced and emerging web technologies and innovations.

MOBILE OS ARCHITECTURE TRENDS

Contributors

Xiao-Feng Li

Software and Services Group,
Intel Corporation

Yong Wang

Software and Services Group,
Intel Corporation

Jackie Wu

Software and Services Group,
Intel Corporation

Kerry Jiang

Software and Services Group,
Intel Corporation

Bing Wei Liu

Software and Services Group,
Intel Corporation

The world is flat, because it becomes increasingly mobile, fast, connected, and secure. People expect to move around easily with their mobile devices, keeping close communications with their partners and family, enjoying the versatile usage models and infinite contents, and without worrying about the device and data management. These all put requirements on the mobile devices, of which the mobile OS is the soul. Based on our years of experience in mobile OS design and an extensive survey of current industry situation, we believe there are several commonalities in future mobile OS architecture, such as user experience, power management, security design, cloud support, and openness design. We develop an analysis model to guide our investigation. In this article, we describe our investigations in the trends of mobile OS architecture over the next decade by focusing on the major commonalities. Based on the findings, we also review the characteristics of today's mobile operating systems from the perspective of architecture trends.

Introduction

Mobile OS design has experienced a three-phase evolution: from the PC-based OS to an embedded OS to the current smartphone-oriented OS in the past decade. Throughout the process, mobile OS architecture has gone from complex to simple to something in-between. The evolution process is naturally driven by the technology advancements in hardware, software, and the Internet:

- *Hardware.* The industry has been reducing the factor size of microprocessors and peripherals to design actual mobile devices. Before the form factor size was reduced enough, the mobile device could not achieve both small size and processing capability at the same time. We had either a PC-sized laptop computer or a much weaker personal data assistant (PDA) in phone size. Mobile operating systems for PDAs usually did not have full multitasking or 3D graphics support. Features like sensors, such as accelerometers, and capacitor-based touchscreens were not available in the past mobile operating systems.
- *Software.* With a laptop computer, the software is mainly focused on the user's productivity, where support for keyboard and mouse that have precise inputs are essential. The software for a personal data assistant, as its name implies, helps the user to manage personal data such as contacts information, e-mail, and so on. The mobile operating systems were not designed for good responsiveness or smoothness with a rich user interface (UI) including both touchscreen and other sensors.

- *Internet.* Along with Internet development, especially after Web 2.0, there is abundant information in the network waiting to be searched, organized, mined, and brought to users. People are increasingly living with the Internet instead of just browsing the Web. More and more people are involved in the development, including information contribution, application development, and social interactions. The mobile operating systems cannot be self-contained, but have to be open systems.

The usage model of past mobile devices is limited. A user mostly runs the device applications for data management and local gaming, only occasionally browses Internet static Web pages or accesses specific services like e-mail. In other words, the possible usages of the device are predefined with the preinstalled applications when the user purchases it. This is largely changed in new mobile devices, where the device is virtually a portal to various usage models. All the involved parties such as service providers, application developers, and other device users continuously contribute and interact through the device with its owner. Figure 1 shows the high-level usage model difference between the past and new mobile devices.

The usage model of past mobile devices is limited.

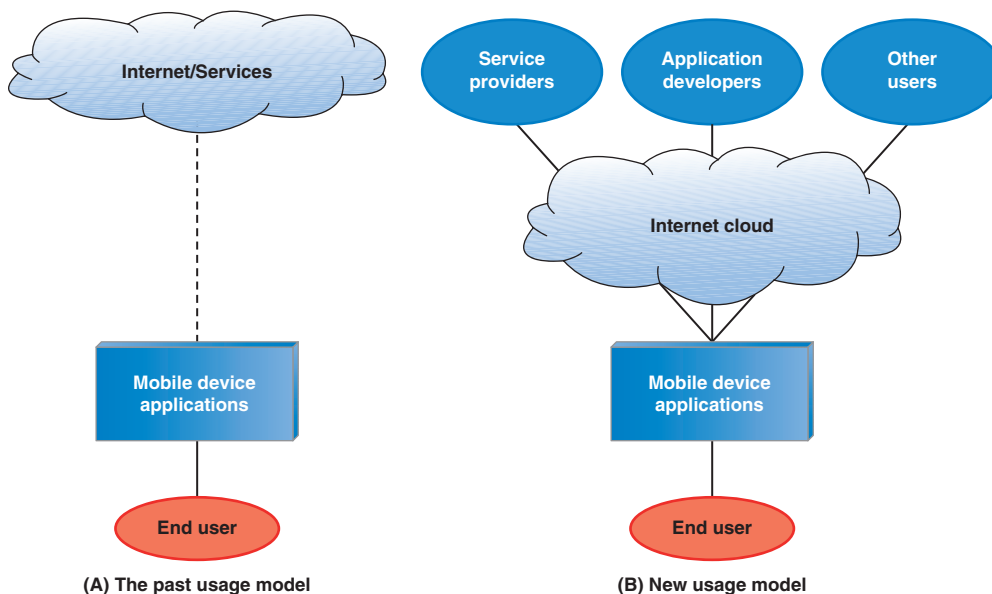


Figure 1: High-level usage models of mobile devices

(Source: Intel Corporation, 2012)

The representatives of current mobile operating systems include Apple's iOS* 5.0, Google Android* 4.0, Microsoft Windows* Phone 7.0, and a few others. In terms of their usage models, they share more similarities than differences:

- All of them have a documented software development kit (SDK) with well-defined APIs that enable the common developers to develop applications for these systems.

“System security is always a key concern to the device users.”

“As of today, the thin client model still has technical challenges in user experience and security.”

- All of them have online application stores for the developers to publish and for the users to download applications, such as Apple App Store, Google Play, and Windows Phone Marketplace.
- All of them have a certain level of multitasking and 3D graphics support. Touchscreens and sensors are just no-brainers. Much effort have been spent to make the user interactions smooth and responsive.
- Browsing experience is far beyond static Web pages. HTML5 is becoming the default so as to run Web-based applications.
- All of the operating systems support device-based payment. Together with enterprise applications and private information, system security is always a key concern to the device users.
- As mobile operating systems, one of key design differences from non-mobile operating systems is the focus on battery life. The systems try best to reduce the active power consumption of the device components and put them into idle whenever possible.

The similarities of the current mobile operating systems reflect the advancement trend in hardware, software, and the Internet. Anticipating the trend of mobile operating systems, we believe those areas are the major focuses of the next generation of mobile OS design, including user experience, battery life, cloud readiness, security, and openness. They are actually conflicting targets to a large extent:

- *User experience and battery life.* To achieve best responsiveness and smoothness, the system expects all the hardware resource available to exploit their best capacity. At the same time, to sustain the battery life as a mobile device, the hardware components should be idle whenever possible.
- *Security and openness.* One does not want to expose all of one's system functionalities to external entities, because that puts the system under security threat. On the other hand, without exposing enough system APIs, it is impossible for the developers to create innovative usages.
- *Cloud readiness.* As more and more services and applications are offered from the cloud, it is natural to consider a thin client device model that trusts the cloud and that offloads computations to the cloud. But as of today, the thin client model still has technical challenges in user experience and security.

In this article, we try to investigate the various aspects of mobile OS design and present our opinions about the future of mobile OS architecture.

The article is arranged as follows. Based on the framework laid out in this section, we use separate sections to discuss the respective topics in text below. They are arranged in user experience, power management, security, openness, and cloud readiness. Finally we have discussions and a summary.

User Experience (UX)

Traditional performance is inadequate to characterize modern client devices. *Performance* is more about the steady execution state of the software stack and

is usually reported with a final score of the total throughput in the processor or other subsystems. *User experience* is more about the dynamic state transitions of the system triggered by user inputs. The quality of the user experience is determined by such things as the user perceivable responsiveness, smoothness, coherence, and accuracy. Traditional performance could measure every link of the chain of the user interaction, while it does not evaluate the full chain of the user interaction as a whole. Thus the traditional performance optimization methodology cannot simply apply to the user experience optimization. It is time to invest in device user interaction optimizations so as to bring the end user a pleasant experience.

User Interactions with Mobile Devices

In a recent performance measurement with a few market Android devices, we found there was a device X behaving uniformly worse than another device Y with common benchmarks in graphics, media, and browsing. But the user perceivable experience with the device X was better than device Y. The root reason we identified was that traditional benchmarks or benchmarks designed in traditional ways did not really characterize user interactions, but measured the computing capability (such as executed instructions) or the throughput (such as processed disk reads) of the system and the subsystems.

Take video evaluation as an example. Traditional benchmarks only measure video playback performance with some metrics like FPS (frames-per-second), or frame drop rate. This methodology has at least two problems in evaluating user experience. The first problem is that video playback is only part of the user interactions in playing video. A typical life cycle of user interaction usually includes at least the following links: “launch player” → “start playing” → “seek progress” → “video playback” → “back to home screen.” Yet good performance in video playback cannot characterize the real user experience in playing video. User interaction evaluation is a superset of traditional performance evaluation.

The other problem is, using FPS as the key metric to evaluate the smoothness of the user interactions cannot always reflect good user experience. For example, when we flung a picture in the Gallery3D application, the device Y had obvious stuttering during the picture scrolling, but the FPS value of device Y was higher than that of device X. In order to quantify the difference of the two devices, we collected the data of every frame during a picture fling operation in the Gallery3D application on both device X and device Y, as shown in Figure 2 and Figure 3 respectively. Every frame’s data is given in a vertical bar, where the x-axis is the time when the frame is drawn, and the height of the bar is the time it takes the system to draw the frame. From the figures, we can see that device X obviously has a lower FPS value than device Y, but with smaller maximal frame time, less frames longer than 30 ms, and smaller frame time variance. This means that, to characterize the user experience of the picture fling operation, those metrics like maximal frame time and frame time variance should also be considered.

“The traditional performance optimization methodology cannot simply apply to the user experience optimization.”

“Good performance in video playback cannot characterize the real user experience.”

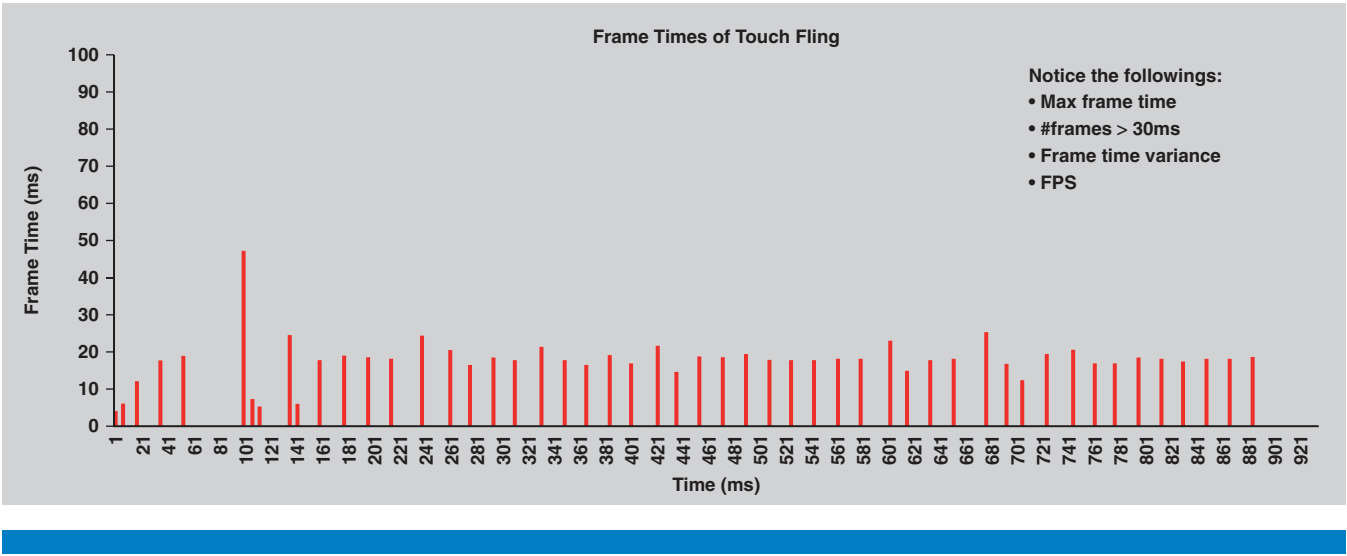


Figure 2: Frame times of a fling operation in Gallery3D application on device X
(Source: Intel Corporation, 2011)

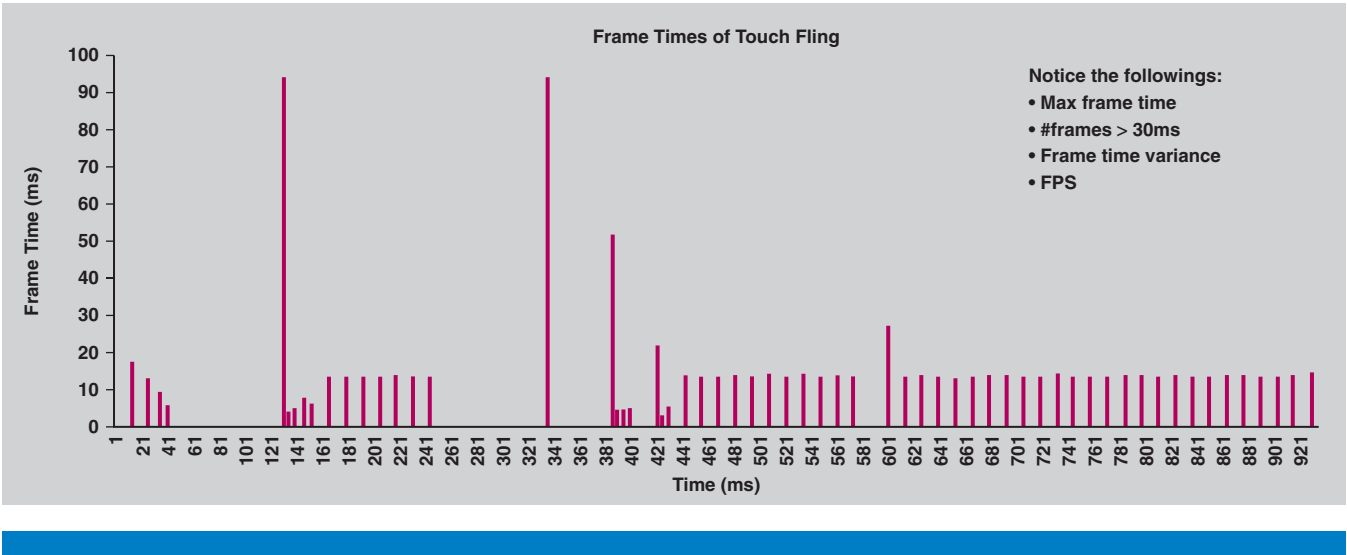


Figure 3: Frame times of a fling operation in Gallery3D application on device Y
(Source: Intel Corporation, 2011)

As a comparison, Figure 4 shows the frame data of a fling operation after we optimized the device Y. Apparently all the metrics have been improved and the frame time distribution became much more uniform.

User experience is more about dynamic state transitions of the system triggered by user inputs. A good user experience is achieved with things such as user perceivable responsiveness, smoothness, coherence, and accuracy. Traditional performance could measure every link of the chain of the user interaction without evaluating the full chain of the user interaction as a whole.

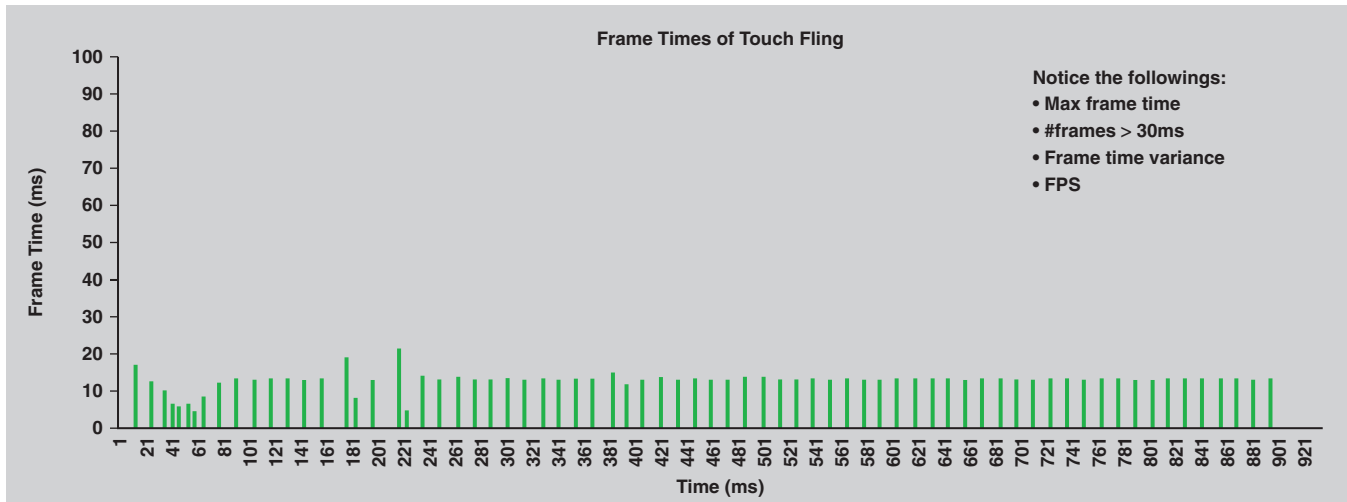


Figure 4: Frame times of a fling operation in Gallery3D application on device Y after optimization
(Source: Intel Corporation, 2011)

Another important note is that user experience is a subjective process; just consider the experience when watching a movie or appreciating music. Current academic research uses various methodologies such as eyeball tracking, heartbeat monitoring, or just polling to understand user experience. For our software engineering purpose, in order to analyze and optimize the user interactions systematically, we categorize the interaction scenarios into four kinds:

- *Inputs to the device from the user, sensor, network, and so on.* This category evaluates if the inputs can trigger the device to action accurately or fuzzily as expected. For touchscreen inputs, it measures the touch speed, pressure, range, and so forth.
- *Device response to the inputs.* This category evaluates how responsive the device is to the inputs.
- *System state transition.* This category especially evaluates how smooth graphics transition on the screen. It can be a follow-up of the device response to some input.
- *Continuous control of the device.* People operating the device not only give a single input, but sometimes also control the graphic objects in the screen, such as to control a game jet-plane, or to drag an application icon. The category is to evaluate the controllability of the device.

Among them, “inputs to the device” and “control of the device” are related to the user experience aspect of *how a user controls a device*. “Device response to the inputs” and “system state transition” are related to the aspect of *how the device reacts to the user*. We can map a user interaction life cycle into scenarios that fall into the categories above; then for each scenario, we can identify the key metrics in the software stack to measure and optimize.

User Interaction Optimizations

As we have described in last subsection, there is no clear-cut and objective way to measure the user experience. We set up following criteria in our measurement of the user interactions:

- *Perceivable*. The metric has to be perceivable by a human being. Otherwise, it is irrelevant to the user experience.
- *Measureable*. The metric should be measurable by different teams. It should not depend on certain special infrastructure that can only be measured by certain teams.
- *Repeatable*. The measured result should be repeatable in different measurements. Large deviations in the measurement mean that it is a bad metric.
- *Comparable*. The measured data should be comparable across different systems. Software engineers can use the metric to compare the different systems.
- *Reasonable*. The metric should help reason the causality of software stack behavior. In other words, the metric should be mapped to the software behavior and it should be possible to be computed based on software stack execution.
- *Verifiable*. The metric can be used to verify an optimization. The measured result before and after the optimization should reflect the change of the user experience.
- *Automatable*. For software engineering purposes, we expect the metric can be measured largely unattended. This is especially useful in regression tests or pre-commit tests. This criterion is not strictly required though, because it is not directly related to user experience analysis and optimization.

Guided by the measurement criteria, we focus on the following complementary aspects of the user experience, how a user controls a device and how a device reacts to a user. How a user controls a device has mainly two measurement areas:

- *Accuracy/fuzziness*. It evaluates what accuracy, fuzziness, resolution, and range are supported by the system for inputs from the touch screen, sensors, and other sources. For example, how many pressure levels are supported by the system, how the sampled touch events' coordinates are close to the fingertip move track on the screen, how many fingers can be sampled at the same time, and so on.
- *Coherence*. It evaluates the drag lag distance between the fingertip and the dragged graphic object in the screen. It also evaluates the coherence between the user operations and the sensor-controlled objects, such as the angle degree difference between the tilting controlled water flow and the device oblique angle.

How a device reacts to a user also has two measurement areas:

- *Responsiveness*. It evaluates the time between an input being delivered to the device and device showing visible response. It also includes the time spent to finish an action.

“How a user controls a device has mainly two measurement areas.”

- *Smoothness.* This area evaluates graphic transition smoothness with the maximal frame time, frame time variance, FPS, frame drop rate, and so forth. As we have discussed, FPS alone cannot accurately reflect the user experience regarding smoothness.

For these four measurement areas, once we identify a concrete metric to use, we need to understand how this metric is related to a “good” user experience. Since user experience is a subjective topic that highly depends on human being’s physiological status and personal taste, there is not always scientific conclusion about what value of a metric constitutes a “good” user experience. For those cases, we just adopt the industry experience values. The Table 1 gives some examples of the industry experience values.

	Best	Good	Acceptable
Response delay	≤ 100 ms	≤ 200 ms	≤ 500 ms
Graphics animation	≥ 120 fps	≥ 60 fps	≥ 30 fps
Video playback	≥ 60 fps	≥ 30 fps	≥ 20 fps

Table 1: Example Industry Experience Values for User Experience
(Source: Intel Corporation, 2011)

Due to human nature, there are two notes for software engineers to pay attention to the user experience optimizations.

The value of a metric usually has a range for “good” user experience. A “better” value than the range does not necessarily bring “better” user experience. Anything beyond the range limit could be invisible to the user.

The values here are only rough guidelines for common cases with typical people. For example, a seasoned game player may not be satisfied with the 120 fps animation. On the other hand, a well-designed cartoon may bring perfect smoothness with 20 fps animation.

Now we can set up our methodology for user experience optimization. It can be summarized into following steps.

Step 1: Receive the user experience *sightings* from the users or identify the interaction issues with manual operations. This can be assisted by high-speed camera or other available techniques.

Step 2: Define the software stack scenarios and metrics that transform a user experience issue into a software symptom.

Step 3: Develop a software *workload* to reproduce the issue in a measureable and repeatable way. The workload reports the metric values that reflect the user experience issue.

Step 4: Use the workload and related *tools* to analyze and optimize the software stack. The workload also verifies the optimization.

Step 5: Get *feedback* from the users and try more applications with the optimization to confirm the user experience improvement.

“A seasoned game player may not be satisfied with the 120 fps animation. A well-designed cartoon may bring perfect smoothness with 20 fps animation.”

“The UX optimization is somehow similar to parallel application optimization.”

“Every application uses an event loop to handle requests.”

Based on this methodology, we have developed an Android Workload Suite (AWS)^[33] that includes almost all the typical use cases of an Android device. We have also developed a toolkit called UXtune^[34] that assists user interaction analysis in the software stack. Different from the traditional performance tuning tools, UXtune correlates the user-visible events and the system low-level events. As the next step, we are porting the work from Android to other systems.

Mobile OS Design for User Experience

Based on our experience with Android, we found the UX optimization is somehow similar to parallel application optimization, only with more complexities, for the following four reasons:

- UX involves multiple hardware components, and multiple software processes, and their interactions;
- UX on a client device has to consider the power consumption, because UX also includes the battery life and device temperature.
- UX has precise timing requirements, such as smoothness where the user expects no frame time variance. Neither faster nor slower is acceptable; hitting the exact time point is required. This point is more like a real-time requirement.
- UX has some subjective factors one has to consider in mobile OS design, such as whether some animation is just a hint or essential to UX, and whether the system can drop some frames to get better response.

One critical lesson learned from our experience is to understand the critical path of certain operations. Different from parallel application tuning, mobile OS design does not always have strict explicit synchronization between the involved hardware components and software threads. For example:

- Every application uses an event loop to handle requests. When a thread A has a request to another thread B, it may not directly invoke the function, but instead posts a message to thread B. The message is then queued in the event loop of thread B waiting for handling. It is then out of the caller's control how fast the event could be handled if there are multiple events in the queue.
- Another example is when a thread A executes a sequence of operations and then posts a message to another thread B for follow-up actions and response to the user. Not all the sequence of the operations by thread A must be done in order. For example, it could post the message to thread B as early as possible so that thread B can respond to the user earlier.

The major point regarding power is that, with user experience, faster is not necessarily better—contrary to traditional performance optimizations. When the system already reaches the user-perceivable best responsiveness, next step optimization is the slower the better within the best perceivable range. For example, if a game has 60 fps without problem, then the mobile OS should try to get both the CPU utilization and CPU frequency as low as possible. We

have to always distinguish the two cases (faster-better and slower-better) clearly. The optimization techniques for the two cases can be quite different.

When multiple cores and GPUs are introduced, the two cases above become more obvious. More cores help “faster-better” scenarios usually, but hurt battery life for “slower-better” scenarios. A mobile OS has to turn on and off the cores with a smart policy, because the on/off latency could be longer than a touch operation (usually in a range of 100 ms to 500 ms).

For parallel application performance tuning, people found “execution replay” useful in debugging. It is usually one multithreaded application replay, that is, within one process. For UX, the interactions cross processes through IPC, and between CPU, GPU, and display controller (DC), are a whole-system-wide collaboration. The traditional replay technique cannot help.

“A mobile OS has to turn on and off the cores with a smart policy.”

Power Management (PM)

Power management has always been a key challenge to mobile OS designers and will be even more so moving forward. Power demands are increasing rapidly on mobile devices as more and more power hungry applications are developed for mobile platforms. However, battery capacity growth could never keep up in the meantime due to both the slow development in battery technologies and the fact that people want more sleek and compact form factors that could fit into a pocket. Power management is becoming an increasingly complex problem on mobile devices and a holistic approach needs to be employed to address it.

“Power management is becoming an increasingly complex problem on mobile devices.”

Processor Power Management

Mobile operating systems have been making steady progress in the power management area in the last decade. Initially the focus of mobile OS power management work had been on processor power management since the processor had long been the most significant consumer of total platform power. Modern processors support dynamic frequency and voltage scaling such as Enhanced Intel SpeedStep® Technology. Such processor capabilities enabled mobile operating systems to adjust processor frequency and/or voltage dynamically at runtime based on the demand of computing power required by the workload that is currently running on the processor. This saves a significant amount of processor power while it is active since consumed power is proportional to the square of core voltage and frequency. The cpufreq subsystem in the Linux kernel is an example of managing processor power while it is active. In addition to dynamic frequency and voltage scaling, modern processors typically support multiple processor idle states with varying amounts of power consumed in those idle states. The deeper the idle state, the more power could be saved although at the expense of longer entry and exit latency. A mobile OS could direct the processor to enter an appropriate idle state based on predicted idleness and QoS constraints posed by other subsystems and user space. Linux’s cpuidle subsystem is an example of power managing a processor while it is idle.

“Modern GPUs are starting to support dynamic voltage and frequency scaling similar to that found on CPUs.”

Device Power Management

The focus of mobile OS power management work has shifted to device power management. In particular, a mechanism has been introduced to manage the power of I/O devices at runtime. Runtime power management for I/O devices could automatically put I/O devices into whatever appropriate low power states they support when the corresponding devices are detected as idle at runtime. In addition to managing the power of I/O devices while they are idle, there are some technology innovations to save I/O device power while they are active. For example, modern GPUs are starting to support dynamic voltage and frequency scaling similar to that found on CPUs. GPU dynamic voltage and frequency scaling could reduce power consumption by as much as 50 percent for mobile 3D graphics in some cases. In addition, I/O devices are becoming smarter in the sense that they can work on their own without CPU intervention. For example, technologies like panel self-refresh could save a significant amount of power while the image is static in the cases like when a user is reading a book on a mobile device. The display panel could keep rendering from its local memory in this case and many hardware components that traditionally must be working while rendering display could be shut down, including CPU, memory, display engine and display port.

Mobile OS Power Management Cases

Android gained momentum and became popular before the infrastructure of device runtime power management was introduced into the Linux kernel and it came up with another approach called *opportunistic suspend* in order to achieve the goal of extending battery life on Android devices. Without device runtime power management capabilities, Android tries to suspend the system aggressively whenever there is no interesting work going on. This is indicated by no one holding a *wakelock*.

Windows 8 introduced a new system power state called *connected standby*. Unlike traditional S3 standby, which halts all system activities, the system is still running though in an extremely low power state and this enables users to stay up to date with the latest information such as their e-mails. Windows 8 connected standby is based on processor idle power management and device runtime power management technologies.

“Software hygiene is the most challenging problem for both suspend- and idle-based power management approaches.”

Software hygiene is the most challenging problem for both suspend- and idle-based power management approaches, and the battery life depends very much on application behaviors on such systems. A recent study says that free Android applications waste 75 percent of power on ads by holding a *wakelock* in the background, thus block system suspend. This is also true for Windows 8, where one power-unfriendly application staying busy for no good reason will prevent the entire system from entering connected standby power state. Some people expect the system power management to be more robust even in the face of such power-unfriendly applications by introducing more capable mechanisms, while others think that putting such a mechanism in place will only lead to the proliferation of more ill-behaved applications.

Openness

Another major trend of mobile operating systems is the openness. In this context, the openness of mobile operating systems means the level of opportunity and freedom that people have to use, contribute to, customize, and innovate for the mobile OS for their purposes. There is already work^[6] that has studied the openness from the developers' perspective. The work has identified the facets that decide the developers' perception of the platform openness. Here we study the trend of openness from the ecosystem perspective. As we believe the openness matters to enable and foster the mobile ecosystem.

“Another major trend of mobile operating systems is the openness.”

Openness to Players of the Mobile Ecosystem

The players of the mobile ecosystem include manufacturers (OEMs) who make and sell mobile devices, the service providers (operators) who provide network connection and other value-added services, consumers (end users), the ISVs who develop commercial applications, and developers from communities who develop applications and even contribute on mobile OS development and evolution if the mobile OS is open-sourced.

The openness of mobile operating systems implies different things to different players in the mobile ecosystem. For operators, the openness of the mobile OS determines how easy their services can be ported, migrated, deployed, and run smoothly across the devices. For mobile device manufacturers, the openness determines how much they can customize the mobile OS itself to run across platforms and differentiate their devices from others, and even more importantly, the openness determines how easily they can build devices with a consistent user experience. For ISVs and community developers, the openness determines how easily they can develop new applications with their creative ideas and how their investment on application development can be maximized through programming once running across devices. For the end users or consumers of mobile devices, the openness means how easily they can get more applications, like the rich applications downloaded from the stores, without worrying too much about inconsistency of user experience and incompatibility of applications across devices. The openness may also give people the chance to participate in the development and evolution of the mobile OS itself during the life cycle of the mobile OS.

“The openness determines how easily they can build devices with a consistent user experience.”

Evolution of Mobile OS Openness

A couple of years ago, most mobile devices were feature phones, and people mostly used the phones for voice calls, as a phonebook, and for text messages. For consumers, the applications they could play were limited to the applications built in devices when the devices were shipped out from the factory. For application developers or third-party ISVs, they didn't have access to any level of source code without a contract with the OS owner. Such a mobile OS was a purely closed system and was typically owned by a mobile device maker. For operators, they had to work closely with OS developers to enable their services, because only the people who developed the mobile OS knew how to develop applications.

“Anyone could have the chance to view all source code, contribute to the code, evolve, and customize the mobile OS itself.”

“Mobile OS openness is a requirement of computing continuum.”

Later, as the mobile phone started to transition from feature phone to smart phone, people expected the smart phone to be able to do more, like browsing the Web and playing music/video, rather than just making calls, storing contact information, and sending text messages. To encourage more developers to develop more applications to meet people's needs, the creators of mobile operating systems simply provided sets of APIs and related tools like SDKs, so that people could develop all kinds of applications for mobile devices. With such openness, application developers gained the freedom to develop applications for mobile operating systems, so it became possible for consumers to buy and install more applications rather than being limited to the pre-built applications. Because application developers and consumers benefitted from such openness, it became almost a “must-have” for most mobile operating systems to provide a set of APIs and an SDK. The iOS from Apple is one of the great examples of a mobile OS providing such a level of openness. In recent years more and more mobile operating systems have made all source code public, in addition to providing APIs and SDKs. Anyone could have the chance to view all source code, contribute to the code, evolve, and customize the mobile OS itself. Compared with the level of openness of just providing APIs and SDK, the open-source mobile OS can provide some additional freedom. For mobile device makers, they may have the freedom to build their own mobile OS based on the open source OS to run on their platforms across devices. For operators, they can easily build and deploy their services across devices running the open source OS and its variants. For developers, the open source mobile OS provides everything they need to easily build their applications. Eventually the end users of mobile devices can benefit from this level of openness, as they have more choices of applications and more devices to choose to run the applications. Lastly, everyone has the freedom to participate in evolving and shaping the open source mobile OS, which is very attractive to the talents from communities around the world. Android OS is another great example of an open source mobile OS. Its great success and segment share growth in the smart phone market during the past few years have indicated to the industry just how successful it has been and how fast it has been growing as an open source mobile OS.

As a summary, the openness of future mobile operating systems is one of the key factors to make mobile platforms friendly to the mobile ecosystem, especially to be attractive to application developers and consumers. Mobile OS openness is a requirement of computing continuum, which expects most software to be built once and running everywhere with a consistent user experience to end users.

Cloud Readiness

The cloud has been widely used by mobile users and most of the cloud services are presented as Web sites and accessed by the browser running on the mobile browsers. More and more cloud services have been provided through web applications, which are installed from an application store and run like native applications on the mobile client. Either with a browser or standalone web application, the following areas should be considered in mobile OS design.

HTML5 Capability

HTML5 capability is essential for web applications to integrate cloud services well and to provide a good user experience.

Web engines being used: we listed the web layout engines and JavaScript engines being used on the mainstream mobile browsers. Chrome* has the potential to be the default browser of Android and bring its success from desktop to mobile. The Webkit is used in iOS and Android.

The HTML5 test web site^[9] scores HTML5 support of browsers on various mobile devices. We can see in Table 2 that iOS, Android, and Windows Phone are all improving their browser's capability to support HTML5. Google made Chrome work on Android 4.0 and showed its ambition to have the lead browser in mobile operating systems. Tizen, the new participant in the mobile OS campaign even got the highest score on its development device released in the first Tizen Developer Conference. We can easily see the intense race of HTML5 support between mobile operating systems.

	Layout engine	JavaScript engine	Score + Bonus
Safari on iOS 5.1	Webkit	Squirrel Fish Extreme (SFX)	324 + 9
Android browser on Android 4.0	Webkit	V8	273 + 3
IE on Windows Phone 8	Trident	Chakra	300 + 6
Tizen 1.0	Webkit	SFX	400 + 15 *
Chrome Beta on Android 4.0	Webkit	V8	364 + 10
Opera Mobile 12	Presto	Carakan	369 + 11
Firefox Mobile 10	Gecko	SpiderMonkey	325 + 9

Table 2: Web Layout Engines, JavaScript Engines, and Their Score from html5test.com

(Source: [9] accessed on May 1, 2012. * The score of Tizen 1.0 is from the latest Tizen development device, tested by Intel Corporation.)

The third-party browsers are in difficult situation and need a strategy to have their own host mobile OS. Opera and Firefox are in such a situation. Due to the fact that they have less control of mobile OS development, they won't be able to win easily if the built-in browser is capable enough for HTML5 support. Firefox has been looking for Boot To Gecko as its host mobile OS. A video on YouTube^[10] also shows the preview of Opera OS on Asus EeePC .

The mobile OS vendors view HTML5 support as more and more important and are making it a core competency. The browser vendors are also looking for the possibility to make it default in the mobile OS.

“HTML5 capability is essential for web applications to integrate cloud services.”

“The web application can be installed and run even offline on the devices.”

Web Applications

The web application defines the client side applications developed with web technologies. It provides rich features by providing APIs for client side development. The web application can be installed and run even offline on the devices. The web application can access local devices and resources as a native application and can be sold in an application store, which benefits much from the cloud service delivery and billing.

The web application is more than a URL accessed from a browser. The related capabilities are being defined in several working groups in W3C. The Web Applications Working Group is the central place related to those works.^[11]

To enable web applications, the mobile OS needs to provide a web application platform, which includes web runtime, web framework, and development tools:

- *The web runtime provides the core capability to run a web application.* It is derived from the browser but is more integrated with the OS runtime. The web runtime provides an HTML5 engine and the APIs to access local devices. It also provides the capabilities to integrate well with the OS runtime, which includes the application life cycle management (install/update/uninstall and launch), OS integration (desktop integration, security policy, OS services access.)
- *The web framework provides rich libraries for web application development.* Examples are jQueryMobile and Sencha. Such frameworks are widely used in iOS and Android.
- *Development tools need to be flexible.* The development tools are very diverse and can be chosen according to the web application developer's preferences. Development tools can be a very concentrated SDK suite like the recently released Tizen SDK 1.0,^[12] browser-based like appMobi XDK^[13], or just a set of tools like RIB and Web Simulator, released on 01.org.^{[14][15]}

As a trend, the mobile OS has to provide a capable and high performance web runtime, a rich web framework, and flexible development tools.

Security is always an important topic for cloud computing. For HTML5 cloud integration in mobile operating systems, the following features must be present:

- *Sandbox support in web runtime.* The web applications should run in separate processes and be managed by the web runtime. Sandbox in browser or web runtime has been supported on most modern mobile operating systems.
- *JavaScript code protection.* JavaScript is a scripting language, so the best way to protect the code is still to run the code on the server side.

“In reality, different mobile operating systems provide different HTML5 support.”

Cross-Platform Capability

HTML5 is well known for its cross-platform capability. But in reality, different mobile operating systems provide different HTML5 support, and the

standardization of HTML5 is still ongoing. PhoneGap has been developed to address the cross-platform HTML5 support by providing its own device APIs. Apple iOS, Android, and Windows Phone all have supported PhoneGap. It is evolving and following W3C. Other Web API providers are also trying to make them into the HTML5 standard in W3C.

The trend is to have a unified HTML5 standard but that is not easy. Mobile OS vendors will implement their ideas in their own way before they go to standard. Apple, Google, and Microsoft are all active participants in the W3C standard definition. For other mobile OS vendors, either following W3C or joining the definition is the trend to make their mobile operating systems survive.

Performance

The performance is complained about most by mobile application developers when they start to build applications with HTML5. The optimization for mobile devices is the most important work to do for HTML5 to really succeed in the mobile area. We consider the following to be the most important areas to do work in optimization for mobile operating systems:

- *Hardware acceleration.* The graphics and video should be accelerated by hardware. WebGL has been enabled on more and more platforms.
- *Multithreading support.* Web Worker should be supported as a key feature in HTML5.
- *JavaScript engine optimization.* JIT (Just In Time) has been enabled in both SFX and V8.
- *Native or hybrid application support.* The capability to reuse existing native libraries will be another approach for web applications. Android NDK provides such a capability and it is widely used in Android applications. But on web applications, it has not been widely used. The NaCL by Chromium is an option to support that.

Cloud Integration

Besides the powerful capabilities provided by HTML5, the seamless integration between the cloud and client is even more important. It is not only for web applications but also for native applications. Important reasons for integration include:

- *Cloud storage seamless integration.* The client should be using the cloud storage just as it uses its native storage. That requires the cloud storage client to be tightly integrated into the mobile OS. Apple has made iCloud deeply integrated into iOS 5. The Google Drive is integrated with Google web services.
- *Cloud APIs accessibility.* On the cloud client side, the mobile operating systems should provide capable and easy-to-use libraries for both web and native applications to access cloud client APIs, which are normally RESTful, SOAP, or Query APIs.

“The optimization for mobile devices is the most important work to do for HTML5.”

“Apple has made iCloud deeply integrated into iOS 5.”

“The synchronization and notification are key features in mobile operating systems.”

- *Account management.* With cloud integration, multiple clients with a single account share the cloud and the synchronization should be managed. Accounts should be tightly integrated in the mobile OS and account APIs should be provided for applications to securely access the corresponding cloud services and the local resources. The synchronization and notification are key features in mobile operating systems to enlarge the usability of cloud integration.

Discussion and Summary

In this section, we first discuss the major mobile operating systems in the market today, and then summarize this article.

Apple iOS

Apple has been the leader in mobile OS design. Its iPhone* and iPad* have prevailed across the world in only a few years. Both products feature the Apple iOS.

User experience: iOS provides good performance and is normally set as the benchmark for other mobile operating systems. Apple is continuously enhancing the UX performance. The iPhone 4S has much better performance boost than its previous generations, especially the Internet and browser.^[1] With more new features added, it adds more performance requirements. An unofficial study^[1] showed the UX performance drops after the upgrade from iOS 4.x to 5.x on iPhone 3GS.

Power management: iOS power optimization seems not able to catch up with the increasing demands on power for new features. Arieso, a mobile network management company, estimates that iPhone 4S users consume twice as much data as the previous iPhone model due to increasing use of online services like the virtual personal assistant Siri, which definitely consumes much more power.^[3]

Openness: iOS is perceived as a closed mobile OS. Research work^[6] defines a concept of perceived platform openness (PPO), where a platform's openness degree is decided by its developers' perception.

Cloud readiness: iOS 5.0 with HTML5 support makes it a good cloud client and the iCloud has been integrated by default as storage.

Google Android

Android is currently a popular operating system for mobile devices and is developed by the Open Handset Alliance led by Google. The goal of the Android Open Source Project is to create a successful real-world product that improves the mobile experience for end users.^[16]

User experience: The Android user experience team defined a set of design principles^[17] with three overarching goals: “Enchant me,” “Simplify my Life,” and “Make me amazing.”^[18] State-of-the-art Android and iOS devices achieved similar results in a set of battery life benchmark tests.^[19]

Power management: Android aggressively suspends devices to save power whenever nothing blocks suspend by holding a *wake lock*.^[20] However, Android allows third-party applications to run in the background, which might hold such a *wake lock* for no good reason and thus suck power quietly.

Security: Each application runs in a sandbox environment to enforce security in Android and this is done by assigning each application a unique user ID and running that application as that user in a separate process.^[21]

Openness: Google releases the Android code as open source under Apache license and the Android Open Source Project is where Android development and maintenance happen. However, Google typically partners with a selected manufacturer to make a flagship device for each new version of Android and only makes the new code publicly available after that device has been released. Fragmentation has become more and more a big concern in the Android ecosystem. Android maintains the Android compatibility program and offers the compatibility test suites to guarantee applications developed for Android run on every Android device.

Cloud readiness: Although Google has a huge lead in the cloud area, it has not put together a comprehensive solution as Apple does with iCloud yet.

Microsoft Windows Phone

Microsoft has released its latest redesigned mobile OS called Windows Phone. Based on their design change between Windows Mobile 6.5 and Windows Phone 7, some characteristics of the newer OS are exposed.

User experience: With touchscreen-based user interaction replacing the previous stylus input, Microsoft decided to break the application compatibility between Windows Phone and Windows Mobile.^[23] Similar to Android's AppWidget design, Windows Phone invents the concept of *Live Tiles* for the home screen.^[28]

Power management: Similar to its design security, Windows Phone's design for battery life can largely benefit from its Windows CE and Windows Mobile experience. One special consideration is that Windows Phone chooses black as the main default color theme, because black pixels do not emit any light, hence saving power for the OLED screen.^[32]

Security: Windows Phone's design is shifted from the original Windows Mobile's enterprise-oriented design to an end-user-oriented one. The security experience accumulated for the enterprise product should be still useful.^[31]

Openness: Before Windows Phone was available in the market, Microsoft released its SDK to enable the developers to program for the new OS.^[24] Windows Phone Marketplace has provided its services to 35 different countries/regions.^[25] The current programming languages are C# and Visual Basic. These are not a surprise to any Windows developers, so the language learning curve is expected to be flat.

“Android aggressively suspends devices to save power.”

“Windows Phone Marketplace has provided its services to 35 different countries/regions.”

Cloud readiness: Windows Phone is approaching cloud readiness at a fast pace. Windows Phone 8 integrates Internet Explorer 10 that is claimed to have full HTML5 support and supports parallel page loading in multiple tabs.^[29] Besides that, Skype is deeply integrated into the OS.^[26] One new concept in Windows Phone are *hubs*, which aggregate various similar service features into one hub. This is supposed to greatly improve the phone's user experience with cloud services.^[27] Furthermore, the software framework design of Windows Phone includes two parts: Screen and Cloud. The Cloud part is especially designed for "Developer Portal Services" and "Cloud Service."

In this article, we have investigated the major aspects of mobile OS design based on the analysis model we have developed, including user experience, battery life, cloud readiness, security, and openness. These should be the areas of focus for next-generation mobile OS design.

The future mobile OS also depends on the available hardware design. We believe a successful mobile system is a result of co-design between software and hardware, together with the progress of the Internet.

References

(The paper references the web sites accessed on May 1, 2012.)

- [1] [http://www.youtube.com/watch?v=ng33wXDkyRM,](http://www.youtube.com/watch?v=ng33wXDkyRM)
[http://www.iphonedownloadbolg.com.](http://www.iphonedownloadbolg.com)
- [2] <http://www.anandtech.com/show/4951/iphone-4s-preliminary-benchmarks-800mhz-a5-slightly-slower-gpu-than-ipad-2>
- [3] <http://www.arieso.com/news-article.html?id=89>
- [4] http://www.apple.com/iphone/business/docs/iOS_Security_Mar12.pdf
- [5] <http://www.techrepublic.com/blog/security/comparing-android-and-ios-security-how-they-rate/5774>
- [6] <http://www.user.tu-berlin.de/komm/CD/paper/090322.pdf>
- [7] <http://www.apple.com/icloud/>
- [8] <http://www.networkworld.com/community/blog/apples-ios-5-and-cloud>
- [9] <http://html5test.com/results/mobile.html>
- [10] <http://www.youtube.com/watch?v=mWSPNDD0tek>
- [11] <http://www.w3.org/2012/webapps/charter/Overview.html>
- [12] <https://developer.tizen.org/sdk>
- [13] <https://chrome.google.com/webstore/detail/onmkoldigcfmebcinpmineoadckallb>
- [14] <https://01.org/projects/rapid-interface-builder-rib>

- [15] <https://01.org/projects/web-simulator>
- [16] <https://www.mylookout.com/mobile-threat-report>
- [17] <http://www.juniper.net/us/en/local/pdf/additional-resources/jnpr-2011-mobile-threats-report.pdf>
- [18] <http://webupon.com/security/information-systems-attacker-motivations/>
- [19] <http://source.android.com/tech/security/index.html>
- [20] <http://www.w3.org/2011/webappsec/>
- [21] <http://www.trustedcomputinggroup.org>
- [22] <http://www.globalplatform.org/specifications.asp>
- [23] <http://channel9.msdn.com/Events/TechEd/NorthAmerica/2010/WPH201>
- [24] <http://www.engadget.com/2010/09/16/microsoft-demos-twitter-and-netflix-apps-for-windows-phone-7-r/>
- [25] http://windowsteamblog.com/windows_phone/b/windowsphone/archive/2011/07/06/windows-phone-around-the-world-language-support-in-mango.aspx
- [26] <http://wmpoweruser.com/microsoft-reps-claiming-windows-phone-8-definitely-coming-to-second-gen-handsets-probably-to-first-gen/>
- [27] <http://www.engadget.com/2010/03/18/windows-phone-7-series-the-complete-guide/>
- [28] <http://www.engadget.com/2010/02/15/windows-phone-7-is-official-and-microsoft-is-playing-to/>
- [29] <http://pocketnow.com/windows-phone/exclusive-windows-phone-7-web-browser-comparison>
- [30] <http://arstechnica.com/microsoft/news/2011/04/windows-phone-7-mango-one-heck-of-an-upgrade.ars>
- [31] <http://arstechnica.com/microsoft/news/2010/03/windows-phone-7-series-in-the-enterprise-not-all-good-news.ars>
- [32] <http://www.neowin.net/news/interview-windows-phone-7-battery-life-copy-paste-multitasking-and-more>
- [33] <http://software.intel.com/en-us/articles/aws-android-workload-suite-for-user-interaction-measurement/>
- [34] <http://software.intel.com/en-us/articles/quantify-and-optimize-the-user-interactions-with-android-devices/>

Author Biographies

Xiao-Feng Li (xiao-feng.li@intel.com): Architect of System Software Optimization Center, Intel Corporation. Xiao-Feng has been working with Intel for 12 years, with extensive technical experience in parallel system, compiler design and runtime technologies, where he has authored about 20 academic papers and 10 U.S. patents. Two years ago, Xiao-Feng initiated the evaluation and optimization efforts for best Android user experience on Intel platforms. Before he joined Intel, Xiao-Feng was a technical manager in Nokia Research Center. Xiao-Feng holds a PhD degree in computer science, and is a member of Apache Software Foundation. His personal homepage can be found at <http://people.apache.org/~xli>.

Yong Wang (yong.y.wang@intel.com): Senior software engineer from Intel's Open Source Technology Center. He has been with Intel for 7 years working on a variety of projects, including virtualization, manageability, OSV enabling, etc. Most recently Yong has been working on power management for a wide range of mobile operating systems such as Moblin, Meego, Tizen and Android. Yong graduated from Beijing University of Aeronautics and Astronautics and holds a master degree of computer science. He enjoys sports and reading in his spare time.

Weihua Jackie Wu (jackie.wu@intel.com): Engineering Manager in Intel's Open Source Technology Center leading a team on Mobile OS and HTML5 tools development. Before that, as a research engineer, Jackie was focused on wireless network and energy efficient communications. Prior to joining Intel in 2004, Jackie was in Chinese Academy of Sciences developing embedded operating system and smartphone products. Jackie received his M.S. and B.S. in Engineering Mechanics from Beijing University of Aeronautics and Astronautics in 2002 and 1999 respectively. He has two US patent applications.

Kerry Jiang (kerry.jiang@intel.com): Software Manager in System Optimization Technology Center (SOTC), Intel Corporation. Kerry has been working with Intel for eight years, four years in open source mobile OS software stack on IA, which includes Android optimization and MeeGo SDK. Before joining Intel, Kerry worked in Motorola on mobile platform and 2G wireless base-station software developments. Kerry holds a master degree in electronics engineering.

Bingwei Liu (bing.wei.liu@intel.com): Engineering manager of Open Source Technology Center, Intel Corporation. Bingwei has been working in Intel for 11 years, with abundant experience in Linux OS, open source software and system engineering. His working scope spans from enterprise to client platforms and now is focused on Mobile OS. Bingwei holds a master degree of computer science.