SOLUTION BRIEF

Data Center: Big Data and Analytics
Security

Apache Spot: A More Effective Approach to Cyber Security

Apache Spot uses machine learning over billions of network events to discover and speed remediation of attacks to minutes instead of months—streamlining resources and cutting risk exposure

This solution brief describes how to solve business challenges and enable digital transformation through investment in innovative technologies.

If you are responsible for ...

- **Business strategy:** You will better understand how a cyber security solution will enable you to meet your business outcomes.
- **Technology decisions:** You will learn how a cyber security solution works to deliver IT and business value.

"Apache Spot opens an important new path for organizations to get past challenges at scale as well as provide a framework for analytics teams to overcome obstacles and create effective value-add analytics."[2]

— Austin Leahy
Principal Data Scientist,
Global Threat Management, eBay

Authors

Jeff Towle
Security Solution Architect Manager
Industry Sales Group, Intel

Ritu Kama
Director, Big Data Products, Intel

Apache Spot

Executive Summary

No industry is immune to cybercrime. The global "hard" cost of cybercrime has climbed to USD 450 billion annually; the median cost of a cyberattack has grown by approximately 200 percent over the last five years,[1] and the reputational damage can be even greater. Intel is collaborating with cyber security industry leaders such as Accenture, eBay, Cloudwick, Dell, HP, Cloudera, McAfee, and many others to develop big data solutions with advanced machine learning to transform how security threats are discovered, analyzed, and corrected.

Apache Spot is a powerful data aggregation tool that departs from deterministic, signature-based threat detection methods by collecting disparate data sources into a data lake where self-learning algorithms use pattern matching from machine learning and streaming analytics to find patterns of outlier network behavior. Companies use Apache Spot to analyze billions of events per day—an order of magnitude greater than legacy security information and event management (SIEM) systems.

Apache Spot's big data design can be used for securing network, application, and the Internet of Things (IoT) cyber security use cases. Apache Spot is the industry's first open source predictive threat deterrence solution. It takes advantage of a wide range of existing and emerging technologies created by Intel for a truly end-to-end cyber security solution. Apache Spot mitigates risk while serving as a powerful competitive advantage to protect classified data and proprietary intellectual property.

More Effective Approach to Cyber Security

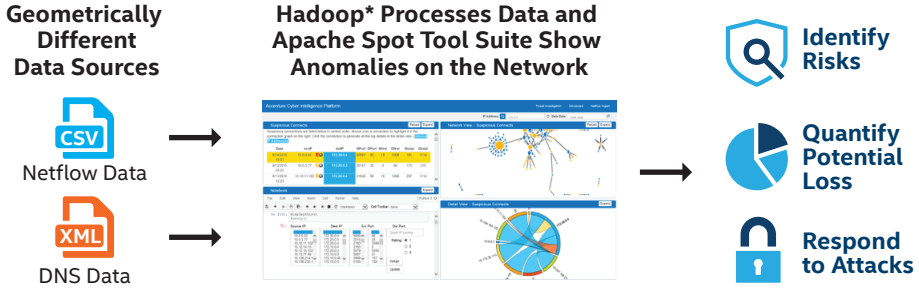

Figure 1. Apache Spot innovates cyber security through machine learning and advanced data science.

Solution Benefits

- Efficient, intelligent analysis of risky network activity
- Identification of unknown cyber threats in near real time
- Better-informed actionable intelligence
- Integrated, open source operating model
- Massively scalable, extensible architecture
- Modular design, fully customizable

Business Challenge: Risk Mitigation Demands Faster Discovery, Actionable Answers

In 2015, nine mega-breaches (each involving more than 10 million records) were reported and the total number of exposed identities increased 23 percent to 429 million compared to the previous year.[3] As of January 2016, Target has incurred USD 291 million in breach-related costs including legal fees, crisis communications, and forensics costs.[4] Virtually every industry is susceptible to cybercrime (see Figure 2).

More Data = More Threats

Increasingly, complex data schemas and the volume and velocity of data has made traditional deterministic approaches to intrusion detection unrealistic. Current approaches to cyber security have not kept up with the expanding threat vectors driven by mobility, social media, and virtualized cloud networks. Legacy security tools cannot handle the huge array of disparate data sources needed to understand what constitutes a risk, quickly determine where those risks are, and determine the appropriate actions needed to remediate or investigate them.

Limited Capabilities

The cyber security industry requires powerful new approaches to quantify and make cyber security risk events actionable. Security analysts must have solutions that create intelligent metrics while eliminating false positives, benchmarked against a known good, baseline security posture. Dwell time data breach detection is 98 days on average.[5] This gap is a critical risk window that must be closed as soon as possible. Each hour can significantly reduce economic, reputational, IP, and classified data loss. Siloed cyber security tools are expensive and lack the innovation, insight, and agility needed to find unknown, catastrophic risks. The volume of heterogeneous data sources and the velocity at which these sources need to be analyzed can only be addressed by massively scalable data lake intelligence and self-learning algorithms that speed identification of what constitutes a credible risky event that requires action.

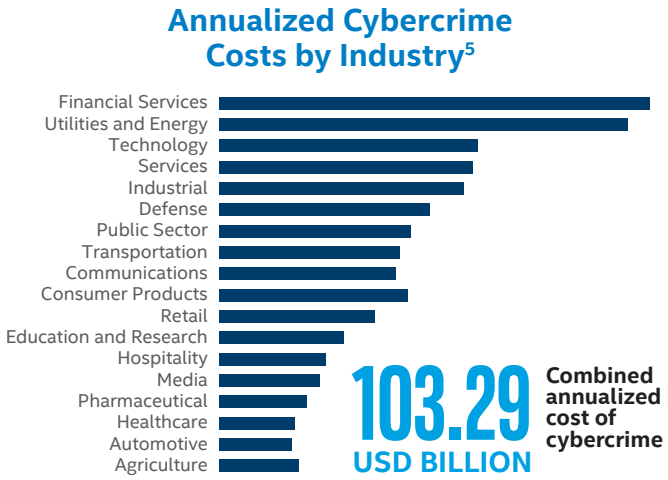

Figure 2. The cost of cybercrime affects every industry and can reach staggering levels.

"You can no longer have a backward-looking signature-based approach for cyber security."[6]

— Tom Reilly, CEO, Cloudera

A Broad Brush: Any Enterprise Can Improve Threat Detection and Risk Mitigation

Apache Spot advanced threat detection is an innovative and industry-tested approach to cyber security. It secures an organization's critical assets against threats through the following capabilities:

- Finding new or unknown threats
- Visualizing fraud detection by creating exception-based baselines
- Identifying compromised credentials
- Detecting policy violations
- Enhancing compliance monitoring and management
- Detecting advanced persistent threats (APTs) and malware from external sources
- Providing visibility into network and endpoint behaviors
- Revealing malicious behavior patterns of zero day threats that are undocumented or for which no existing evidence (signature) is available

Solution Value: Rapid Insight into Suspicious Connection Activity

Cyber security is ultimately a business risk presented through technical threat vectors. Until the Snowden and Target corporate hacks, businesses defined their growth strategies without considering the effectiveness of cyber security controls. Now, without a cyber security reference architecture that is designed to help meet business goals, reputation and business growth objectives are at risk. In fact, the real financial risk of immature cyber security controls has been demonstrated recently by credit rating agencies such as Standard and Poor and Moody's inclusion of cyber security in their credit analyses.

"Moody's views material cyber threats in a similar vein as other extraordinary event risks, such as a natural disaster, with any subsequent credit impact depending on the duration and severity of the event, the implications could start taking a higher priority in credit analysis."[7]

— Moody's Investors Services

ONE SOLUTION FOR ALL ANALYST LEVELS

Senior Analyst: "Allows me to customize the data and scripts to my environment."

Junior Analyst: "Automatically alerts me to actionable suspicious events, helps me investigate, and suggests ways to optimize my network in a timely fashion.

Decision Maker: "Tells me what happened in a way I can understand so I can make decisions as a result."

Protect the Data

A reasonable and appropriate security posture is especially important during digital business transformation. As companies adopt more advanced and agile business models built on software-defined infrastructure (SDI), cloud-based services, and the Internet of Things (IoT), they must capture logging and metadata events from these data sources to perform analytics and risk analysis in real time. Apache Spot is well positioned to secure any organization's risk posture in these modern, complex environments.

Narrow the Threat Quickly

When combined with built-in security technologies at the processor, storage, networking, cloud, and sensor levels, Apache Spot acts as the foundation for a massively scalable and highly adaptive end-to-end cyber security solution. Currently, Apache Spot collects network flow data and domain name system (DNS) log event information.[8] It then aggregates the data and reports suspicious connections and outlier intrusion patterns. The correlated data is placed in an Apache Hadoop* cluster on the Cloudera Enterprise Data Hub*. Apache Spot then uses operational analytics tools and machine learning algorithms developed by Intel to detect, score, and take action on the highly suspicious connection activity (see Figure 3). A continuous refinement process is used to build risk profiles to assist security administrators with investigations and forensic research.

Going far beyond the traditional signature-based, deterministic cyber security approach, Apache Spot can operate as a batch or near-real-time cyber security analytical platform that deters and controls emerging threat vectors. Companies using Apache Spot, such as eBay, can expect the following benefits:

- Detect unknown threats at multiple levels in near real time
- Create an accurate risk profile baseline for contextual awareness to perceived threats
- Reduce "false positives" by using the risk baseline to train algorithms, eliminating repeat consideration
- Analyze billions of events to easily—and affordably—scale as needed
- Automatically learn what is "normal" for a particular network and then continuously improve
- Provide interactive visualizations that help quickly identify suspicious patterns
- Integrate with other analytic processes or platforms for additional use cases

Grow the Technology

As an open source initiative from Intel, Apache Spot benefits from continuous improvements from the open source community. Innovation and agility are strengthened through industry collaboration and knowledge sharing. Apache Spot integrates with other cyber security tools and threat management libraries, which helps maximize the value of legacy cyber security investments and extends them with a more valuable forensic pattern analysis.

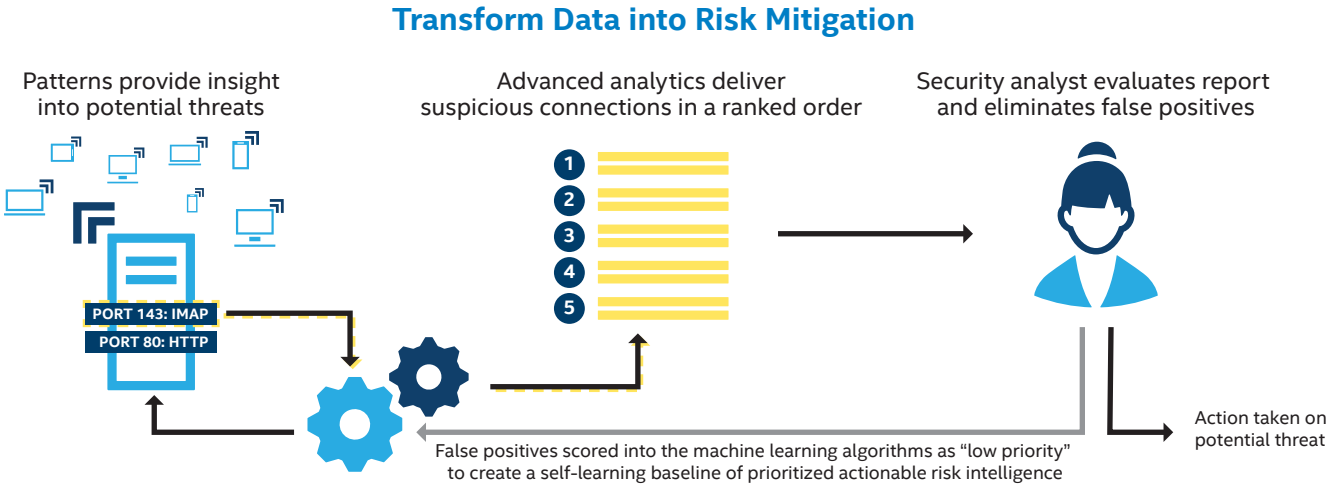

Figure 3. Apache Spot helps companies automate cyber security life cycle management to align highly trained cyber resources to the suspicious events that require immediate action.

Solution Architecture: Near-Real-Time or Batch Security Discovery

Apache Spot uses an open source library and runs on Cloudera Enterprise Data Hub. Intel and Cloudera have tuned the hardware and software solution to match the huge volumes of data that need to be processed to identify security threats either in batch or in near real time. Apache Spot components can be deployed on bare-metal systems or on public clouds, and Apache Spot can be set up on an appliance for rapid prototyping in any data center. Intel has established a robust industry ecosystem with organizations such as Accenture to help deploy Apache Spot in cloud environments. Additionally, an Apache Spot deployment configuration can be adjusted to reflect the necessary data volumes and analytic processing windows for the use cases being addressed.

Apache Spot's power lies in the ability to process data from widely disparate sources, correlate the data to identify anomalies, and deliver actionable business intelligence. Instead of relying on knowledge of known threats, Apache Spot can process years of non-deterministic data to uncover hidden indicators of compromise to enrich event sampling methodologies and speed investigations. Further machine learning and visual analytics can find hidden, deep-level patterns of risky or outlier behavior.

As shown in Figure 4, Apache Spot runs on an infrastructure consisting of bare-metal Intel® Xeon® processor-based servers, the Linux* OS, and a Java*-based virtual machine. The primary components of Apache Spot include the following:

- **Parallel data ingestion framework.** The system runs on Cloudera Enterprise Data Hub and uses optimized, open source decoders to load network flow and DNS data into a Hadoop Distributed File System* (HDFS). The decoded data is stored in multiple searchable formats.
- **Machine learning.** The system uses a combination of Apache Spark* and optimized C code to run scalable machine learning algorithms. The machine learning component filters suspicious traffic from normal traffic and characterizes the unique behavior of an organization's network traffic for optimization and other network analysis use cases.
- **Operational analytics.** Other techniques applied to the data include noise filtering, whitelisting, and heuristics, which produce a short list of the most likely patterns that may be security threats. This helps reduce false positives and non-threatening indicators to make Apache Spot's results actionable as quickly as possible.
- **Interactive visualization dashboard.** Cyber security analysts can drill down to explore Apache Spot's results and initiate appropriate responses.

Rounding out the solution are a set of unified services for orchestration; operations; and resource, security, and data management.

Technology Innovation for Fast, Scalable Cyber Security

- Intel® architecture provides an affordable, massively scalable foundation for cyber security analytics.
- Cloudera Enterprise Data Hub* is a unified platform that can cost-effectively collect and store huge quantities of data and provide actionable insight to a diverse range of users.
- Open source tools such as Apache Spark* and Kudu* enhance Apache Spot's big data processing capabilities.
- Intel® Math Kernel Library and Intel® MPI Library speed results through direct processor access.
- Intel's Solution Integrator network extends industry expertise to help with implementation details as well as the analysis, interpretation, and response to cyber security alerts.

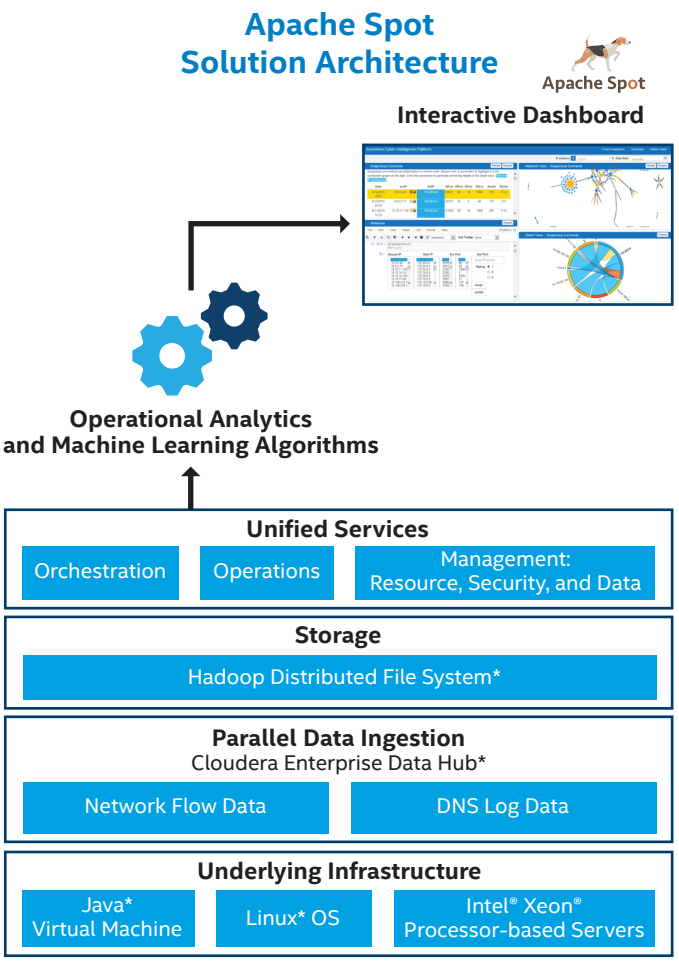

Figure 4. After ingesting disparate sources of data, Apache Spot uses parallel processing, machine learning, advanced analytics, and an intuitive interactive dashboard to help enterprises focus on the riskiest threats.

Conclusion

Increasingly, large amounts of confidential and classified information are used in modern data-sharing environments. Legacy signature and rules-based cyber security controls are no longer sufficient to guard against advanced threats. Companies must proactively identify threats that are not currently known or understood with flexible, intelligent, and massively scalable security intelligence tools.

Compared to traditional signature-based cyber security solutions, Apache Spot accelerates the exposure of suspicious connections and previously unseen attacks. Its innovative approach to cyber security provides massively scalable analytics in batch mode or near real time, helping to significantly reduce false positives with self-learning algorithms. The data is filtered and analyzed to identify suspicious data flow patterns that do not match normal business-based connection criteria from a known good risk profile that is unique to each enterprise. Companies such as eBay are using Apache Spot to create a predictive threat posture to guard against malicious attacks before catastrophic damage can be done.

Apache Spot helps future-proof an organization's cyber security because it is based on an open design built to ingest disparate heterogeneous data sources into a Hadoop platform. Companies like Cloudera, Accenture, Dell, HP, and McAfee are already committed to taking Apache Spot to the next level. Intel encourages more companies to contribute to the Apache Spot project and help accelerate business value through ingestion of additional data sources and the development of even smarter algorithms.

Find the solution that's right for your organization. Contact your Intel representative or visit **intel.com/analytics.**

Solutions Proven By Your Peers

Intel Solution Architects are technology experts who work with the world's largest and most successful companies to design business solutions that solve pressing business challenges. These solutions are based on real-world experience gathered from customers who have successfully tested, piloted, and/or deployed these solutions in specific business use cases. Solution architects and technology experts for this solution brief are listed on the front cover.

Learn More

You may also find the following resources useful:

- **Apache Spot:** spot.apache.org
- **Cloudera Enterprise Data Hub:** cloudera.com/content/dam/cloudera/Resources/PDF/solution-briefs/Cloudera-EDH-ExecutiveBrief.pdf

Solution Provided By:

Apache Spot

intel

[1] Source: hamiltonplacestrategies.com/news/cybercrime-costs-more-you-think#Paper

[2] Source: vision.cloudera.com/open-network-insight-changing-infosec-data-science-forever

[3] Source: symantec.com/content/dam/symantec/docs/reports/istr-21-2016-en.pdf

[4] Source: datasecuritylaw.com/blog/targets-cyber-insurance-a-100-million-policy-vs-300-million-so-far-in-costs

[5] Ponemon Institute, 2015 Cost of Cyber Crime Study: United States, October 2015. www.ponemon.org/blog/2015-cost-of-cyber-crime-united-states

[6] Source: datanami.com/2016/05/09/oni-may-best-hope-cyber-security-now

[7] Source: bankinfosecurity.com/moodys-warns-cyber-risks-could-impact-credit-ratings-a-8702

[8] Because Apache Spot is an open source project, it benefits from the input from many cyber security experts. It is anticipated that future releases of Apache Spot will include additional data sources beyond network flow and DNS data.

1016/JGAL/KC/PDF ♻ Please Recycle 334459-001US