WhiteHat SECURITY

A Big Data Case Study on Using a Risk-Based Approach for Information Security and Fraud Analytics

bluelava CONSULTING, LLC

Gaining visibility, meaningful information security, and fraud data in seconds

This document presents a big data case study based on work performed between 2009 and 2013 when the founder of Blue Lava Consulting was employed as the CISO for a Global Fortune 100 organization. The materials presented in this case study are not proprietary and they aligned to the business requirements used by the organization (referred to hereafter as "the Company") during this time period.

Blue Lava partnered with WhiteHat Security to create this case study. WhiteHat Security's SaaS-based Sentinel was the Company's application security testing solution of choice and an integral part of it's Information Security and Fraud program.

Special thanks to the InfoSec and IT Operations teams who built this InfoSec and Fraud big data solution and the former boss for constantly challenging us to be the best we could. Without his support and belief in the teams working on this project, this effort would not have been so successful.

A gentle reminder: a successful Information Security and Fraud program continues to evolve and mature as long as its leadership team nurtures and supports the program and the culture that's required to make it succeed.

WhiteHat SECURITY

Contents

Introduction

Organizations are rich with sensitive data—intellectual property, financial reports, personally identifiable information (PII), credit card data, trade secrets, client data—the list goes on and on. The chances any given organization will become a cybercriminal's target are now greater than ever.

Cybercrime continues to be more and more rampant as cyber criminals are targeting companies and sensitive data at an alarming rate.

Companies must evolve to use an integrated risk-based approach with big data for Information Security (InfoSec) and Fraud analytics as part of an overall strategy that compliments the overall InfoSec program. The decision of how a big data solution for InfoSec and Fraud analytics will be used in an environment will rely on a number of factors. For this case study, the big data for InfoSec and Fraud analytics complemented the overall InfoSec program the team built from the ground up and was treated as a separate effort, requiring additional funding. The investment was successful.

The topics addressed in this case study will provide detailed information regarding the use of big data for InfoSec and Fraud analytics. The information contained in this document can be applied to a large enterprise, or the approach can be scaled back to support a smaller organization.

This case study provides five key insights:

1. How real-time InfoSec and Fraud analytics can protect an organization
2. Building and deploying a risk-based InfoSec and Fraud analytics solution (referred to hereafter as the "Big Data Solution" (BDS))
3. Building the budget for supporting a BDS
4. The total cost of ownership (TCO) of implementation and support
5. Building and deploying a BDS: the proactive approach

This case study does not cover pricing for professional services, hardware, and software.

Executive Summary

In 2009, the Company began re-innovating its InfoSec program. During this time, the Company's InfoSec team started to collect data from several traditional security systems while conducting penetration (pen) tests and vulnerability assessments across multiple websites and supporting systems.

Security events were increasing, which caused the InfoSec team to spend a significant amount of time researching root causes.

Worse yet, after conducting multiple research efforts in response to anomalous website events, it was discovered that cyber criminals were bypassing traditional InfoSec and Fraud monitoring solutions. Although the cyber criminals were bypassing traditional monitoring tools because their behavior looked like legitimate web traffic, a data loss did not occur. How could this be?

Due to the evolution of cyber criminal tactics, it was more difficult to identify the business owner when something malicious occurred. The lines were blurred in determining if the malicious activities were an InfoSec issue or a Fraud issue. By 2010, the InfoSec team treated all malicious InfoSec and Fraud events as one and the same.

Something different had to be done, as handling the growing number of security events introduced a variety of challenges for the Company. The InfoSec team assembled a plan to research options for doing something radical to solve these complex challenges.

This was the genesis of the InfoSec and Fraud BDS, which saved the company millions of dollars through operational efficiencies and by preventing malicious activities from occurring.

Key Findings

1. Cyber criminals are evolving – if you don't understand your environment, cyber criminals will take advantage of weaknesses in systems, data, and controls.
2. If you don't implement a solution as a means to understand all data elements in your environment, you will not be successful in understating how the evolutionary strides made by cyber criminal evolutions are being used to by-pass your Information Security and Fraud Controls.
3. Legacy risk models need to be re-evaluated – new risk models that quantify losses in dollars are available and must be considered.

Recommendations

1. Use big data Information Security analytics to reduce the noise and false positive rates from monitoring systems – this will allow your teams to focus on the most important events.
2. Start small and pick a project where you can see results. For example, working with application security vulnerabilities and understanding how high-risk countries are targeting your systems is critical. By using solutions like WhiteHat Security's Sentinel, your organization will be able to scan and identity Internet-facing applications quickly.
3. Ensure the Big Data Information Security analytics platform is leveraged across multiple teams in order to maximize the investment.
4. Recruit the right resources to ensure you have the right collective skillsets to ensure success with the system.

Company Overview

The Company is an online division of a Fortune 100 Global retailer. The online division has hundreds of websites and a co-branded credit card. The IT environment for this organization is a separate entity focusing on the online component. To provide an integrated experience, all systems are required to have the ability to extend to traditional brick-and-mortar point-of-sale (POS) terminals, inventory, and pricing systems.

The organization's business and technical teams are very forward-thinking and embrace innovation. Other organizational attributes include the following:

1. Is comprised of hundreds of internal developers as well as on-site contractors and third-party developers working in a fast-paced agile environment
2. Has several offices located in the Midwest, West Coast, Israel, and India
3. Operates multiple geographically-dispersed data centers
4. Has an IT Operations team managing all functions of alerting and monitoring for all online entities
5. Is comprised of a variety of business components that require adherence to the following regulatory and compliance-related items:
   a. Payment Card Industry Data Security Standard (PCI DSS)
   b. Health Insurance Portability and Accountability Act (HIPAA)
   c. US and International personally identifiable information (PII) protection laws
   d. Sarbanes-Oxley (SOX) Act

InfoSec Program Overview

In Q3 of 2009, the Company began migrating its data centers as a means to consolidate all IT Operations under centralized, in-house management. As part of this effort, the Company wanted to build an InfoSec and compliance team since these functions were once outsourced to third-party organizations.

Once the new data centers were operational, the InfoSec team initiated programs that started collecting the following data from several sources:

1. Application security scanning from WhiteHat Sentinel Source tor pre-production / static application security testing and WhiteHat Sentinel for production / dynamic application security testing.
2. File integrity monitoring
3. Firewalls
4. Intrusion detection/prevention systems (IDS/IPS)
5. Network firewalls
6. Network vulnerability scanning
7. Operating System (OS) vulnerability scanning

Historically, to determine how cyber criminal tactics were evolving, the InfoSec team relied on distilling trends from previous years' pen test results and the Security Information and Event Management (SIEM) system.

The number of security events were increasing year-over-year. This caused the InfoSec team to spend a significant amount of additional time researching the root causes behind the security events—resulting in less time to focus on being strategic and proactive.

The InfoSec and IT Operations teams manually researched these events by reviewing information in the following manner:

1. Review two dashboards in the IT Operations event console
2. Log into the SIEM to review potential issues
3. Review the log files in supporting systems that were not feeding data into the SIEM (eg. applications, network, OS)

Once this research was completed, additional team members with expertise in performance monitoring, network and systems engineering, and fraud protection also had to be consulted to discuss the findings.

Research requests on average consumed from 5 to 50 analyst hours determining what type of anomalous website traffic behavior was occurring, correlating the behavior to the available data sources, quantifying the potential loss exposure in dollars, and reporting to management for follow-up action. When research at this level was performed, resources working on revenue-generating projects were pulled away from what they were doing, causing friction between the business units and the InfoSec team.

After conducting multiple research projects in response to website events, it was discovered cyber criminal behavior had successfully circumvented the traditional InfoSec and Fraud monitoring tools, going unnoticed for a period of time. Although the cyber criminals were bypassing traditional monitoring tools because their behavior looked like legitimate web traffic, a data loss did not occur. What was even more challenging about the revelation, due to the evolution of cyber criminal activities, it was difficult to identify the business owner when something malicious had occurred. After realizing this type of behavior was malicious in nature, the InfoSec team treated InfoSec and Fraud events and alerts as one and the same.

Observed cyber criminal activities included the following:

1. Site scraping: a malicious technique used to acquire and harvest content and store the data in local databases. For example, a cyber criminal could take advantage of an organization investing in translating its website to multiple languages. The cyber criminal's automatic script runs through the legitimate website and captures the translated language to create a phishing site that looks exactly like the original website. Once the site is scraped, the cyber criminal will then launch a well-crafted phishing attack to lure unassuming consumers to the phishing site and ask them to provide credentials.
2. Slow methodical crawls through the website: this technique is used for automatically checking the website to identify hidden directories, files, or dormant login pages. The risk here is the disclosure of potentially-sensitive information hidden within the website.
3. Architecture probing: entities may scan a website to identify what development framework (eg. J2EE), OS, network, and databases may exist underneath or behind the website. A cyber criminal will attempt to exploit unpatched or vulnerable applications with the goal of gaining unauthorized access to systems, taking control of the systems, or accessing sensitive information.
4. The use of Tor traffic to visit the site: Tor is a network of virtual tunnels that allows an individual to protect their identity and remain anonymous during a visit to a website. By using Tor, it is difficult to differentiate the 'good' traffic from the 'bad' traffic within a computing environment since the Tor user is truly anonymous.
5. Business logic flaws: business logic flaws are unique because cyber criminals don't always exploit application, network, or OS vulnerabilities—in many cases, the advanced cyber criminal will identify a weakness in a business process within an application. For example, when someone visits a web page to register to win a gift card, that person should be allowed to register one time. A cyber criminal will create a script (oftentimes referred to as a 'bot') to register multiple times in order to increase their chances of winning.
6. Anomalous behavior associated with emerging technologies: emerging threats are constantly evolving as cyber criminals are identifying and exploiting weaknesses in new technology.
7. Distributed denial of service (DDoS) attacks: a distributed denial of service is an attack where multiple compromised systems are used to target a single system. The single system under attack is rendered useless, therefore impacting the user experience and revenues generated by the system.

Early Research in 2010

The main driver for the research in 2010 was a theory that existed within the InfoSec team: there were relationships among different types of malicious activities as cyber criminals visited the online websites. For example, if a high-risk country that generated little or no revenue tried to perform a directory traversal on a particular website host server in an attempt to identify the structure of the server, was there a relationship between this activity and the site's application vulnerabilities? Was there also a way to identify or predict this type of malicious activity?

Early research in the area of big data platforms for InfoSec and Fraud in 2010 was weak to non-existent since the market for the InfoSec team's high-level requirements could not be met. Additionally, the InfoSec team could not find a successful implementation of an InfoSec and Fraud BDS.

One thing was certain: reviewing the materials and reports on a daily basis did not demonstrate how serious the number of InfoSec events was until the information was compared on a month-by-month basis. At the end of the year, it was very clear—cyber-criminal activities were increasing and showing no signs of slowing down.

Another Increase in Security Events

The graph below demonstrates the trending of the Company's InfoSec events from 2010-2011. The events were tracked and reported on by the SIEM, which was used to monitor InfoSec events. This 50% increase in security events made it clear that a new course of action was needed to protect the Company's brand, reduce the risk of loss exposure, and protect the supporting infrastructure.

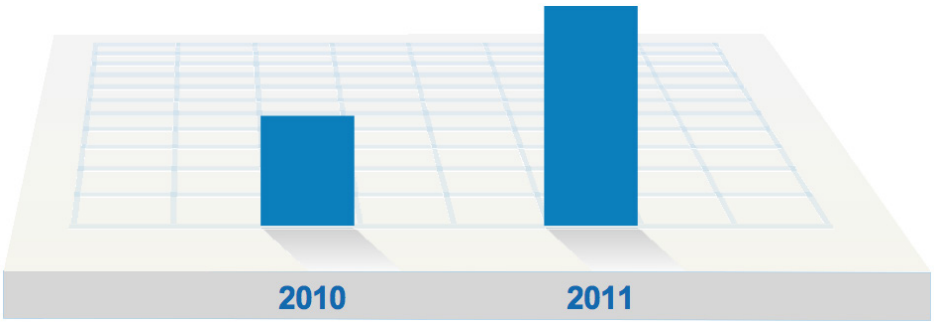


Figure 1. Security Events Per Year

Benefits of Preliminary Research Conducted in 2010

The Company's InfoSec team focused on creating an internal threat landscape dashboard with supporting industry-recognized solutions. By doing this, the InfoSec team clearly identified what to invest in (eg. people, process, and technology) to address cybercriminal activity. This exercise also yielded a clear understanding of how much data was utilized in the environment.

There were some benefits of identifying what the pulse was for a BDS in 2010. One of the results of the preliminary research was solving the complex issue of determining whether website users were legitimate (good) or malicious (bad). This was problematic for a number of reasons.

First, blocking legitimate revenue-generating traffic is always a career-limiting move. Secondly, blocking IP addresses or countries would lead to cyber criminals dropping the original IP address and having another IP address issued within seconds, which allowed the malicious activity to continue.

Initially, when researching 'good' or 'bad' traffic on a website, resources had to be pulled from revenue-generating projects to research a specific event. These research efforts generally took days or even weeks to conduct. By the time the IT Operations team could review the results from an investigation, it was already old news, and the teams were always several steps behind understanding the intent of the website traffic.

The issue of identifying a solution for monitoring and alerting on 'web session intelligence' surfaced through this research in 2010. The InfoSec team made the investment into this technology to reduce the amount of time needed for researching these types of events.

One observation during this time period was that the web session intelligence software was unable to be ingested natively by the SIEM.

Realizing the Need for a Different Solution

With the number of security events continually increasing each year, the Company was forced to review its security, alerting, and SIEM strategy. The amount of data reported from the SIEM solution increased but did not provide the information the InfoSec team needed. Since the collected data was disparate and located in multiple silos, the InfoSec team needed to review data from the SIEM and then manually review the data in other systems. The data needing to be researched was located in several locations belonging to multiple business owners.

Determining the root cause of an event could take hours or days. Many of the teams relied on their own log files or proprietary systems that did not feed the SIEM. The following is a list of teams and supporting systems that could have been required during an investigation:

1. Application Engineering
2. Fraud
3. InfoSec
4. Network Engineering
5. IT Operations
6. Performance Engineering
7. System Engineering

The InfoSec team had been compiling data on a monthly basis, demonstrating that security events were increasing. The results were captured and reported as follows:

1. Daily events
2. Weekly events
3. Monthly events
4. Quarterly events
5. Yearly events

Analysis of the events indicated that cyber criminals had evolved their techniques by creating advanced bots that performed highly-automated functions. Here are two examples of highly-advanced, automated functions that were used in 2011:

1. Generating high-velocity account registrations for a registration page (e.g. over 500,000 registrants within milliseconds).
2. Creating account registrations with email addresses such as name01, name02, name03, and name04—clearly a pattern of malicious behavior.

The graph below demonstrates the Company's increase in InfoSec events from 2010-2012. Security events grew another 50% from 2011-2012.

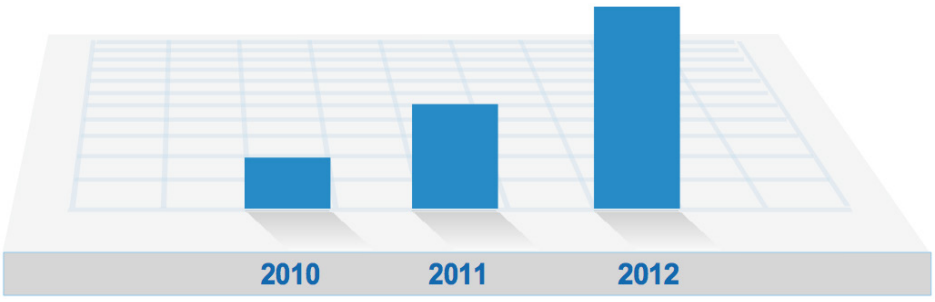


Figure 2. Increase in InfoSec Events

The growing number of security events introduced two new challenges:

1. While the number of security events progressively climbed, the response techniques and processes for investigating security events did not change, causing analysts to use the same amount of time to review each security alert.
2. Cyber-criminal techniques were evolving at a pace that the Company's existing InfoSec and Fraud systems could not identify in near real-time.

Thinking Outside the Box: Evaluating a Big Data Strategy

The InfoSec team wanted to achieve the following high-level objectives by creating a BDS:

1. Use existing technology the teams had already purchased.
2. Determine if a technology existed that could ingest a massive amount of structured and unstructured information with the ability to combine, correlate, index, and search through the data.
3. Preserve the data for a set number of months, which was critical since the data would need to be evaluated for seasonal activities hour-over-hour, day-over-day, week-over-week, month-over-month, and year-over-year.
4. For future technical purchases, new investments were required to deploy an application-programming interface (API) to extract information. This was critical because it would ensure all information could be moved into a single platform. It was imperative to have this open approach to creating the platform since the price of disk space was dipping to a more affordable price per GB. For technology that did not have such APIs, it became part of a discussion with vendors to ensure an API would be included in future releases.
5. Produce meaningful executive and technical dashboards with supporting metrics.
6. Generate actionable alerts in seconds as opposed to hours or days.

During initial discussions, the InfoSec team was unsure how the architecture of the BDS would be configured. For the system to work properly, the BDS had to ingest, correlate and process a large and ever-changing list of different types of events, distributed across two broad categories:

1. Traditional security and fraud events
2. Non-traditional security and fraud events

Traditional InfoSec and Fraud Events

Traditional InfoSec and Fraud event analysis would include gathering event information from the following sources:

1. Cyber threat intelligence monitoring
2. File integrity monitoring (FIM)
3. Firewalls
4. Fraud analytics
5. Geographical location (country code)
6. Geographical location (IP address)
7. Intrusion detection systems (IDS)
8. Intrusion protection systems (IPS)
9. Open source intelligence (OSINT)
10. Web application firewalls (WAFs)

Non-Traditional InfoSec and Fraud Events

Non-traditional InfoSec and Fraud event analysis would involve collecting information from the following sources:

1. Application security vulnerability data
2. Behavioral analytics (web session intelligence software)
3. Intelligence and security community insight
4. Network vulnerability data
5. OS vulnerability data
6. System CPU utilization reports
7. System disk utilization reports
8. System memory utilization report

Application Security Vulnerability Data

The InfoSec team evaluated several application security solutions in order to determine, which would meet the company's requirements. This was critical to evaluate as part of the big data platform since there were hundreds of Web sites that were exposed to the Internet.

After two months of evaluation, the company standardized on the WhiteHat Security's Sentinel which provides application security scanning, This selection was made for the following primary reasons:

1. The application security scans are non-intrusive
2. Data gathered from the scans is rich with meaningful data
3. All vulnerabilities are validated by a team of subject matter experts (SMEs)

Non-Intrusive Scans

Because of the nature of this company's environment, it was critical to have an application security scanning solution that was non-intrusive – not having an impact on the end user as well as the developer environments.

Data Gathered is Meaningful

Information that was presented by the application security scans had to be displayed in a meaningful way – for this organization, it was critical to have the vulnerability clearly displayed with information about how to resolve the vulnerability.

Validation of False Positives

The biggest challenge for the company was always working through the massive amounts of data when reviewing application security scan results. This is because most application security scanning solutions do not validate the false positives when scanning is completed. The effort to review the scans is then placed on the engineers, developers, and InfoSec teams. WhiteHat Sentinel was the only solution that validated the false positives and produced the results in a timely manner, which reduced the amount of time required by the stakeholders.

WhiteHat Sentinel Onboarding Process

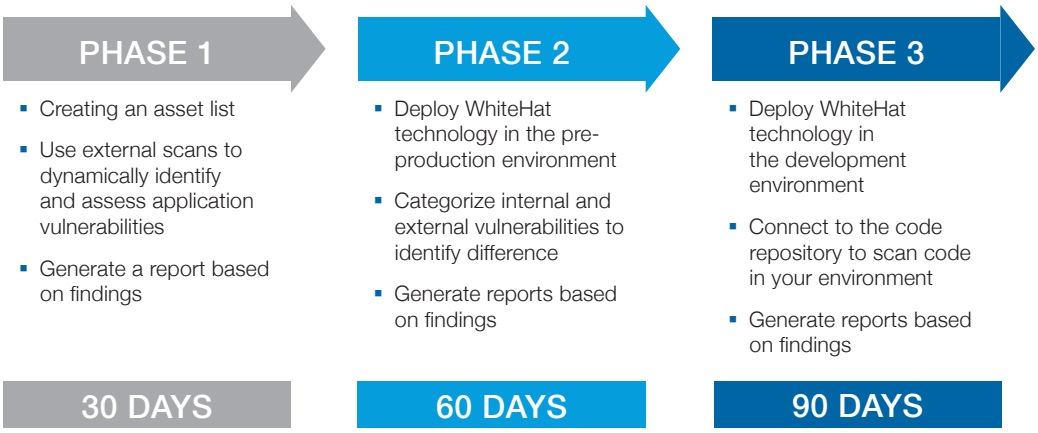


Figure. 3 WhiteHat Sentinel Onboarding Process

Phase 1 consisted of initial external scans on all Internet-facing applications. This was critical because the company had limited visibility into all Web sites owned and operated by the company. By starting with external, Internet applications, WhiteHat was able to create a master asset list, which identified all entities belonging to the company. This master asset list was then used to determine which WhiteHat solutions would be required.

Once external scans were successful, the organization moved to Phase 2, which allowed the company to move internal scans inside the organization further back through the development process. Because of this approach, WhiteHat was able to demonstrate that all source code was not the same in multiple environments.

In Phase 3, the company was able to move application security scanning to the development area by using source code analysis (SCA), further demonstrating the value of having source code scanned for application security vulnerabilities while source code is being checked in by developers.

By utilizing multiple offerings from WhiteHat, the company was able to demonstrate that all areas of development were successfully being scanned for application security vulnerabilities.

The diagram below displays the required high-level architecture:

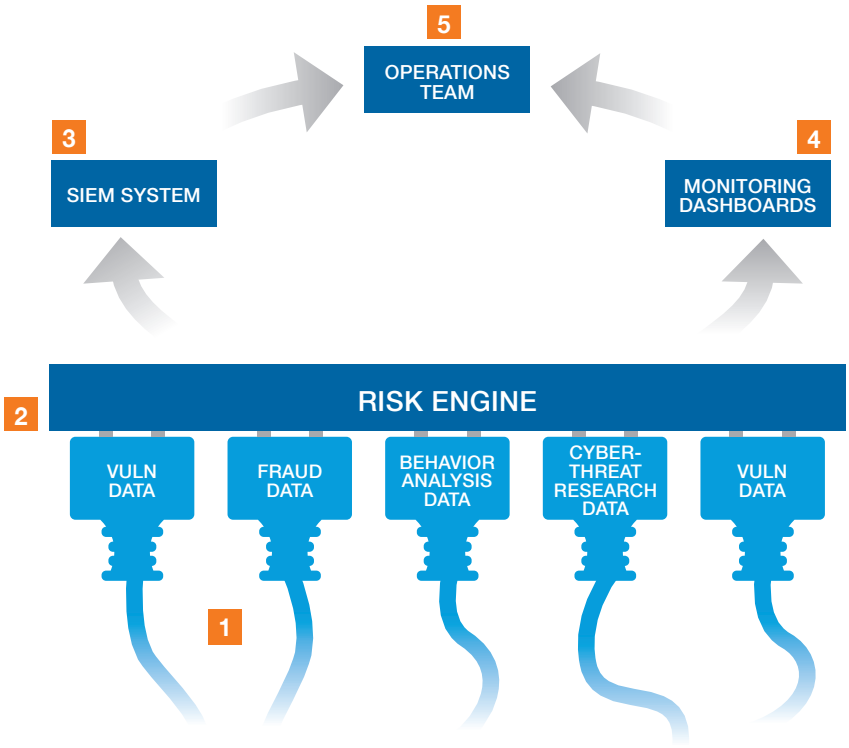


Figure 4. High-level architecture

The diagram included the concept of a power strip on the bottom (1). Of particular concern was the ability of the platform to ingest and correlate all InfoSec data, because, as one can imagine, the data generated by a multitude of InfoSec tools could be massive.

Another requirement emerged from the research and design discussions: a correlation and risk engine (2) would need to be included as part of the platform. The correlation and risk engine would quantify loss exposure in dollars—something the leadership team, the business, and executives required. From the correlation and risk engine, it was envisioned that data could then be processed through a SIEM system (3), which would eventually provide a 'heads-up' dashboard (4) for the IT Operations team (5) to review.

Technology Catches Up: The Information Security and Fraud Big Data Solution

Research conducted in mid-2012 proved there were multiple commercial solutions that could meet or exceed the requirements for building the BDS envisioned in 2010. The main requirements for the platform were modified to include the outcome of the research findings, with the additional requirements to keep the data onsite—the biggest driver for this solution.

Several big data companies were contacted in order to evaluate emerging technologies to support big data efforts. The goal was to determine if any products would meet the Company's business requirements.

During the evaluation period, when different aggregation and collection tools were utilized, a major issue surfaced. The team discovered that the SIEM technologies being evaluated could not ingest the different types of data elements the InfoSec team needed to process. The SIEM solutions weren't able to immediately identify an attack that took advantage of a new method or a never publicly-disclosed vulnerability. Multiple vendors and integrators confirmed that a SIEM would never be able to natively ingest the data the InfoSec team envisioned processing. SIEM vendors recommended creating an individual 'rule' or 'signature' for each of the different data elements the InfoSec team needed to ingest.

The approach and methodology used to address this inability to process the data by the SIEM vendors followed these steps, which introduced a variety of challenges:

1. Evaluate the data elements that needed to be processed through the SIEM
2. Develop a 'rule' or 'signature' for the SIEM solution
3. Test the 'rule' or 'signature' in a vendor-controlled lab
4. Fine-tune the 'rule' or 'signature' in a vendor-controlled lab
5. Deploy the 'rule' or 'signature' in a customer test environment
6. Deploy the 'rule' or 'signature' in a customer production environment
7. Validate the 'rule' or 'signature' in a customer production environment

Most SIEM vendors wanted to charge a professional service fee for the consulting work required to build a 'rule' or 'signature.' Additional fees for these services were priced by 'rule' or by 'signature.'

There were two main issues with this approach:

1. Timeliness of the threat and attack data: if an organization has to wait hours or days for a 'rule' or 'signature' to be created, what would the loss exposure be while the malicious activity continues during this period?
2. The number of 'rules' or 'signatures' that need to be created: There were thousands of non-traditional security rules that needed to be created and converted to a SIEM format—everything from behavior analytics, fraud rules, and system state. It was originally thought that having the data from the rules feeding into the SIEM would make the system perform faster and allow the team to better manage the rules. This wasn't the case—the traditional SIEM solutions were incapable of ingesting different data types.

From a business perspective, the approach of paying for custom 'rules' or 'signatures' wasn't practical and didn't scale. The InfoSec team decided to incorporate the traditional SIEM solution into a much larger InfoSec and Fraud BDS.

The process to evaluate the big data vendors took two months. Several vendors were considered; however, only one vendor met or exceeded all business requirements.

After the technology to support the core BDS was chosen, a new architecture emerged; the new solution utilized the following high-level design as shown in Figure 5:

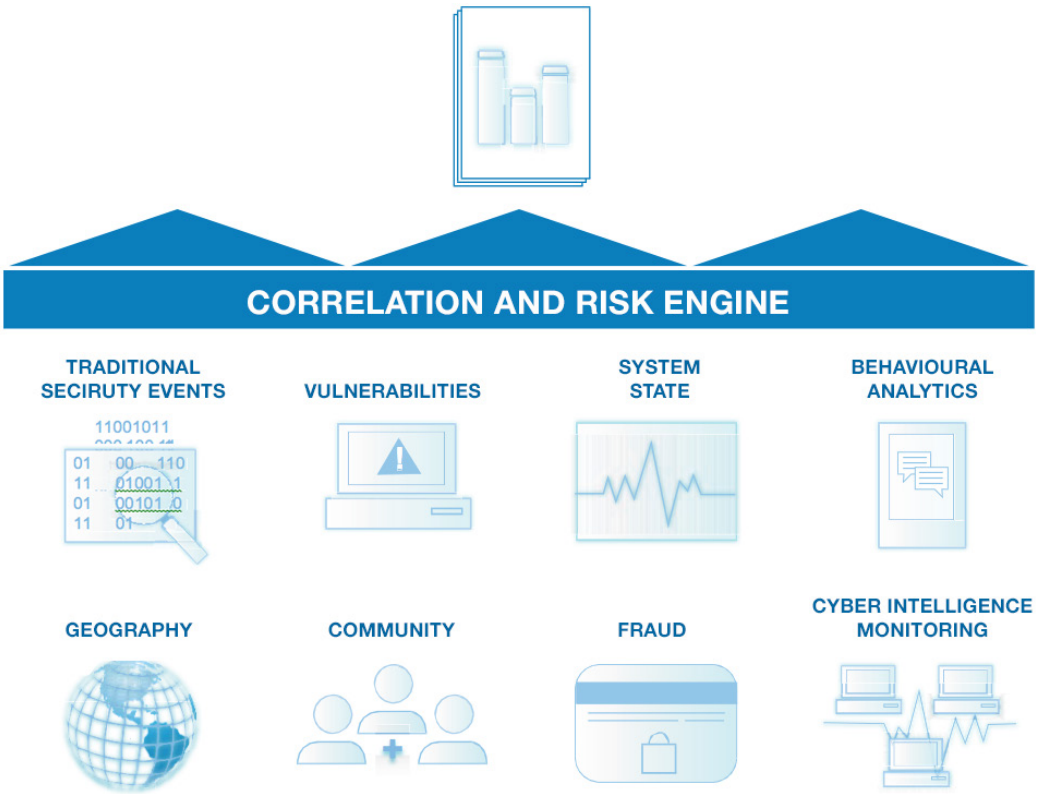


Figure 5. Correlation and Risk Engine

Many enhancements were made to the platform since the original concept. First, since traditional security events tracked by the SIEM were unable to process the data required by the InfoSec team, the SIEM data was added to the list of types of data to include for ingestion and correlation. Secondly, the technology used for the correlation and risk engine evolved into a combination of automatic and manual functions. The core functionality of quantifying loss exposure in dollars remained the same. This was largely because decisions elevated at this level had to be physically crosschecked by a human. Even though there were manual processes involved, this approach was faster than the legacy processes.

From the numerous inputs into the platform, data was fed and correlated into a risk engine, which fed into a dashboard for the end user in the IT Operations Center. Website events and alerts were displayed in near real-time directly within the dashboard. The data presented to the IT Operations team through this dashboard was also integrated with the team's existing dashboards.

DESCRIPTION	TYPES OF DATE ELEMENTS
Traditional InfoSec data typically housed in a SIEM	▪ File Integrity Monitoring (FIM) ▪ Firewalls ▪ Intrusion Detection Systems (IDS) ▪ Intrusion Prevention Systems (IPS) ▪ Network Traffic ▪ Web Application Firewalls (WAF)
Financial Reporting	▪ Financial data demonstrating how revenues were impacted in near realtime. This was used to trend minute-over-minute and hour-over-hour and then compared against the previous year. Having this data available was critical for understanding how potential loss exposure was going to impact the company in dollars
Security Vulnerabilities	▪ Application Security Vulnerabilities ▪ Network Vulnerabilities ▪ OS Vulnerabilities ▪ Penetration Testing Results
System State	▪ CPU Usage ▪ System Memory Usage ▪ System Disk Usage
Web Session Intelligence	▪ Behavior Scores—based on good or bad behavior ▪ High Velocity Traffic ▪ Man-in-the-Browser attacks ▪ Man-in-the-Middle attacks ▪ Man-in-the-Mobile attacks ▪ Page Clicks ▪ Page Views
Geography	▪ Geolocation ▪ IP Address ▪ Source Country
Community	▪ Ability to Anonymously Share Event Data Across Industries
Fraud Data	▪ Transaction-Based Fraud Data And Analytics
Cyber Threat Intelligence	▪ External threat research that provides early warning intelligence on what cyber criminals were plotting against the Company (e.g. account take over (ATO), gift card abuse, and DDoS attacks)

Table 1: A partial list of data elements used for collection, correlation, and reporting

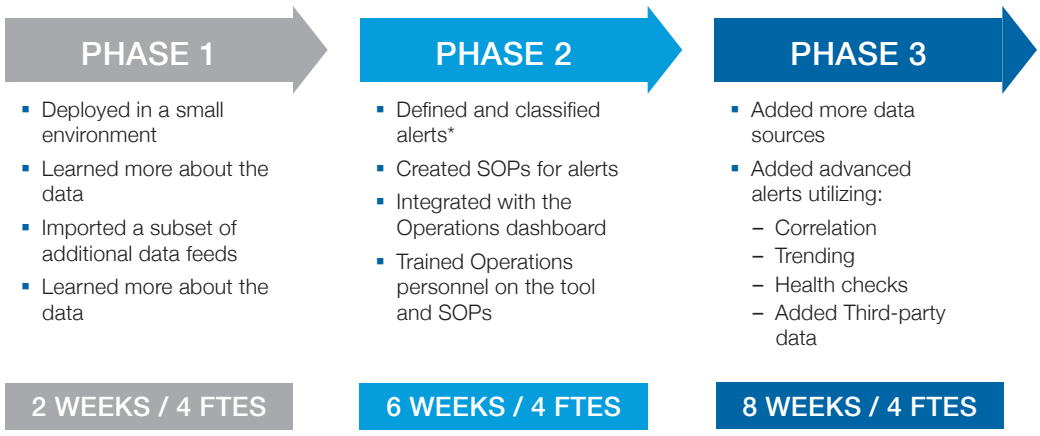
Managing the Project in Phases

Four full-time resources worked on this project for 16 weeks. The four-person team included:

1. InfoSec Architect
2. Developer
3. InfoSec Engineer
4. Senior Analyst

Using a phased approach, the InfoSec team attempted to break the project down into small, manageable deliverables. In doing so, the team quickly realized how easy it was to deploy the real-time InfoSec and Fraud BDS. One of the key requirements for success of this project was the integration of reporting dashboards into existing tools that were being used by the IT Operations Analysts. The InfoSec team did not want to create a completely different dashboard that the team would have to rely on for monitoring, alerting, and researching.

The process used for this approach is shown in the figure below.



** The types of alerts discovered through Phase 1 prompted the team to create an event-classification system based on several industry-recognized frameworks.*

Figure 6. Managing the Project in Phases

Phase 1

The InfoSec team deployed the BDS in a small, controlled lab environment with the primary goal of learning how data elements were to be ingested, correlated, stored, indexed, and searched. As the data elements were identified, it was clear that the BDS was going to be easier to implement than originally anticipated. Additional data feeds were added to the test platform.

The WhiteHat Sentinel application programming interface (API) was a powerful feature for ingesting the application security vulnerabilities seamlessly into the BDS.

This served two purposes:

1. The InfoSec team wanted to see how difficult it would be to ingest different types of data elements.
2. The InfoSec and IT Operations teams wanted to learn more about the data elements in order to analyze different types of patterns and trends within behaviors.

Ingesting data into the BDS was easy because of the ability to utilize vendor APIs.

Because all data sets were ingested without any major issues, the InfoSec team accelerated the deployment plans for the BDS and extended the scope of the project.

Data feeds for Phase 1 included the following data elements:

1. Traditional InfoSec data typically housed in the SIEM (e.g. firewall, WAF, IDS/IPS, file Integrity monitoring)
2. Application security vulnerabilities
3. Network security vulnerabilities
4. OS security vulnerabilities

RESOURCES	HOURS
InfoSec Architect	80
Developer	80
InfoSec Engineer	80
Senior Analyst	80
Phase 1 Total Hours	320
Phase 1 Cost	$40,000

**These figures assume a 40-hour work week with internal resources billing at $125 per hour.*

Table 2. The internal development costs for Phase 1

Phase 2

The second phase of the project involved the following functions:

1. Adding additional data elements
2. Defining and classifying the alerts
3. Creating the standard operating procedures (SOPs) for the alerts
4. Integrating the alerts within the IT Operations team dashboards
5. Training IT Operations personnel

Additional data elements added for Phase 2:

1. Web session intelligence
2. Fraud data
3. System state

Defining and classifying alerts was crucial since there needed to be a new way to review the combination of events. When the BDS was developed, there were few event-classification systems that met the business requirements for the InfoSec team so a proprietary event-classification system was created.

Events were manually reviewed through Phase 2 to see if there were strong relationships with other independent events. Combined events that indicated malicious behavior were categorized as an alert.

Alerts were broken down into the following categories:

1. Critical
2. Major
3. Minor
4. Informational

Another powerful use of WhiteHat Sentinel was being able to take the application security vulnerabilities and correlate the open vulnerabilities against high-risk countries that were attempting to exploit the site through malicious attacks.

The following is an example of a combination of events that would be escalated as a single alert:

- **WAF alert (behavior) + open appsec vuln (XSS) + IDS alert (attack) + geographical location (high-risk country) + behavior analytics score (greater than 80%) = Critical Alert**
- The information processed by the BDS used the proprietary defined data structure created for correlating or linking information together. Based on these new types of alerts, hundreds of new SOPs were designed and created to support combinations of events.
- The information processed by the BDS used the proprietary defined data structure created for correlating or linking information together. Based on these new types of alerts, hundreds of new SOPs were designed and created to support combinations of events.

RESOURCES	HOURS
InfoSec Architect	240
Developer	240
InfoSec Engineer	240
Senior Analyst	240
Phase 2 Total Hours	960
Phase 2 Cost	$120,000

**These figures assume a 40-hour work week with internal resources billing at $125 per hour.*

Table 3. The internal development costs for Phase 2

Phase 3

During Phase 3, the team extended ingesting other data elements:

1. System state
2. CPU activity
3. Disk activity
4. Memory usage
5. Fraud scores (post-authorization)
6. Cyber threat intelligence
7. Community

The biggest challenge during this time period was to ensure the events were tied together across multiple disciplines. To achieve this, the team successfully correlated events to demonstrate how different events occurred over time.

This, coupled with the ability to measure the frequency of the event, the threat capability (e.g. cyber-criminal behavior vs. script kiddie behavior), and how computing systems responded (e.g. increase in CPU usage, drive activity, or memory utilization) were essential to identifying the relationships across the different environments.

RESOURCES	HOURS
InfoSec Architect	320
Developer	320
InfoSec Engineer	320
Senior Analyst	320
Phase 3 Total Hours	1280
Phase 3 Cost	$160,000

**These figures assume a 40-hour work week with internal resources billing at $125 per hour.*

Table 4. The internal development costs for Phase 3

Total Cost of Ownership

Total Cost of Ownership (TCO) for the BDS can be broken down into two areas:

1. Training
2. Support

Training

Training was built into the approach and methodology as the BDS was being designed, developed, and deployed. Part of the strategy was to include the IT Operations team as part of weekly meetings to update everyone on the progress and to discuss how the alerts were going to be reclassified with the new event- classification framework.

Training was provided to 16 IT Operations center analysts every two weeks in one-hour sessions for the first 60 days. There was a commitment to provide this training every two months thereafter as the BDS expanded and new threats emerged.

Support for the BDS

Once the BDS was operational, support consisted of the follow elements:

1. One senior InfoSec subject matter expert (SME) spending 25% of their time reviewing weekly behavior patterns and fine-tuning the system.
2. One mid-level InfoSec analyst spending 15% of their time performing cyber threat research and tuning the system.
3. One system-monitoring developer assigned for future enhancements, maintenance, and support.
4. Monthly meetings for health checks from the vendors' professional services team.

Future costs that were considered when building out this new platform are listed in the table below:

RESOURCES	MONTHLY HOURS
InfoSec Architect	16
Developer	160
InfoSec Engineer	16
Senior Analyst	4
Ongoing Monthly Support	196
Monthly Cost	$24,500
Yearly Cost	$294,000

**These figures assume a $125 internal hourly bill rate.*

Table 5. Ongoing BDS Support

Operational Efficiencies: Reviewing Critical Web Application Firewall Alerts Before and After the Big Data Solution

Hundreds of powerful use cases demonstrated the business value of the BDS once it was running and operationalized. All stakeholders were able to confidently analyze information faster than before while quantifying loss exposure in dollars.

Thousands of WAF alerts occur in this environment on a daily basis. It's impossible to manually review all alerts flagged as malicious, so it's imperative to identify which WAF alerts are critical.

Before the BDS, multiple resources had to work for several hours to review a critical WAF alert. The types of manual reviews conducted before the BDS consisted of the following teams and systems being utilized for a WAF alert review.

TEAM	LOG FILES (CDN)	LOG FILES FIREWALL	LOG FILES SYSTEMS	LOG FILES WAF	SIEM
Application Engineering			■		
InfoSec		■		■	■
IT Operations	■	■	■	■	■
Network Enginerring		■		■	
Performance Engineering			■		

This manual review also consisted of multiple conference calls.

Table 6. Firewall alerts before the data solution.

After the BDS platform was deployed, two resources could research multiple critical WAF alerts in less than 10 minutes.

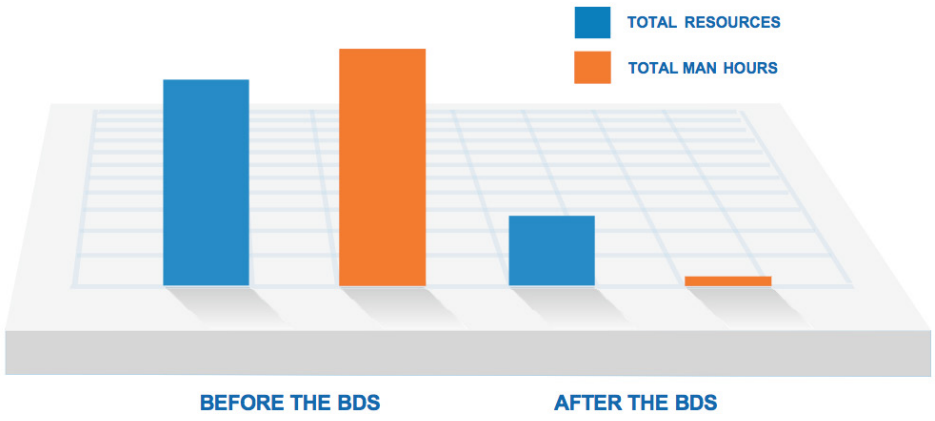


Figure 7. Before and after BDS

Predicting a Distributed Denial of Service (DDoS)

A successful distributed denial of service (DDoS) attack will render a site inoperable and unable to conduct its business functions by overloading system resources. The BDS identified and alerted teams to behaviors indicative of an active DDoS.

The precursor to one type of DDoS in this environment was indicated when excessive queries on women's shoes and clothing were simultaneously performed from more than three source countries. During one of the more advanced DDoS attacks the Company experienced, the InfoSec and IT Operations teams gathered information showing that malicious traffic had increased to 861,000 connections per minute. The 861,000 connections per minute were generated from the RussKill bot. One of the most disturbing aspects of this type of DDoS was that the traditional InfoSec tools reported the behavior as appearing 'normal' in the environment. A combination of web session intelligence and cyber threat intelligence solutions alerted that the behavior was anomalous. This allowed the InfoSec and IT Operations teams to correlate and predict future DDoS events.

A DDoS affects more than just online revenues for a retailer—it also affects the entire daily operation. The table below tracks how disruptive a DDoS can be for an online retailer:

TEAM TO RESEARCH, REMEDIATE AND ADDRESS THE ISSUE	RESOURCES	TIME	TOTAL HOURS	ESTIMATED COSTS
Call Center Support	6	120	720	$77,400
Development	4	120	480	$51,600
Incident Management Teams	4	160	640	$68.800
InfoSec	4	240	960	$103.200
Legal	2	120	240	$25,800
Loss Prevention / Fraud	2	120	240	$25,800
Network Teams	4	120	480	$51,600
Operations	10	80	800	$86,000
Program Team	16	120	1920	$206,400
QA	4	120	480	$51,600
System Engineering	6	120	720	$77,400
Total Per Incident	62	1440	7680	$825,600

*These figures do not include lost revenues or opportunity costs. These figures are just representative of the impact on operations. *These figures assume a $107.50 blended hourly bill rate.*

Table 7. Before the BDS

TEAM TO RESEARCH, REMEDIATE AND ADDRESS THE ISSUE	RESOURCES	TIME	TOTAL HOURS	ESTIMATED COSTS
Call Center Support	1	10	10	$1,075
Development	1	10	10	$4,300
Incident Management Teams	1	40	40	$4,300
InfoSec	1	30	60	$6,450
Legal	1	20	20	$2,150
Loss Prevention / Fraud	1	4	4	$430
Network Teams	1	40	40	$4,300
Operations	2	20	40	$4,300
Program Team	1	10	10	$1,075
QA	1	10	10	$1,075
System Engineering	1	10	10	$1,075
Total Per Incident	12	204	254	$27,305

**These figures assume a $107.50 blended hourly bill rate.*

Table 8. After the BDS

Business Logic Abuse—The Rebate King

Business logic abuse occurs when an exploit takes advantage of a flaw in the programming of an application.

One example is a cyber criminal nicknamed “The Rebate King” who discovered a business logic flaw in the Company’s marketplace functionality.

Within a marketplace, vendors can – for a fee – host their items for sale by using an existing infrastructure, marketing engine, and templates. By using a marketplace, vendors are able to set up a storefront quickly and generate immediate website traffic.

For this use case, The Rebate King would normally perform these steps (first as a vendor and then as a customer). These steps were taking advantage of a loyalty program to earn points as well as to receive a paid commission and an instant cash rebate.

As a Vendor:

1. The Rebate King created a vendor account, ID, and password.
2. The Rebate King would then enter the marketplace site as a vendor to post and sell items.
3. In this example, The Rebate King posted high-end plotter cartridges for sale at $1,500 each.
4. The Rebate King also received a commission for hosting products on the marketplace site.
5. The ‘vendor’ in this case would then log out and then log back in as the ‘customer’ by visiting the online store—the vendor and the customer being one and the same.

As a Customer:

1. The Rebate King (now as the customer) would log in and purchase 10 plotter cartridges and receive an instant 10% rebate on the individual item, which is managed by a third-party provider.
2. The 'customer' attempted to use stolen credit cards for this purchase.
3. The 'customer' would attempt to receive an instant $150 rebate per item purchased ($1,500 is the purchase price per item for this example).
4. The 'customer' would also attempt to collect points to be used for a loyalty program.

The intention of the program was to never allow behavior like this; however, the BDS identified the malicious behavior and allowed the InfoSec team to make the necessary corrections to resolve the issue.

Single Sign-On Abuse

In this next example, cyber criminals used a consolidated user ID and password list acquired from previous external, non-Company breaches. Since many users use the same ID and password for logging in to multiple sites, these cyber criminals were testing to see if this list of credentials was valid on the Company's websites. If the credentials were valid, they could be used at multiple sites (eg. retail, finance, insurance, utilities, communications).

Several source IP addresses were identified as attempting to log into a Single Sign-On (SSO) page—a service that enables users to log into one page and gain access to multiple Company online environments. The activity identified in this example was not like that of a typical user. When a typical user lands on an SSO page, it's usually from conducting other activities on a website such as searching, shopping, reading, or through some sort of re-direction from another website that is integrated as part of the business functionality. The activity identified by the BDS in this example indicated that the attackers were attempting to log into the system with tens of millions of stolen user IDs and passwords.

The team validated this activity with other global security practitioners through the function called 'Community' mentioned earlier in this document. The InfoSec and IT Operations teams watched the behavior to ensure there were no successful logins.

Application Security Vulnerability: Attempted Exploits on a Ruby on Rails System

This is the most malicious example that will be discussed in this case study. If successful, the outcome would have been catastrophic, and cyber criminals would have literally controlled all of the Company's data centers and supporting systems.

When an application security vulnerability is discovered, cyber criminals will probe the architecture of the application and supporting systems to perform reconnaissance as a means to determine the resistance strength of a website. This behavior is not only challenging to isolate—because it appears to be normal behavior—but also truly frightening to imagine how much damage can be done if a cybercriminal penetrates a company's infrastructure.

This particular example is about an Open Source framework called Ruby on Rails (RoR). Figure 8 below was created to walk through the workflow of what occurred during this attack.

The key takeaway from this example is the usefulness of the BDS in providing actionable information in seconds as opposed to hours, days, or weeks.

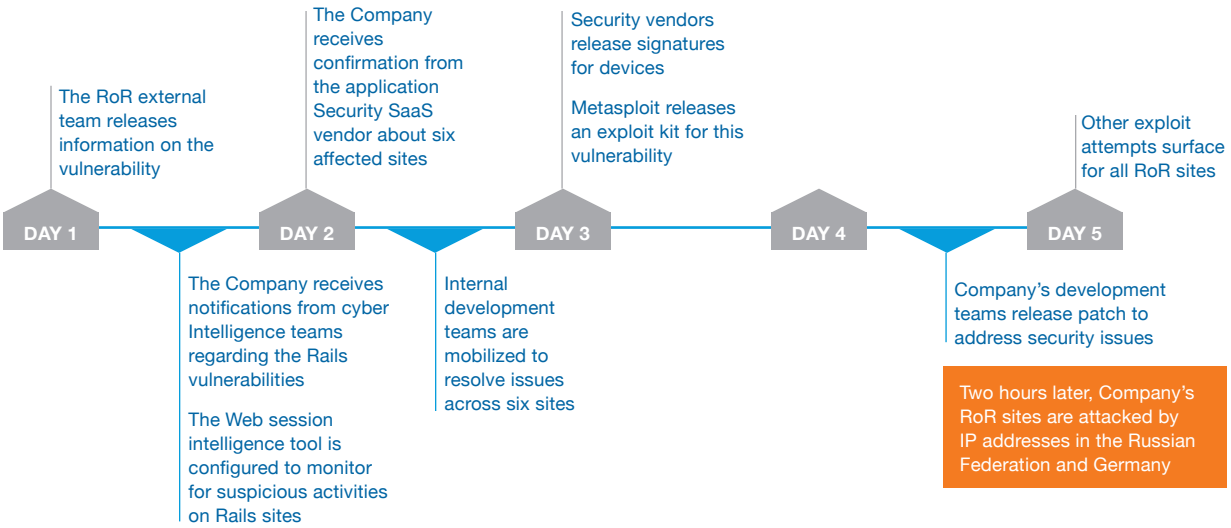


Figure 8. Ruby on Rails (RoR) Open Source Framework

Day 1:

There was a global public announcement regarding a critical application vulnerability, which stated RoR had a zero-day vulnerability. The Company's InfoSec team mobilized to perform research, which confirmed the number of websites that were going to be affected—six total sites were identified as running RoR.

Shortly after the RoR security announcement, the Company received confirmation from a third-party cyber threat intelligence company that cyber criminals were planning to exploit the RoR vulnerability on a global scale. During this time, there were approximately 250,000 RoR websites throughout the world.

The Company's InfoSec team created alerts specific for this type of behavior in the web session intelligence software. This was to identify and isolate any malicious behavior against the six affected websites.

Day 2

The Company received confirmation from WhiteHat that there were six affected Company RoR sites. Internal development teams were mobilized to determine the best way to resolve the identified vulnerabilities.

Day 3

Multiple security vendors released signatures for IDS, IPS, and firewall solutions. The InfoSec team needed to coordinate with multiple internal teams to ensure the updates were functioning properly.

Several hours after the vendors' signatures were released, Metasploit released the exploit kit for the RoR vulnerability.

Day 4

The Company's internal development teams deployed a solution to resolve the RoR security issue.

Within two hours of the fix being promoted to production, the IT Operations team received an alert that identified IP addresses located in Germany and the Russian Federation were attempting to exploit the RoR vulnerability on all six Company websites. The traffic of the exploitation attempt was at a high velocity and behavior scores from the web session intelligence software confirmed that this web behavior was malicious. Highly-skilled cyber criminals were conducting an attack.

Day 5

During Day 5, the IT Operations personnel identified additional attempts to exploit the RoR vulnerability.

Benefits for the Business

The InfoSec team that owned the rollout and support of this platform was able to break down silos of information within the business unit. Prior to the deployment of the BDS, multiple data silos existed because various teams created their own information repositories.

There are numerous benefits to breaking down these silos for InfoSec and Fraud teams. Additional benefits to the organization included providing rich data to other stakeholders in the business unit, as listed in the table below:

TEAM	VALUE
Application Engineering	Evaluating website and application speed, reliability, and overall performance.
Business Owners	Trending information such as comparing sales revenues from one year to another.
Marketing	Assessing the effectiveness of marketing efforts by analyzing all website traffic.
Operations	Identifying good or bad traffic on websites and responding to malicious traffic in near real-time.

Table 9. Benefits for the business

Lessons Learned

In 2009, the InfoSec team made a conscious decision to collect and retain as much data as possible to determine if certain relationships—when correlated—would yield more information about malicious activity.

It is imperative that each organization determines its own strategy for utilizing a BDS. How will it be funded? Who will the business owner be for such a solution? A successful InfoSec and Fraud program's success is hinged on the organization aligning to the business needs and using a risk-based approach where issues are quantified in dollars.

Security Events Are Increasing

Unfortunately, the number of security events is not decreasing. The figure below demonstrates the number of security events increasing over the past four years. The graph demonstrates security events increasing 2x between 2010-2011, 2x between 2011-2012, and then 6x between 2012-2013.

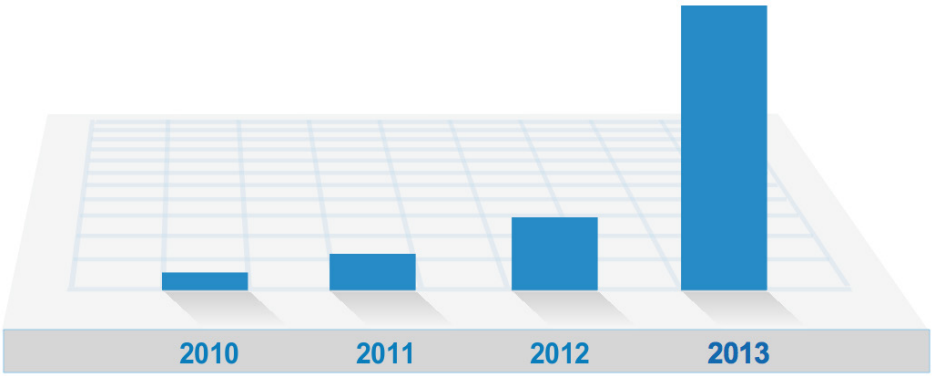


Figure 9. Number of Security Events Increasing Over Time

Risk Models Need to Be Re-Evaluated

The traditional formula used for determining risk should be addressed:

Some would argue this formula became extinct years ago. It is Blue Lava's position that the formula is merely stale and no longer worth using within a progressive InfoSec and Fraud program. In short, this out-of-date formula can no longer keep up with emerging threat methods.

New risk models are available to quantify losses in dollars that the business teams understand.

Implement a Risk-Based Approach to InfoSec and Fraud Analytics

It is critical for organizations to move towards a risk-based approach in building out its InfoSec and Fraud BDS. One key ingredient is to identify the risks and loss exposure by using a monetary value. Executives and business owners are asking for this in their reports. Aligning to the business will be a key component to success.

The InfoSec team standardized on the Factor Analysis of Information Risk (FAIR) for the framework used to quantify losses in dollars. This decision was made after extensive research on risk frameworks was conducted. The FAIR solution is the only solution that quantified the loss exposure in dollars, which aligned to the business. Table 10 provides a comparison of different types of risk frameworks and how FAIR compares to them.

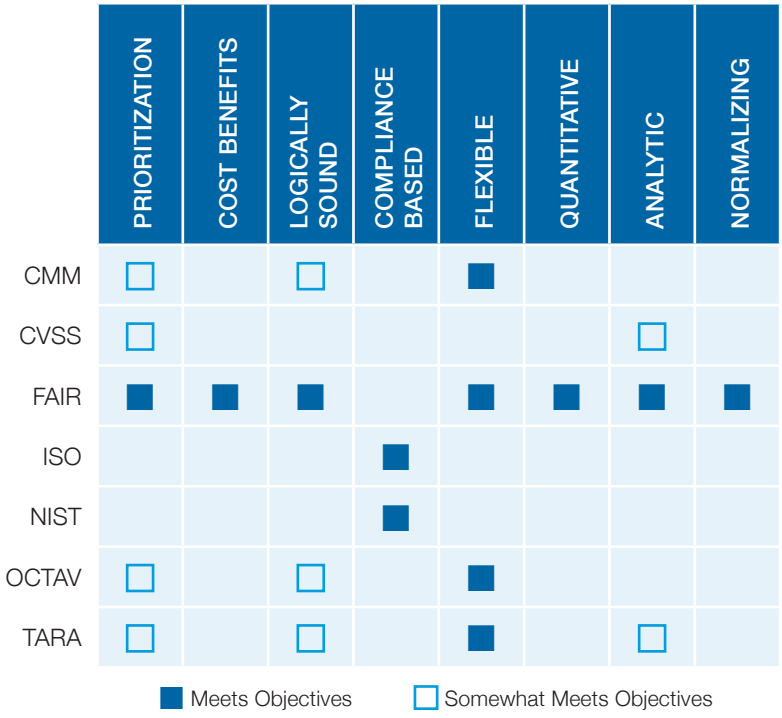


Table 10. Benefits for the business

Teams Need Data in Seconds (Or Less)

The real-world examples mentioned in this case study highlight the need to have data available in seconds to multiple teams. This becomes possible when integration of several data sources occurs across multiple disciplines. This approach allowed the InfoSec and IT Operations teams to make better decisions about how to address different types of malicious behavior. There are many use cases from this effort where the BDS was able to 'learn' about events, and by using proprietary scripts, systems and applications were able to perform 'self-learning' or 'self-healing.'

Daily Meetings: Critical Early in the Development

During Phase 1, the four-member team incorporated twice-a-day 30-minute meetings to review efforts. The morning meeting discussed the goals for the day. The afternoon meeting discussed progress that had been made and needed enhancements to address the following day. As the team entered into Phase 2, the meet- ings were reduced to one 30-minute meeting each afternoon. Having the daily meetings aided in the success of this effort.

Storing System Data Elements

For the BDS to perform as fast as it did, the system processed data on local disks and then periodically archived the data to network-attached storage (NAS). This approach required more funding to implement; however, the trade-off was the data was available in seconds as opposed to minutes, hours, or days.

Trending Cyber Criminal Behavior

When deploying data in this type of environment, it is highly recommended that organizations consider a strategy for understanding cyber criminal trends. Organizations will find seasonal patterns and outlying behavior that jump out at them through the dashboards and visualization tools they implement. Develop a strategy to trend cyber criminal behavior over time.

Include Vendor Professional Services in Your Budget: You Can't Do This Alone

Organizations will need subject-matter-experts at some point—it's inevitable. The point when the Company's InfoSec team needed coaching was in the latter part of Phase 1 and then during the first part of Phase 2. The InfoSec team needed some vendor time to ensure the approach was well thought-out. The InfoSec team did not want to 'go back' and retrofit the solution because of something that was overlooked during the architecture, design, or development phases.

Where Do I Start?

What's exciting about the approach used in this case study is that most organizations have this data. The data may reside in separate systems or environments; however, the data is out there. It is imperative to build relationships and work together in order to achieve this goal.

One approach to use for taking on a big data effort for InfoSec and Fraud may look something like the following phased rollout.

	PHASE 1	PHASE 2	PHASE 3
SIEM Data	■		
Application Security	■		
Network Security Vulnerabilities	■		
OS Security Vulnerabilities	■		
Risk Framework	■		
Web Session Intelligence		■	
Fraud Data		■	
System State		■	
Cyber Threat Intelligence			■
Community			■

Table 11. InfoSec and Fraud phased rollout

Your Industry Isn't the Only One Affected

The malicious examples referred to in this case study are real and there were hundreds of other use cases that could have been referenced in this case study. If any of the malicious attempts were successful, the impact against the Company would have been devastating. One key point to these events: the organization experiencing an attack is not the only one affected. Each of the events experienced by an InfoSec and IT Operations team is tied to multiple industries. If a system is compromised and sensitive data resides on the systems when a cyber criminal acquires it, it's not just one company that is tied to the event. The event is tied to a company, the supporting financial institutions, and insurance companies that underwrite the cyber liability insurance, and so on. When things go bad, it causes a massive ripple effect through multiple industries.

About Blue Lava Consulting, LLC

Blue Lava Consulting works in a strategic partnership with organizations to assess Information Security programs, Information Security risks, and to build an efficient set of Information Security and fraud controls. Our experience providing Information Security coaching, IT risk management, and research allows us to tailor our strategies in delivering superior results with the optimum balance of business resiliency and agility. The Blue Lava team is disciplined to work with organizations to provide a detailed and comprehensive knowledge transfer through our engagements. Blue Lava clients throughout the world include emerging technology companies and Fortune 500 organizations, which will attest to our knowledge and experience in these areas.

About WhiteHat Security

Founded in 2001 and headquartered in Santa Clara, California, WhiteHat Security is the leader in application security, enabling businesses to protect critical data, ensure compliance, and manage risk. WhiteHat is different because we approach application security through the eyes of the attacker. Through a combination of technology, more than a decade of intelligence metrics, and the judgment of real people, WhiteHat Security provides complete web security at a scale and accuracy unmatched in the industry. WhiteHat Sentinel, the company¹s flagship product line, currently manages tens of thousands of websites including sites in highly regulated industries, such as top e-commerce, financial services, and healthcare companies. For more information on WhiteHat Security, please visit www.whitehatsec.com

WhiteHat Security, Inc. | 3970 Freedom Circle | Santa Clara, CA 95054 | 1.408.343.8300 | www.whitehatsec.com