Big security for big data
# Table of contents

3  Big security for big data  
3  Data explosion  
3  Data collection  
5  Data integration  
5  Data analytics  
5  In real time versus the past  
6  Born for faster speed  
6  Real-time threat evaluation  
7  Pattern matching  
7  Statistical correlation  
7  Monitor and respond  
8  Conclusion  
8  About HP Enterprise Security  
8  Protect your business
Figure 1. Data is generated at a much faster pace

Big security for big data

We are children of the information generation. No longer tied to large mainframe computers, we now access information via applications, mobile devices, and laptops to make decisions based on real-time data. It is because information is so pervasive that businesses want to capture this data and analyze it for intelligence.

Data collection

Your first job is to aggregate all the information from every device into one place. This means collecting information from cloud, virtual, and real appliances: network devices, applications, servers, databases, desktops, and security devices. With Software-as-a-Service (SaaS) applications deployed in the cloud, it is important to collect logs from those applications as well since data stored in the cloud can contain information spanning from human resource management to customer information. Collecting this information gives you visibility into who is accessing your company’s information, what information they are accessing, and when this access is occurring. The goal is to capture usage patterns and look for signs of malicious behavior.

Typically, data theft is done in five stages: First, hackers “research” their target in order to find a way to enter the network. After “infiltrating” the network, they may install an agent to lie dormant and gather information until they “discover” where the payload is hosted, and how to “acquire” it. Once the target is captured, the next step is to “exfiltrate” the information out of the network. Most advanced attacks progress through these five stages, and having this understanding helps you look for clues on whether an attack is taking place in your environment, and how to stop the attacker from reaching their target. The key to determining what logs to collect are to focus on records where an actor is accessing information or systems.

Data explosion

The multitude of devices, users, and generated traffic all combine to create a proliferation of data that is being created with incredible volume, velocity, and variety. As a result, organizations need a way to protect, utilize, and gain real-time insight from “big data.” This intelligence is not only valuable to businesses and consumers, but also to hackers. Robust information marketplaces have arisen for hackers to sell credit card information, account usernames, passwords, national secrets (WikiLeaks), as well as intellectual property. How does anyone keep secrets anymore? How does anyone keep secrets protected from hackers?

In the past when the network infrastructure was straightforward and perimeters used to exist, controlling access to data was much simpler. If your secrets rested within the company network, all you had to do to keep the data safe was to make sure you had a strong firewall in place. However, as data became available through the Internet, mobile devices, and the cloud having a firewall was not enough. Companies tried to solve each security problem in a piecemeal manner, tacking on more security devices like patching a hole in the wall. But, because these products did not interoperate, you could not coordinate a defense against hackers.

In order to meet the current security problems faced by organizations, a new paradigm shift needs to occur. Businesses need the ability to secure data, collect it, and aggregate into an intelligent format, so that real-time alerting and reporting can take place. The first step is to establish complete visibility so that your data and who accesses the data can be monitored. Next, you need to understand the context, so that you can focus on the valued assets, which are critical to your business. Finally, utilize the intelligence gathered so that you can harden your attack surface and stop attacks before the data is exfiltrated. So, how do we get started?
Four steps to security intelligence

**1. Data collection** from cloud, virtual, and real devices for complete visibility into data and its accessibility.

**2. Data integration** through automation and rule-based processing. HP ArcSight normalizes and categorizes log data into over 400 meta fields.

**3. Data analytics** which involves combining logs from multiple sources and correlating events together to create real-time alerts.

**4. HP ArcSight Correlation Optimized Retention and Retrieval (CORR) Engine** serves as a foundation for threat detection, security analysis, and log data management.

**Benefits**

The CORR-Engine helps security analysts to:
- Detect more incidents
- Address more data
- Operate more efficiently
- Evaluate threats in real time
- Find threats faster
- Collect relevant information about user roles, critical assets and data in real time and uses it to reduce false-positives

**Results**

- Security event monitoring is simple, intelligent, efficient, and manageable
- HP ArcSight Security Event Information Management (SIEM) processes events faster making security information available in real time
**Data integration**

Once the machine data is collected, the data needs to be parsed to derive intelligence from cryptic log messages. Automation and rule-based processing is needed because having a person review logs manually would make the problem of finding an attacker quite difficult since the security analyst would need to manually separate attacks from logs of normal behavior. The solution is to normalize machine logs so that queries can pull context-aware information from log data. For example, HP ArcSight connectors normalize and categorize log data into over 400 meta fields. Logs that have been normalized become more useful because you no longer need an expert on a particular device to interpret the log. By enriching logs with metadata, you can turn strings of text into information that can be indexed and searched.

**Data analytics**

Normalized logs are indexed and categorized to make it easy for a correlation engine to process and identify patterns based on heuristics and security rules. It is here where the art of combining logs from multiple sources and correlating events together help to create real-time alerts. This preprocessing also speeds up correlation and makes vendor-agnostic event logs, which give analysts the ability to build reports and filters with simple English queries.

**In real time versus the past**

Catching a hacker and being able to stop them as the attack is taking place is more useful to a company than being able to use forensics to piece together an attack that already took place. However, in order to have that as part of your arsenal, we have to resolve four problems:

- How do you insert data faster into your data store?
- How do you store all this data?
- How do you quickly process events?
- How do you return results faster?

**Figure 3.** Performance improvements of ESM with CORR-Engine over ESM with Oracle

- **Detect more incidents**
  - Up to 3x the current performance (events per second [EPS]) using the same hardware
  - Faster query 15x

- **Address more data**
  - Up to 20x the current capacity for correlated events using the same disk space

- **Operate more efficiently**
  - Frees up security analyst cycles for proactive monitoring
  - No DBA needed
At HP ArcSight, we have been evolving our solution for over 12 years. When we created our first SIEM product, Oracle’s database was the best data store. However, as the problem space of our SIEM customers evolved over the years and big data became prevalent; it was important to redesign our solution to handle the new challenges. The data store now needs to capture more events, compress and archive more data, and execute searches much faster.

**Born for faster speed**

When we originally introduced this technology into our logger solution, customers could see the benefits. HP ArcSight’s CORR-Engine (correlation optimized retention and retrieval) is uniquely architected to enable a single instance to capture raw logs at rates of above 100,000 events per second, compress, and store up to 42 TB of log data per instance and execute searches at millions of events per second.\(^2\) By creating our own data store that utilized both column and row-store technology, we were able to marry the significant performance benefits with the flexibility of free-form unstructured searches, all while providing a very intuitive, easy-to-operate user interface.

The CORR-Engine serves as a foundation that provides the speed needed for today’s threat detection, security analysis, and log data management. By processing more events, it can soon identify the meaning of any event by placing it within context of what, where, when, and why that event occurred and its impact on the organization.

Our correlation delivers accurate and automated prioritization of security risks and compliance violations in a business relevant context. Real-time alerts show administrators the most critical security events occurring in the environment, along with the context necessary to further analyze and mitigate a breach. Using CORR-Engine, administrators and analysts are able to:

**Detect more incidents**
- The new architecture allows event correlation rates of up to 3x the current performance using the same hardware.

**Address more data**
- The new architecture enables storage capacity of up to 20x the current capacity for correlated events using the same disk space.\(^2\)

**Operate more efficiently**
- The use of a common data store allows both the real-time correlation application and the log management application to use the same set of data, providing a seamless workflow that includes detection, alerting, and forensic analysis and reporting.

**Find threats faster**
- The graph above shows the multiples of improvement when we switched from RDBMS to our own-patented data store utilizing our new CORR-Engine.

This new capability allows users to search for any string or “keyword” located in the database, regardless of the event type or source. The HP ArcSight CORR-Engine indexes both raw (unstructured) and normalized (structured) event data to provide rapid search capabilities. With our combined flat-file and RDBMS technology, HP ArcSight can return search results in excess of millions events per second for both structured and unstructured data.

As a result of using this new data store, security administrators could focus on finding malicious activities, not on tuning or managing the database. Also, central to our ability to process more events in real time, the new CORR-Engine permitted additional parallel processing capabilities, up to 80 CPU cores, big enough for the biggest organizations on the planet. By adding parallel processing power, HP ArcSight can handle more events, faster in an easy-to-use interface.

**Real-time threat evaluation**

HP ArcSight also makes use of actor information as a variable in its threat formula that collects information regarding identity management user roles, critical assets, vulnerability data, and “watch lists” in real time and uses this information to reduce false-positives and monitor critical infrastructure in memory. For example, if a Microsoft® SQL Server injection attack is targeting an Oracle database, HP ArcSight immediately lowers the severity of the attack, knowing that Oracle is not susceptible to MS SQL attacks. However, if a privileged user is accessing a critical piece of infrastructure after regular working hours and inserting a USB thumb drive into their system, this may generate a number of low severity events. Pieced together, HP ArcSight would immediately raise the severity of this activity based on the understanding of the user’s privileges and the assets criticality. This would start the alert process and start monitoring activity and workflow processes for a potential security breach.

*Figure 4. Correlation is the key to making sense of 1s and 0s*
**Pattern matching**

HP ArcSight has an expansion pack: Threat Detector which allows customers to mine through archived data looking for relationships between events that would have been missed by real-time correlation.

As an example, a low-and-slow attack takes place when an attacker purposely lowers the threshold on their attack to avoid detection. Such an evasive technique might be when the attacker is using a dictionary attack to guess a user’s password. They would not try to brute-force the authentication system all at once, as the system would lock out the user’s account after a series of unsuccessful login attempts. So the attacker uses a scripted stealth method of only attempting to login twice while trying to guess the password, then sleeps for five minutes and continues to invoke two attempts every five minutes. This means there would be 576 unsuccessful login attempts daily, but since most correlation rules look for brute-force methods, only a routine that would mine through historical data would be able to match this pattern. Threat Detector would detect this attack and then allow customers to introduce new rules that would block the attacker going forward.

**Statistical correlation**

HP ArcSight’s multidimensional correlation engine combines real time, in memory event log data with asset awareness, asset vulnerability, and identity correlation to assist operating teams with immediate detection of threats. The powerful correlation engine allows you to maintain a state of situational awareness by processing millions of log events in real time. We help to prioritize critical events so that your security administrator can review only those events that need specialized attention. With built-in network asset and user models, HP ArcSight is uniquely able to understand who is on the network, what data they are seeing, and which actions they are taking with that data.

HP ArcSight Enterprise Security Manager (ESM) uses a heuristic analytics model to keep a baseline of activity from events received by ESM and monitors any increases in attack, target, protocol, or user activity using a percentage threshold. The statistics that are calculated are used by ESM to determine spikes in the baseline average as well as other deterministic activity such as anomalous behavior, session reconciliation, effectiveness of IDS and firewalls as well as monitoring DHCP lease activity. This statistical baseline is also used for determining anomalous user or application-usage behavior.

**Figure 5. Smart correlation**

![Smart correlation diagram]

**Intelligent threat and risk detection**
- Sophisticated correlation technologies
- Pattern recognition and anomaly detection to identify modern known and unknown threats
- The more you collect, the smarter it gets

**Monitor and respond**

HP ArcSight proactively alerts and notifies you when malicious activity has occurred in your environment. However, because of the ability to process events quickly, we can alert your analysts in real time. For example, if we detect a distributed denial of services (DDoS) attack, we can send an email to you and your team, and notify you via your mobile device. A priority 1 escalation alerts your team so that a response can be mobilized against a prioritized security event. For example, if the Tier 1 team doesn’t acknowledge a notification within a certain timeframe, HP ArcSight can automatically escalate this to your Tier 2 team, tying into your existing response processes and procedures.

Once you’ve received a notification, you can start to analyze and investigate your environment using our easy-to-use data driven capabilities. Our dashboards help you visualize where your data is located and provide specialized views from business oriented to geographical oriented to systems oriented. From the dashboard, we can drill into the supporting events, drill into any level of detail, and customize the view and presentation of that data. And with our strong visualization capabilities, you can easily understand the significance of the data.

IT must be able to respond quickly, efficiently, and accurately to help minimize damage to the enterprise. HP ArcSight Threat Detector follows a simple three-step methodology:

- Discover the systems on your network
- Analyze what actions we should take and which offer the best results
- Provide guidance on what to do

By using HP ArcSight Threat Detector, you can:

- Reduce your response time from hours to seconds
- Simulate response actions before applying changes
- Cut off threats at the most effective choke points
- Automatically document all changes for audit or rollback
Conclusion

In today’s business environment, having access to the right information means making the right decision critical to surviving. Businesses need to protect their intelligence as it accumulates much faster because of big data. With HP ArcSight ESM, you can process big data events at faster speeds, get results in real time so that your business is getting the security information when it needs it the most in real time. With HP ArcSight CORR-Engine, security event monitoring is simple, intelligent, efficient, and manageable.

About HP Enterprise Security

HP is a leading provider of security and compliance solutions for the modern enterprise that wants to mitigate risk in their hybrid environment and defend against advanced threats. Based on market-leading products from HP ArcSight, HP Fortify, and HP TippingPoint, the HP Security Intelligence Platform uniquely delivers the advanced correlation, application protection, and network defenses to protect today's hybrid IT infrastructure from sophisticated cyber threats.

HP Services

HP ESP Global Services take a holistic approach to building and operating cyber security and response solutions and capabilities that support the cyber threat management and regulatory compliance needs of the world's largest enterprises. We use a combination of operational expertise—yours and ours—and proven methodologies to deliver fast, effective results and demonstrate ROI. Our proven, use-case driven solutions combine market-leading technology together with sustainable business and technical process executed by trained and organized people.

Learn more about HP ESP Global Services at hpenterprisesecurity.com.

Protect your business

Find out how to strengthen your security intelligence with HP ArcSight. Visit hp.com/go/hpesm.

---


Source: ESM 6.0c: Beta-Test, HP ArcSight QA and Dev team, August 2012