SANS ANALYST PROGRAM

Sponsored by NitroSecurity

Benchmarking Security Information Event Management (SIEM)

A SANS Whitepaper – February 2009

Written by J. Michael Butler

SIEM Benchmarking Process

The Baseline Network

SIEM Storage and Analysis

Advisors:

Stephen Northcutt, *SANS CEO*

J. D. Hietala, *GIAC, Senior SANS Analyst Principal, Compliance Research Group*

Dave Shackleford, *SANS GIAC Technical Director; Chief Security Officer, Configuresoft; Director, Configuresoft's Center for Policy and Compliance*

Jerry Shenk, *GIAC, Senior SANS Analyst; Senior Security Analyst, D&E Communications*

Introduction

Critical business systems and their associated technologies are typically held to performance benchmarks. In the security space, benchmarks of speed, capacity and accuracy are common for encryption, packet inspection, assessment, alerting and other critical protection technologies. But how do you set benchmarks for a tool based on collection, normalization and correlation of security events from multiple logging devices? And how do you apply these benchmarks to today's diverse network environments?

This is the problem with benchmarking Security Information Event Management (SIEM) systems, which collect security events from one to thousands of devices, each with its own different log data format. If we take every conceivable environment into consideration, it is impossible to benchmark SIEM systems. We can, however, set one baseline environment against which to benchmark and then include equations so that organizations can extrapolate their own benchmark requirements. That is the approach of this paper.

Consider that network and application firewalls, network and host Intrusion Detection/Prevention (IDS/IPS), access controls, sniffers, and Unified Threat Management systems (UTM)—all log security events that must be monitored. Every switch, router, load balancer, operating system, server, badge reader, custom or legacy application, and many other IT systems across the enterprise, produce logs of security events, along with every new system to follow (such as virtualization). Most have their own log expression formats. Some systems, like legacy applications, don't produce logs at all.

First we must determine what is important. Do we need all log data from every critical system in order to perform security, response, and audit? Will we need all that data at lightning speed? (Most likely, we will not.) How much data can the network and collection tool actually handle under load? What is the threshold before networks bottleneck and/or the SIEM is rendered unusable, not unlike a denial of service (DOS)? These are variables that every organization must consider as they hold SIEM to standards that best suit their operational goals.

Why is benchmarking SIEM important? According to the National Institute of Standards (NIST), SIEM software is a relatively new type of centralized logging software compared to syslog. Our *SANS Log Management Survey*[1] shows 51 percent of respondents ranked collecting logs as their most critical challenge – and collecting logs is a basic feature a SIEM system can provide. Further, a recent *NetworkWorld* article[2] explains how different SIEM products typically integrate well with selected logging tools, but not with all tools. This is due to the disparity between logging and reporting formats from different systems. There is an effort under way to standardize logs through Mitre's Common Event Expression (CEE) standard event log language.[3] But until all logs look alike, normalization is an important SIEM benchmark, which is measured in events per second (EPS).

Event performance characteristics provide a metric against which most enterprises can judge a SIEM system. The true value of a SIEM platform, however, will be in terms of Mean Time To Remediate (MTTR) or other metrics that can show the ability of rapid incident response to mitigate risk and minimize operational and financial impact. In our second set of benchmarks for storage and analysis, we have addressed the ability of SIEM to react within a reasonable MTTR rate to incidents that require automatic or manual intervention.

Because this document is a benchmark, it does not cover the important requirements that cannot be benchmarked, such as requirements for integration with existing systems (agent vs. agent-less, transport mechanism, ports and protocols, interface with change control, usability of user interface, storage type, integration with physical security systems, etc.). Other requirements that organizations should consider but aren't benchmarked include the ability to process connection-specific flow data from network elements, which can be used to further enhance forensic and root-cause analysis. Other features, such as the ability to learn from new events, make recommendations and store them locally, and filter out incoming events from known infected devices that have been sent to remediation, are also important features that should be considered, but are not benchmarked here. Variety and type of reports available, report customization features, role-based policy management and workflow management are more features to consider as they apply to an individual organization's needs but are not included in this benchmark. In addition, organizations should look at a SIEM tool's overall history of false-positives, something that can be benchmarked, but is not within the scope of this paper. In place of false positives, Table 2 focuses on accuracy rates within applicable categories.

These and other considerations are included in the following equations, sample EPS baseline for a medium-sized enterprise, and benchmarks that can be applied to storage and analysis. As appendices, we've included a device map for our sample network and a calculation worksheet for organizations to use in developing their own EPS benchmarks.

[1] www.sans.org/reading_room/analysts_program/LogMgt_June08.pdf
[2] www.networkworld.com/reviews/2008/063008-test-siem.html
[3] http://cee.mitre.org/cee.html

SIEM Benchmarking Process

The matrices that follow are designed as guidelines to assist readers in setting their own benchmark requirements for SIEM system testing. While this is a benchmark checklist, readers must remember that benchmarking, itself, is governed by variables specific to each organization. For a real-life example, consider an article in eSecurity Planet, in which Aurora Health in Michigan estimated that they produced 5,000–10,000 EPS, depending upon the time of day.[4] We assume that means during the normal ebb and flow of network traffic. What would that load look like if it were under attack? How many security events would an incident, such as a virus outbreak on one, two or three subnets, produce?

An organization also needs to consider their devices. For example, a Nokia high-availability firewall is capable of handling more than 100,000 connections per second, each of which could theoretically create a security event log. This single device would seem to imply a need for 100,000 minimum EPS just for firewall logs. However, research shows that SIEM products typically handle 10,000–15,000 EPS per collector.

Common sense tells us that we should be able to handle as many events as ALL our devices could simultaneously produce as a result of a security incident. But that isn't a likely scenario, nor is it practical or necessary. Aside from the argument that no realistic scenario would involve all devices sending maximum EPS, so many events at once would create bottlenecks on the network and overload and render the SIEM collectors useless. So, it is critical to create a methodology for prioritizing event relevance during times of load so that even during a significant incident, critical event data is getting through, while ancillary events are temporarily filtered.

Speed of hardware, NICs (network interface cards), operating systems, logging configurations, network bandwidth, load balancing and many other factors must also go into benchmark requirements. One may have two identical server environments with two very different EPS requirements due to any or all of these and other variables. With consideration of these variables, EPS can be established for normal and peak usage times. We developed the equations included here, therefore, to determine Peak Events (PE) per second and to establish normal usage by exchanging the PE_x for NE_x (Normal Events per second).

List all of the devices in the environment expected to report to the SIEM. Be sure to consider any planned changes, such as adding new equipment, consolidating devices, or removing end of life equipment.

[4] www.esecurityplanet.com/prodser/article.php/3453311/Using-SIM-Software-to-Deal-with-Security-Overload.htm

First, determine the PE (or NE) for each device with these steps:

1. Carefully select only the security events intended to be collected by the SIEM. Make sure those are the only events included in the sample being used for the formula.
2. Select reasonable time frames of known activity: Normal and Peak (under attack, if possible). This may be any period from minutes to days. A longer period of time, such as a minimum of 90 days, will give a more accurate average, especially for "normal" activity. Total the number of Normal or Peak events during the chosen period. (It will also be helpful to consider computing a "low" activity set of numbers, because fewer events may be interesting as well.)
3. Determine the number of seconds within the time frame selected.
4. Divide the number of events by the number of seconds to determine PE or NE for the selected device.

Formula 1:

$$\frac{\text{\# of Security Events}}{\text{Time Period in Seconds}} = \text{EPS}$$

The resulting EPS is the PE or NE depending upon whether we began with peak activity or normal activity. Once we have completed this computation for every device needing security information event management, we can insert the resulting numbers in the formula below to determine Normal EPS and Peak EPS totals for a benchmark requirement.

Formula 2:

1. In your production environment determine the peak number of security events (PEx) created by each device that requires logging using Formula1. (If you have identical devices with identical hardware, configurations, load, traffic, etc., you may use this formula to avoid having to determine PE for every device):

$$[\text{PEx (\# of identical devices)}]$$

2. Sum all PE numbers to come up with a grand total for your environment
3. Add at least 10% to the Sum for headroom and another 10% for growth.

So, the resulting formula looks like this:

Step 1: $(PE_1+PE_2+PE_3...+ (PE_4 \times D_4) + (PE_5 \times D_5)...) = SUM_1$ [baseline PE]

Step 2: $SUM_1 + (SUM_1 \times 10\%) = SUM_2$ [adds 10% headroom]

Step 3: $SUM_2 + (SUM_2 \times 10\%) =$ Total PE benchmark requirement [adds 10% growth potential]

Once these computations are complete, the resulting Peak EPS set of numbers will reflect that grand, but impractical, peak total mentioned above. Again, it is unlikely that all devices will ever simultaneously produce log events at maximum rate. Seek consultation from SMEs and the system engineers provided by the vendor in order to establish a realistic Peak EPS that the SIEM system must be able to handle, then set filters for getting required event information through to SIEM analysis, should an overload occur.

We have used these equations to evaluate a hypothetical mid-market network with a set number of devices. If readers have a similar infrastructure, similar rates may apply. If the organization is different, the benchmark can be adjusted to fit organizational infrastructures using our equations.

The Baseline Network

A mid-sized organization is defined as having 500–1000 users, according to a December guide by Gartner, Inc., titled "Gartner's New SMB Segmentation and Methodology." Gartner Principal Analyst Adam Hils, together with a team of Gartner analysts, helped us determine that a 750–1000 user organization is a reasonable base point for our benchmark. As Hils puts it, this number represents some geo and technical diversity found in large enterprises without being too complex to scope and benchmark. With Gartner's advice, we set our hypothetical organization to have 750 employees, 750 user end points, five offices, six subnets, five databases, and a central data center. Each subnet will have an IPS, a switch and gateway/router. The data center has four firewalls and a VPN. (See the matrix below and Appendix A, "Baseline Network Device Map," for more details.)

Once the topography is defined, the next stage is to average EPS collected from these devices during normal and peak periods. Remember that demanding all log data at the highest speed 24x7 could, in itself, become problematic, causing a potential DOS situation with network or SIEM system overload. So realistic speeds based on networking and SIEM product restrictions must also be considered in the baseline.

Protocols and data sources present other variables considered determining average and peak load requirements. In terms of effect on EPS rates, our experience is that systems using UDP can generate more events more quickly, but this creates a higher load for the management tool, which actually slows collection and correlation when compared to TCP. One of our reviewing analysts has seen UDP packets dropped at 3,000 EPS, while TCP could maintain a 100,000 EPS load. It's also been our experience that use of both protocols in single environment.

Table 1, "Baseline Network Device EPS Averages," provides a breakdown of Average, Peak and Averaged Peak EPS for different systems logs are collected from. Each total below is the result of device quantity (column 1) x EPS calculated for the device. For example, 0.60 Average EPS for Cisco Gateway/Routers has already been multiplied by the quantity of 7 devices. So the EPS per single device is not displayed in the matrix, except when the quantity is 1.

To calculate Average Peak EPS, we determined two subnets under attack, with affected devices sending 80 percent of their EPS capacity to the SIEM. These numbers are by no means scientific. But they do represent research against product information (number of events devices are capable of producing), other research, and the consensus of expert SANS Analysts contributing to this paper.

Table 1: Baseline Network Device EPS Averages

Qty	Type	Description	Avg EPS	Total Peak EPS	Average Peak EPS
750	Employees/Endpoints (Windows XP)	Desktops & laptops at 5 locations	Included at domain servers	Included at domain servers	Included at domain servers
7	Cisco Catalyst Switches	One at each location, one in DMZ and one in the Trusted network	5.09	51.88	26.35
7	Cisco Gateway/Routers	One at each location	0.60	380.50	154.20
5	Windows 2003 Domain Servers	One at each location	40.00	404.38	121.75
3	Windows 2003 Application Servers	In high availability cluster at data center	1.38	460.14	230.07
3	MS SQL Database Servers running on Windows 2003 Server	High availability cluster at data center	1.83	654.90	327.45
6	Microsoft Exchange Servers	One at each location with two (cluster) at the data center	3.24	1,121.50	448.60
3	MS IIS Web Servers on Windows 2003	High availability cluster at data center	1.17	2,235.10	1,117.55
2	Windows DNS Servers	At data center – failover	0.72	110.80	110.80
2	Linux Legacy Application Servers	At data center	0.12	43.60	21.80
1	Linux MySQL Database Server	One in Trusted network for legacy application	0.12	21.80	21.80
7	NitroGuard IPS	One at each location, one in DMZ and one in the Trusted network	40.53	5,627.82	1,607.95
1	Netscreen Firewall	Netscreen facing the Internet	0.58	2,414.00	2,414.00
3	Cisco Pix Firewalls	Between the data center and the other four sites, in front of Trusted network, between Trusted and the DMZ	39.00	1,734.00	1,178.00
1	Cisco VPN Concentrator	Located at data center Facing the Internet	0.83	69.45	69.45
1	Squid Proxy	Located at data center	14.58	269.03	269.03
		Totals:	**149.79**	**15,598.90**	**8,118.80**

A single security incident, such as a quickly replicating worm in a subnet, may fire off thousands of events per second from the firewall, IPS, router/switch, servers, and other infrastructure at a single gateway. What if another subnet falls victim and the EPS are at peak in two subnets? Using our baseline, such a scenario with two infected subnets representing 250 infected end points could theoretically produce 8,119 EPS.

We used this as our Average Peak EPS baseline because this midline number is more representative of a serious attack on an organization of this size. In this scenario, we still have event information coming from servers and applications not directly under attack, but there is potential impact to those devices. It is important, therefore, that these normal logs, which are useful in analysis and automatic or manual reaction, continue to be collected as needed.

SIEM Storage and Analysis

Now that we have said so much about EPS, it is important to note that no one ever analyzes a single second's worth of data. An EPS rating is simply designed as a guideline to be used for evaluation, planning and comparison. When designing a SIEM system, one must also consider the volume of data that may be analyzed for a single incident. If an organization collects an average of 20,000 EPS over eight hours of an ongoing incident, that will require sorting and analysis of 576,000,000 data records. Using a 300 byte average size, that amounts to 172.8 gigabytes of data. This consideration will help put into perspective some reporting and analysis baselines set in the below table. Remember that some incidents may last for extended periods of time, perhaps tapering off, then spiking in activity at different points during the attack.

While simple event performance characteristics provide a metric against which most enterprises can judge a SIEM, as mentioned earlier, the ultimate value of a well-deployed SIEM platform will be in terms of MTTR (Mean Time To Remediate) or other metrics that can equate rapid incident response to improved business continuity and minimal operational/fiscal impact.

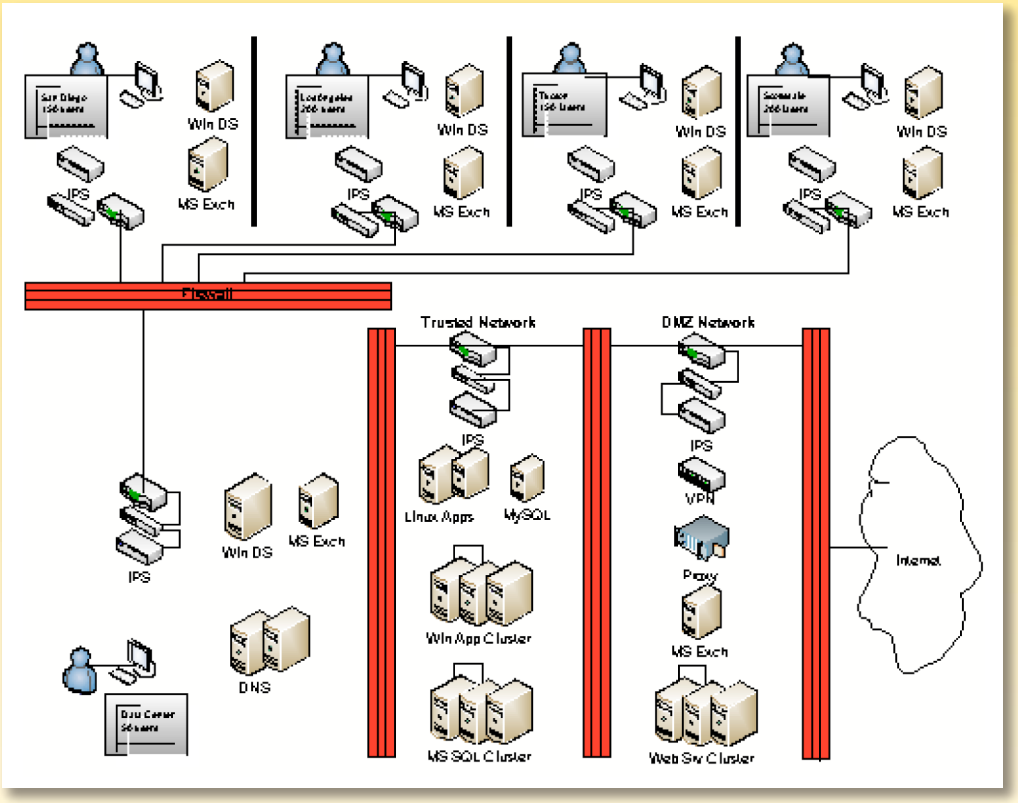
It should be noted in this section, as well, that event storage may refer to multiple data facilities within the SIEM deployment model. There is a local event database, used to perform active investigations and forensic analysis against recent activities; long-term storage, used as an archive of summarized event information that is no longer granular enough for comprehensive forensics; and read/only and encrypted raw log storage, used to preserve the original event for forensic analysis and nonrepudiation—guaranteeing chain of custody for regulatory compliance.

Table 2: Storage and Analysis Benchmarks

Feature	Benchmark	Settings	Explanation
Storage			
Storage format/ Database interface: - Native to SIEM - Database - Flat files - SAN storage	SIEM is capable of storing entire raw logs in multiple formats without hindering collection speeds.	Complete log data can be viewed in native SIEM system, SQL, DB2, Oracle, MySQL, or other appropriate databases, if necessary. Collection rates are scalable as specified.	Storage database must be acceptable to organization: Collection rates should not be unreasonably delayed during peaks, no matter what storage format. Some SIEM products have methods of storage that use flat files with indexing done in databases that may provide faster results than traditional databases.
Compression of log data	Compression ratio	Test the difference between original data size in bytes and compressed size. Use a compression target of an 8:1 ratio. Collection rates as specified.	Log data should be compressed, with minimal impact to SIEM system, for better storage. Actual compression may be skewed by the fact that normalized data can take up more space that raw data. In those cases, an 8:1 ratio may not be achievable.
Encryption (online/archive)	SIEM data is encrypted with minimal impact on access and encryption/decryption speeds.	Online and archived data should not be readable without proper authorization. The algorithm meets industry standards (AES, 3DES, etc.). No data is lost when encryption is used. Encrypt/decrypt rates meet specifications.	Logs tell attackers a lot about a network. They also contain personal and/or proprietary information in chains and fragments. If encryption is not possible, the vendor must provide proof that the logs are adequately protected.
Long-term storage	Should scale to store raw and normalized data for 90 days to 7 years, based on regulatory and business requirements.	Using only normal EPS from our baseline network, 2.4 billion events will be collected in 6 months, 33.1 billion collected in 7 years. If we use 300B as the average message size, 7 years of data will require 9.9TB of space (if there are no incidents).	The requirement for space will vary depending upon the number of incidents during the time period stored. Adequate overhead must be calculated. Also consider anticipated infrastructure changes in the company business plans.
Real-time analysis	A SIEM system must allow real-time monitoring of events.	Validate that data are displayed within 1 minute of actual event, within 2–3 minutes under load. In our baseline, analysis would have to scale to 8,119 EPS during peak. Set rate of accuracy at 90%. (In our baseline that's 812 incorrect events per second.)	During incidents, quick analysis of data is important. Monitoring thresholds should align with maximum EPS. Accuracy rates have been improved by correlation, heuristics, behavior-based anomaly detection, etc. But accuracy is still a large problem because it is subject to variances that can't be anticipated.
Real-time management	A SIEM system must allow management in near real-time.	Validate reaction time to events for accuracy (90%) and speed (within 60 seconds of event correlation).	A SIEM's value is notification of events and incidents quickly enough to react and mitigate risk.
Historic analysis	Expect same analysis capability as with current data within 15 minutes or less.	Correlation and conclusions drawn in 15 minutes or less with same level of accuracy as in nonarchived systems. Using our baseline 7-year storage calculations, this means correlating against up to 9.9TB or more of data.	When setting this benchmark, note that not all historic data needs be drilled down at once. Using specified queries reduces the search volume considerably.
Normalize log data	Approximately 90 % of systems requiring management can be normalized with out-of-box tools at expected collection rates.	Verify that log data collected is transformed to a standardized format within average peak EPS. In our scenario, this would be 8,119 EPS or better.	Logs have disparate formats. Normalization allows us to compare data from the systems.
Correlation	90% accuracy. Speed to match avg peak baseline EPS.	Correlation of live events is achieved with 90% accuracy. Collection rates meet required EPS at normal and peak times. In our baseline this equates to 8,119 EPS or better.	If analysis cannot be accomplished in real-time, then reaction to an event may come too late
Taking action: Quarantine, block, route, and control accounts, services, configurations, processes and privileges.	MTTR: Verify that SIEM system is able to take actions as appropriate to the event within a specified timeframe and with appropriate accuracy.	MTTR = 3-5 minutes with a 99% accuracy rate based on organization requirements. These are drawn against our baseline environment in a Peak Average of 8,119 EPS.	SIEM systems should not be trusted to take action with perfect accuracy. Suggested actions should not be in categories prone to false positives. Use these features only when you are confident that they are accurate in the applied environment.
Failover	In the case of system outage, SIEM and its reliably fails over to the backup system.	Failover is instantaneous. During a failover event, data is continually collected and accessible within seconds.	A secondary system must continue to collect data. Access to secondary system must be nearly immediate to troubleshoot the loss event.

Appendix A: Baseline Network Device Map

This network map is the diagram for our sample network. Traffic flow, points for collecting and/or forwarding event data, and throttle points were all considered in setting the benchmark baseline in Table 1.



Appendix B: EPS Calculation Worksheet

Use this list along with your and your peers' experience and other references as resources to set benchmarks for your infrastructure. The Avg., Peak, and Avg. Peak columns are intentionally left blank for you to fill in your own benchmark numbers:

Feature	Benchmark	Settings	Explanation	Avg.	Peak	Avg. Peak
System logs collected	Relevant and critical logs are collected. Should be able to handle peak threshold	Performance/speed of collection, normalization.	Should be able to intake at speed of log traffic from subnets and central network.			
Network devices: - Firewalls - VPNs/SSL - IAM –Switches/ routers - Web proxies)	Source, destination, calls, connections, access, traffic, and other security-related log data can be collected and normalized at specified rate.	Log/Event data collected to SIEM at peak specified EPS per device without dropping events.	Should be able to intake at speed of log traffic from subnets during normal and peak traffic (during events).			
End Points: - Servers - O/S's - Security - NAC - NICs	Collection from end point security-related data at specified EPS.	Log/Event data collected to SIEM at peak specified EPS per device without dropping events.	End point security information can indicate where to remediate and drop event data from systems in remediation.			
Commercial apps: - HR/workflow - Business critical - Contain PI data	Security-related data from commercial applications is collected as needed.	Log/Event data collected to SIEM at peak specified EPS per device without dropping events.	Not all applications can feed into every SIEM. Collection rates can be high for some applications.			
- Custom apps - Legacy apps - Mainframe - Midrange	Security-related data from custom/legacy applications and systems are collected as needed.	RACF, Top Secret, or other security system events appear in SIEM. Collection rates as specified.	Custom/legacy applications tend to have more vulnerabilities and less patching than other systems.			
- Databases - Third party DB - Monitoring tools/ database session logs	Access logs and other security-related data from databases collected.	Compatibility with systems needing coverage. Collection rates as specified.	DB2, SQL, Oracle, MySQL, etc., are central to the business and related events and should never be dropped.			
Backup systems	Backup systems' log data is collected.	Collection rates as specified.	Backups may not be accessed that often, but this and other security data should be available when needed.			
Virtual system logs (applies to any of the above systems that are virtualized)	Virtual machines (VMs) and VM managers (VMMs) are held to the same performance and monitoring standards as physical devices.	Coverage and collection rates as specified.	Virtual systems and managers (e.g.,. HyperV) require the same monitoring as physical systems.			

About the Author

J. Michael Butler, GCFA CISA GSEC, is an Information Security Consultant with LPS, a leading provider of technical services to the mortgage industry. Butler's responsibilities have included computer forensics, information security policies (aligned to ISO and addressing federal and state disclosure laws), enterprise security incident management planning, internal auditing of information systems and infrastructure, service delivery, and distributed systems support. He has also been involved in authoring SANS security training courseware, position papers, articles, and blogs. Butler has over 27 years of experience in the computer industry.

SANS would like to thank this paper's sponsor: