Global Information Assurance Certification Paper

Logging and Reporting : A view from the top

Rick Hislop

August 17, 2003
GSEC Practical Assignment Version 1.4b

Abstract

In today's age of technology, we can find ourselves overwhelmed by the amount of day-to-day information that we're required to process. There's the company e-mails, subscription newsletters, online presentations and seminars, client/vendor requests, and the list goes on – each requiring a portion of our attention and time to deal with. Why would we want to compound the issue by retrieving a seemingly incomprehensible amount of data from the devices, applications, and services that we use on a daily basis?

> *You can't know where you're going unless you know where you've been and where you are.*

There are a number of compelling reasons why we need to process this data into useful information. We have certain legal obligations to securely record and store the data we collect and for how long – this depends on the type of data and the specific legal requirements in your area and situation. As a company, we need to know what's being used and by whom, how it's being used, and how much of it is being used so that we can determine its financial viability, security, suitability and usability.

One of the largest barriers in implementing a proper logging and reporting infrastructure is dealing with the vast amount of data from so many different sources. In an ideal world, all logging would be done in a consistent and orderly manner – this is simply not the case. For every device, application, and service, there is a unique way of recording the data it collects. The challenge is to find a way to manipulate the data so that it can be reported on in an automated fashion so that you can extract the information that you need. Unfortunately, as of the writing of this paper, there is no super-product that will turn the sea of data into useful information.

Even though we can't buy off-the-shelf software to solve this particular puzzle in its entirety, we can use existing technology to get the most out of the data available to us to provide us with the information we need to know. This paper hopes to update and expand on the excellent work done by Stewart Allen : "Importance of Understanding Logs from an Information Security Standpoint".[1]

[1] Stewart Allen : "Importance of Understanding Logs from an Information Security Standpoint"

Why Log … Why Report?

Definitions [2]

data

<data, data processing, jargon> (Or "raw data")
Numbers, characters, images, or other method of recording, in a form which can be assessed by a human or (especially) input into a computer, stored and processed there, or transmitted on some digital channel. Computers nearly always represent data in binary.

Data on its own has no meaning, only when interpreted by some kind of data processing system does it take on meaning and become information.

People or computers can find patterns in data to perceive information, and information can be used to enhance knowledge. Since knowledge is prerequisite to wisdom, we always want more data and information. But, as modern societies verge on information overload, we especially need better ways to find patterns.

information

1: a message received and understood that reduces the recipient's uncertainty [syn: info]
2: a collection of facts from which conclusions may be drawn; "statistical data" [syn: data]
3: knowledge acquired through study or experience or instruction

To coin the phrase “Knowledge is Power” is not an understatement – we *need* to know. Having a solid data collection and reporting infrastructure in place can enable us to determine the current and past state (depending on how long you’ve been logging data) of your network, devices, applications, and services. Logging from a security perspective is of the utmost importance but that reason is somewhat myopic. Logging extends far beyond the security context and should be an integral part of business operations as a whole.

As the technology industry matures, greater reliance is being placed on the logging and reporting of data. Inadequate logging and reporting can negatively impact companies and in some cases, result in their untimely demise. Law courts, federal bodies, and even your own company require accurate data to be stored for specific periods of time and in a secure and accurate fashion. Do you have the required data stored on an employee that was terminated 3 months ago . . . a year ago? Can you determine if they accessed a sensitive resource or abused a corporate service? What about the last time your corporate web server was unavailable? Without logging, you will never be able to find out and that poses a substantial risk to your company or organization.

The logging and collection of data is only part of the equation. Logs simply record specific events and though you can manually review the logs to generate a specific piece of information, reviewing the logs on multiple machines and

[2] Dictionary.com

devices can be a monumental task. Reporting, in its truest sense, enables you to take the data from multiple sources and distill it into something meaningful – convert data into information.

Reporting has the ability to automate log review tasks that would normally take an inordinate amount of time and effort. If you can't get the information you need in a timely fashion, collecting the data in the first place is almost moot. Given today's tight budgets and limited resources, we need to leverage every tool we can.

Reporting provides us with a number of benefits. It can allow us to know how our resources are being used, who is using those resources and when, system outages and alerts, resource availability and performance, and security events. If a device is under-utilized, you could buy a smaller one and re-deploy the existing one to another area that needs it. Reporting can allow you to make the most out of your equipment and services – this translates to better savings as more efficient use is being made of your existing technology.

From a security perspective, reporting can provide insight into the risks facing your company from both internal and external threats. You would have the ability to find out how many people have been trying to compromise your corporate services, the type of traffic that your network has been passing, and attempts at accessing restricted resources such as files, mailboxes, and management applications. By correlating the information, you can determine if the attempt was malicious or not or use the information to further strengthen your security infrastructure and policies.

What can you log?

Logging is the collection of data from various applications, services, or devices – practically everything you use today either does or has the ability to generate some form of a log file. As imposing as that sounds, if you take stock of what you have and categorize them, you will be in a better position to deal with the number and diversity of logs you collect.

Most companies look at the 'well-known' logs generated by Syslog or Microsoft Event Viewer. Both of these services have the ability to collect data from a number of different sources – still, most focus on logging a finite amount of data that is available to them.

To give you an idea of what you can log, below is a partial list of possible sources.

- Devices (stand-alone and network attached)
  - Switches, routers, wireless access points (WAP's)
  - PBX switches and voicemail systems

    - Appliances such as anti-virus, firewalls, load-balancers, N/HIDS and content filters
    - Radius, RAS, and VPN devices
    - Printers and fax machines
    - Security systems and card locks
    - Environmental systems
- Applications
    - Anti-virus
    - Web & E-mail Content filters
    - Firewall and VPN
    - H/NIDS (host/network based Intrusion Detection System)
    - Web / FTP / Mail / Proxy / Directory / DNS / Database servers
    - Device and software inventory systems
    - Backup applications
    - CRM, ERP, and document management systems
- Logging Services
    - Syslog and variations
    - Microsoft Event Viewer
    - SNMP
    - Custom/Proprietary logging implementations

Logging services will collect data from many of the sources listed above but definitely not all. It is important to understand what you want to know and from what device, service, or application. Some logging services have the ability to combine data sources thus minimizing the management of separate logging solutions. By taking an inventory of your desired data sources, you are helping to prepare the foundation for your logging and reporting infrastructure.

One of the best ways to determine your desired data sources is to create a framework for system data and be able to prioritize it. A case study done by Dan Rathbun[3] illustrates a technique used to prioritize data according to its purpose:

Facility Name	**Purpose**
Local0	Perimeter Devices (various devices)
Local1	Unassigned
Local2	Interior networked servers
Local3	Peripheral devices (printers, scanners)
Local4	Workstations (future use)
…	…

Though the original purpose was for use by Kiwi Syslog, the concept can be applied to your logging framework. Organizing your data sources into categories will allow you to plan, maintain, and expand your logging infrastructure.

[3] Dan Rathbun: "Case Study: Using Syslog in a Microsoft & Cisco Environment"

What's in a log?

The contents of log files are just as diverse as the devices, applications, and services they come from. Some log files follow an industry standard (W3C[4] or NCSA) while others are simply in the form that the vendor felt was best to make available to you. Because of this, the actual contents of the log files will vary.

Log files can contain event specific information as it pertains to the device, application, or service. Some examples of the events are:

- Security related (authorized / unauthorized access, authentication passes / failures)
- Access related (file, application, service)
- General (device or application specific events)
- Environmental and system health
- Device usage (bandwidth, traffic errors and alerts)

Event data may contain some or all of the following:

- Date and time of the event
- Device, application, service or UserID associated with the event
- Event ID or severity code
- Description of the event

Some examples of log file entries are shown below:

Windows Event Log

7/17/2003 11:42:25 AM atapi Error None 9 N/A MONSTAW2K
The device, \Device\Ide\IdePort1, did not respond within the timeout period.
7/17/2003 11:41:24 AM SNMP Information None 1001 N/A
MONSTAW2K The SNMP Service has started successfully.
7/17/2003 11:41:17 AM VMnetx Information None 34 N/A
MONSTAW2K (\Device\Vmxctr) Connecting VMnet1.
7/17/2003 11:40:51 AM Service Control Manager Error None 7000
N/A MONSTAW2K The MySql service failed to start due to the following error: The system cannot find the path specified.

Syslog

```
Aug  5 13:38:45 proxysys unix: stdin is </pci@1f,4000/ebus@1/se@14,400000:a
> major <20> minor <0>
Aug  5 13:38:45 proxysys unix: stdout is </pci@1f,4000/ebus@1/se@14,400000:
a> major <20> minor <0>
Aug  5 13:38:45 proxysys unix: cpu1: SUNW,UltraSPARC-II (upaid 1 impl 0x11
ver 0xa0 clock 450 MHz)
Aug  5 13:38:45 proxysys unix: cpu2: SUNW,UltraSPARC-II (upaid 2 impl 0x11
ver 0xa0 clock 450 MHz)
Aug  5 13:38:45 proxysys unix: cpu 2 initialization complete - online
```

[4] W3C "Extended Log Format"

Time plays an important role in logging for without it, you'll have a very difficult time correlating events across multiple systems. According to NTP.ORG, "NTP [or Network Time Protocol] is a protocol designed to synchronize the clocks of computers over a network"[5]. Network time servers provide an accurate time reference for your servers and network devices. See NTP.ORG for a list of publicly available time servers and for information on how to implement NTP within your organization.

Types of reporting

Reporting allows you to extract data from the logs and turn it into useful information. Raw data is cumbersome – it can't summarize how many events occurred in a specific time period nor correlate one event against another. We need to be able take data from single or multiple sources and compile that into concise, readable, and accurate information.

There are two main types of reporting practices: Real-Time and Historical. With Real-Time reporting, you are typically focusing on data that has occurred within a period of seconds, minutes, or even hours. This type of information can give you immediate notification of when things go awry or provide you with a current snapshot of your network health. Reports of this nature are usually more granular and the logs from which the reports are drawn can grow very quickly.

Historical reports, on the other hand, are great for summarizing or trending data. The data, whether summarized or not, can be from a day, a week, a month, or even older. You have the ability to determine trends and create summaries. By having access to the necessary data, you could see how your network bandwidth has been affected by the implementation of new ACL's on a core router the week before or who the Top 20 users were of your RAS server for the previous month. This brings up an important question.

How long do you hold onto the data for? Data retention policies vary widely from organization to organization and department to department and even system to system. The single best answer to this question is to consult your legal department and Chief Security Officer or other person responsible for the data. A balance needs to be struck amongst the legal, auditing, forensic, and fiscal aspects of the storage of data.

An excerpt from the Record Retention Policy of Access Advancing Education Technologies illustrates a framework for data retention. [6] For a more complete look at their work, please visit their website.

[5] NTP.ORG
[6] ACCESS Advancing Education Technologies "Record Retention Policy"

Series Number	Classification	Sub-Classification	Title	Description	Retention Period
GIT-OP-06	Information Technology	Computer Operations and Technical Support	System Backup Files	Copies of master files or databases, application software, logs, directories, and other records needed to restore a system in case of a disaster or inadvertent destruction.	Retain for 3 system backup cycles, then destroy.
GIT-OP-07	Information Technology	Computer Operations and Technical Support	System Users Access Records	Electronic or textual records created to control or monitor individual access to a system and its data created for security purposes, including but not limited to user account records, security logs, and password files.	Retain until no longer of adminsitrative value to agency, then destroy.

There are a number of tools available to create reports but as previously mentioned, there aren't any that "do it all". Some vendor products focus exclusively on a particular application, service, or device while others are more general in their approach and have the capability to be more versatile though for a substantial cost. Freeware products have similar capabilities as the vendor products but they may not provide support, require adept skills to install, configure, and maintain, and usually lack the "polished look and feel" that you would normally get from an off-the-shelf product.

The following list is only a portion of what is available but it should help you to determine the types of products that can be implemented for various reporting requirements. Make sure that you understand what functionality the products are capable of before purchasing them.

Vendor
- WebTrends
- NetIQ Log Manager
- Crystal Reports
- SolarWinds / Orion
- EventTracker for Windows

Freeware
- Big Brother
- Nagios
- SWATCH
- Custom PERL scripts

To find more, see your regular search engine.

Reporting Challenges

There are a number of challenges in reporting on your collected data. The main challenges you will be faced with are the different formats the data is collected in from various sources, storage and backup requirements, hardware availability, and budgetary constraints.

One of the biggest issues that you will encounter is that the data that you collect is stored in different formats. A reporting software package may be suitable for reporting on web server usage but may be completely unsuitable for reporting on logs created by your network router. Data transformation is what will allow you to consolidate your logs into a central repository such as a database. By having a consistent view of the data, running reports will be that much easier.

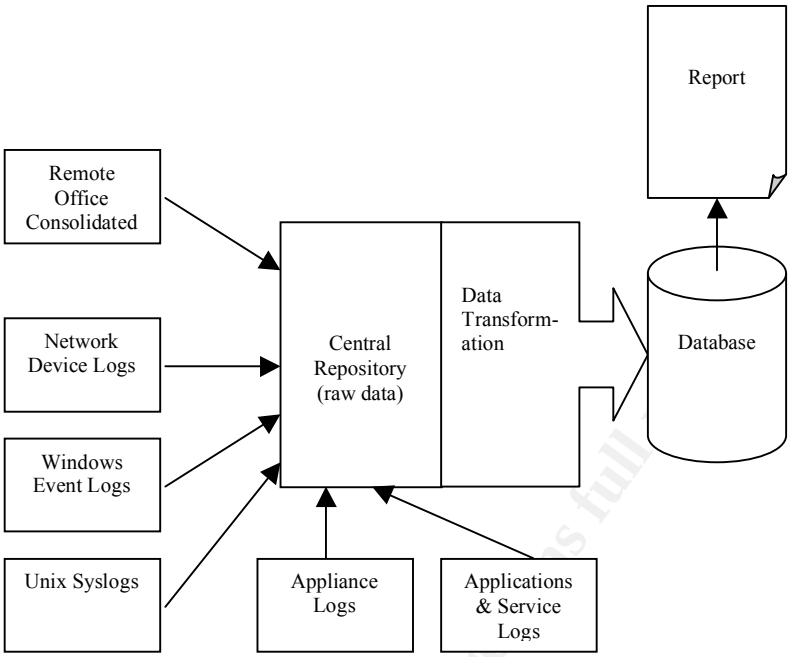
Another barrier to be overcome is the storage of data and the computing power to process it. Logs can grow to enormous sizes and can require a large investment in data storage equipment such as RAID's, NAS's, or even SAN's – the amount of data collected and your retention period (before it gets backed up to offline storage) will play an important role in determining your required storage capacity. Given the ever-changing nature of the logged data, you may need to expand your backup infrastructure to accommodate the volume of data – only full backups can be performed.

Budgetary constraints can hinder your efforts in establishing your reporting infrastructure. Costs can easily skyrocket with the purchase of new hardware, OS and software licenses, and skilled labor. The argument for allocating the necessary time and funds towards your reporting infrastructure are based on how much they ("people in control of the funds") want to know and how soon they need to know it.

Logging Infrastructure

When creating your logging and reporting infrastructure, there are a number of factors that will determine its design. Factors to consider include budget, hardware and software availability, skill-set availability, existing network infrastructure, security considerations, and what you will be logging.

The following diagram represents a basic set up for your logging infrastructure. You may have many more devices to include and surplus hardware to use. The goal is to illustrate the main functional concepts of setting up your logging infrastructure.



Microsoft has a good document outlining how to handle data, data transformation, and analysis methodology. See Microsoft's "Log Capture and Analysis"[7] for more information.

All logging devices, services, and applications submit their data to a central repository. They can submit using a number of different means including, FTP, TFTP, secure FTP, NetBIOS/SMB shares, and the services' proprietary transmission capabilities. You should begin to build an appreciation for the amount of data that can be logged.

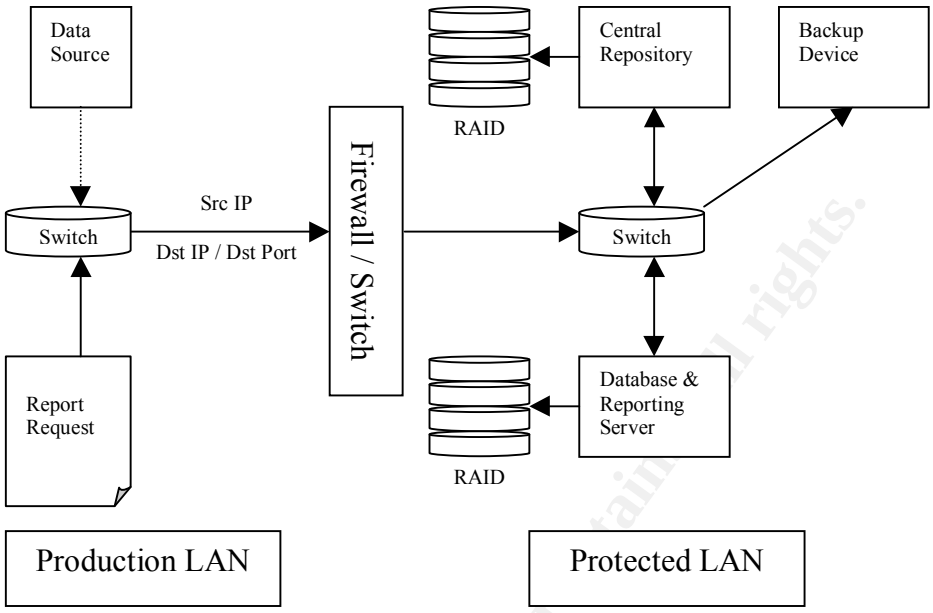
A data transformation component is what will take the data from all of the data sources and format it in such a way that it can be easily referenced in a database. Data transformation is necessary if you want to take data in different formats and convert it to a standardized format. This can be quite involved and a lot of people will overlook this in favor of using specialized software packages to process specific data types. The problem with this is that it requires multiple software packages that are usually quite expensive.

Once the data has been transformed, it is then dumped into a database where it can be accessed using a variety of reporting tools. The reporting software can be anything that can reference the database and present the information to you in a desirable form – printed, or online / web-based.

In understanding the functional requirements of your logging and reporting infrastructure, you will have a better idea of what is needed on the technical side.

[7] Microsoft "Log Capture and Analysis"

Below is a simple technical diagram highlighting some of the important aspects you should consider in implementing your logging infrastructure.



One of the most important aspects of your logging infrastructure is its security. The data reported by each device needs to be secure and unaltered. Should someone manage to compromise one of your devices, a complete and unaltered record of that should be stored in your central repository which itself will be highly secured. By placing your central repository behind a firewall or switch with very restrictive rules/ACL's, you can make it very difficult for your logged data to be compromised. Hackers are quite adept in their ability to cover their tracks by corrupting logged data on both the device and centralized data store. Another way some organizations have used to mitigate the risks of data corruption is to split the collection of data across multiple storage targets. This can be a fairly robust yet very expensive alternative.

The Central Repository and Database Server should be configured to use RAID or hybrid RAID for hardware fault tolerance and backups of the data should be taken regularly. Some organizations have even implemented server clustering and WORM (write once, read many) technologies for safeguarding their data and improving system performance.

Aside from the technological aspects of your logging and reporting infrastructure, you should also consider the physical security side. Physical access to these machines should be limited as well as console and remote access. Should one of your logged devices become compromised, your central repository will have the only reliable copy of that log. CERT has put together some best practices in building a secure logging infrastructure. [8]

[8] CERT.ORG "Manage logging and other data collection mechanisms."

Summary

Logging and reporting perform a key function within business. It gives us access to the information we need to determine how our devices, applications, and services are running, how they are being used and by whom. We have an accurate record that can be used for auditing or forensics as well as satisfy our legal obligations.

Even though it's so important, few organizations implement a comprehensive logging and reporting infrastructure. Dealing with the vast amounts of data and financial investments necessary can be quite overwhelming. The diversity in data formats can make consolidating logging and reporting a seemingly impossible task. By transforming the data into a consistent format, you will be able to centralize your data collection into a central repository / database and minimize the software necessary to report against it.

By creating a framework for your logging and reporting, you can begin to take control of your data. Understanding what you have and what you need to know are important steps in establishing a solid infrastructure that can be easier to create, maintain, and expand. A similar framework can also be used to organize your data retention policies across multiple locations, departments, and devices.

No one software package will allow you to collect and report on everything. Packages today are fairly specialized and if you want to buy your solutions, you will require a substantial investment in purchasing them. Freeware solutions exist, but they must be extended to report on data outside their normal scope and can require a significant amount of time and skill to implement. Despite the challenges, if you are able to plan, organize, and finance, you're logging and reporting infrastructure, you will be well on your way to knowing your organization.

References

[1] Allen, Stewart. "Importance of Understanding Logs from an Information Security Standpoint". October 5, 2001. URL: http://www.sans.org/rr/paper.php?id=200

[2] Dictionary.com http://dictionary.reference.com/

[3] Rathbun, Dan. "Case Study: Using Syslog in a Microsoft & Cisco Environment". June 27, 2003. URL: http://www.sans.org/rr/paper.php?id=1100

[4] W3C "Extended Log Format". URL: http://www.w3.org/TR/WD-logfile.html

[5] Ntp.org: Home of the Network Time Protocol
URL: www.ntp.org

[6] ACCESS "Data Retention Policy"
URL: http://www.access-k12.org/record_retention.htm

[7] Microsoft. "Log Capture and Analysis" URL:
http://www.microsoft.com/technet/treeview/default.asp?url=/technet/itsolutions/ecommerce/maintain/monitor/logcanda.asp

[8] CERT.ORG "Manage logging and other data collection mechanisms."
URL: http://www.cert.org/security-improvement/practices/p092.html

Kiwi Syslog Daemon
URL: www.kiwisyslog.com

RFC 3164 (rfc3164) "The BSD Syslog Protocol". August 2001
URL: http://www.faqs.org/rfcs/rfc3164.html

CERT.ORG. "Identify and enable Web-server-specific logging mechanisms"
URL: http://www.cert.org/security-improvement/practices/p077.html

Wrozek, Brian. "Electronic Data Retention Policy" August 15, 2001.
URL: http://www.sans.org/rr/paper.php?id=514

NetIQ Log Manager
URL: http://www.netiq.com/products/sm/log.asp

EventTracker for Windows
URL: http://www.eventlogmanager.com/

WebTrends Reporting Series
URL: http://www.netiq.com/products/wrc/default.asp

Crystal Reports
URL: http://www.crystaldecisions.com/products/crystalreports/default.asp

Nagios
URL: http://www.nagios.org/

Big Brother
URL: http://www.bb4.com/

SWATCH
URL: http://sourceforge.net/projects/swatch/

Last Updated: May 23rd, 2017

Upcoming Training

SANS Stockholm 2017	Stockholm, Sweden	May 29, 2017 - Jun 03, 2017	Live Event
SANS Houston 2017	Houston, TX	Jun 05, 2017 - Jun 10, 2017	Live Event
Security Operations Center Summit & Training	Washington, DC	Jun 05, 2017 - Jun 12, 2017	Live Event
Community SANS Ottawa SEC401	Ottawa, ON	Jun 05, 2017 - Jun 10, 2017	Community SANS
SANS San Francisco Summer 2017	San Francisco, CA	Jun 05, 2017 - Jun 10, 2017	Live Event
SANS Charlotte 2017	Charlotte, NC	Jun 12, 2017 - Jun 17, 2017	Live Event
SANS Rocky Mountain 2017 - SEC401: Security Essentials Bootcamp Style	Denver, CO	Jun 12, 2017 - Jun 17, 2017	vLive
Community SANS Portland SEC401	Portland, OR	Jun 12, 2017 - Jun 17, 2017	Community SANS
SANS Secure Europe 2017	Amsterdam, Netherlands	Jun 12, 2017 - Jun 20, 2017	Live Event
SANS Rocky Mountain 2017	Denver, CO	Jun 12, 2017 - Jun 17, 2017	Live Event
SANS Minneapolis 2017	Minneapolis, MN	Jun 19, 2017 - Jun 24, 2017	Live Event
SANS Columbia, MD 2017	Columbia, MD	Jun 26, 2017 - Jul 01, 2017	Live Event
SANS Cyber Defence Canberra 2017	Canberra, Australia	Jun 26, 2017 - Jul 08, 2017	Live Event
SANS Paris 2017	Paris, France	Jun 26, 2017 - Jul 01, 2017	Live Event
SANS London July 2017	London, United Kingdom	Jul 03, 2017 - Jul 08, 2017	Live Event
Cyber Defence Japan 2017	Tokyo, Japan	Jul 05, 2017 - Jul 15, 2017	Live Event
Community SANS Phoenix SEC401	Phoenix, AZ	Jul 10, 2017 - Jul 15, 2017	Community SANS
SANS Cyber Defence Singapore 2017	Singapore, Singapore	Jul 10, 2017 - Jul 15, 2017	Live Event
Community SANS Minneapolis SEC401	Minneapolis, MN	Jul 10, 2017 - Jul 15, 2017	Community SANS
SANS Los Angeles - Long Beach 2017	Long Beach, CA	Jul 10, 2017 - Jul 15, 2017	Live Event
SANS Munich Summer 2017	Munich, Germany	Jul 10, 2017 - Jul 15, 2017	Live Event
Mentor Session - SEC401	Macon, GA	Jul 12, 2017 - Aug 23, 2017	Mentor
Mentor Session - SEC401	Ventura, CA	Jul 12, 2017 - Sep 13, 2017	Mentor
Community SANS Atlanta SEC401	Atlanta, GA	Jul 17, 2017 - Jul 22, 2017	Community SANS
Community SANS Colorado Springs SEC401	Colorado Springs, CO	Jul 17, 2017 - Jul 22, 2017	Community SANS
SANSFIRE 2017	Washington, DC	Jul 22, 2017 - Jul 29, 2017	Live Event
SANSFIRE 2017 - SEC401: Security Essentials Bootcamp Style	Washington, DC	Jul 24, 2017 - Jul 29, 2017	vLive
Community SANS Charleston SEC401	Charleston, SC	Jul 24, 2017 - Jul 29, 2017	Community SANS
Community SANS Fort Lauderdale SEC401	Fort Lauderdale, FL	Jul 31, 2017 - Aug 05, 2017	Community SANS
SANS San Antonio 2017	San Antonio, TX	Aug 06, 2017 - Aug 11, 2017	Live Event
SANS Prague 2017	Prague, Czech Republic	Aug 07, 2017 - Aug 12, 2017	Live Event