



The NSF Cybersecurity Center of Excellence

## Security Log Analysis Training

*Vlad Grigorescu, Warren Raquel, Adam Slagell & Jeannette Dopheide*



2016 NSF Cybersecurity Summit  
August 16th, 2016



[Adam]

Thank people (doing something diff combining) & explain logos

Warren Raquel is a Senior Security Engineer at the National Center for Supercomputing Applications. His duties include security operations, incident response and security awareness for NCSA, Blue Waters and XSEDE. He has given talks and taught classes on Digital Forensics and Incident Response, two fields in which he has specialized in for the last decade.

Vlad Grigorescu is also on the Security team at the National Center for Supercomputing Applications, where he currently splits his time between Blue Waters operational security and Bro development. Over the past 10 years, he's focused on network monitoring tool development, malware analysis, security architecture, and penetration testing in a variety of roles at the University of Illinois, Carnegie Mellon University, and Broala.

## Center for Trustworthy Cyberinfrastructure

The mission of CTSC is to provide the NSF community with a coherent understanding of cybersecurity, its importance to computational science, and the resources to achieve and maintain an appropriate cybersecurity program.



[Adam]

Talk briefly about CTSC, disambiguate from CACR.

Activities: summit, webinar, vulnerability alerts, engagements, risk assessments

## The Bro Center of Expertise

Bro is a powerful network analysis framework used for security monitoring and network traffic analysis. The Bro Center of Expertise is a central point of contact for institutions funded by the National Science Foundation seeking advice on how best to use Bro at such institutions.



[Adam]

What is Bro, who uses Bro, and who supports it

Also promote the NSF center

# Security Log Analysis Tutorial Goals



- Today you can expect us to:
  - Take you thru the Log Analysis Lifecycle
  - Provide log analysis examples of real attacks with Bro.logs
  - Encourage interactive Q&A throughout
- You should take away
  - Ideas to improve your security logging & monitoring
  - Methods you can generalize to explore and connect events across logs

4

CTSC

[Adam]

Talk about the audience; not expecting security experts, but people who no Linux & comfortable on command line

No laptop needed: Interactive demos at points and chance to download files used in some examples, but no exercises per say

Really hit the point that it needs to be interactive to get the most out of it, and we are flexible on schedule.

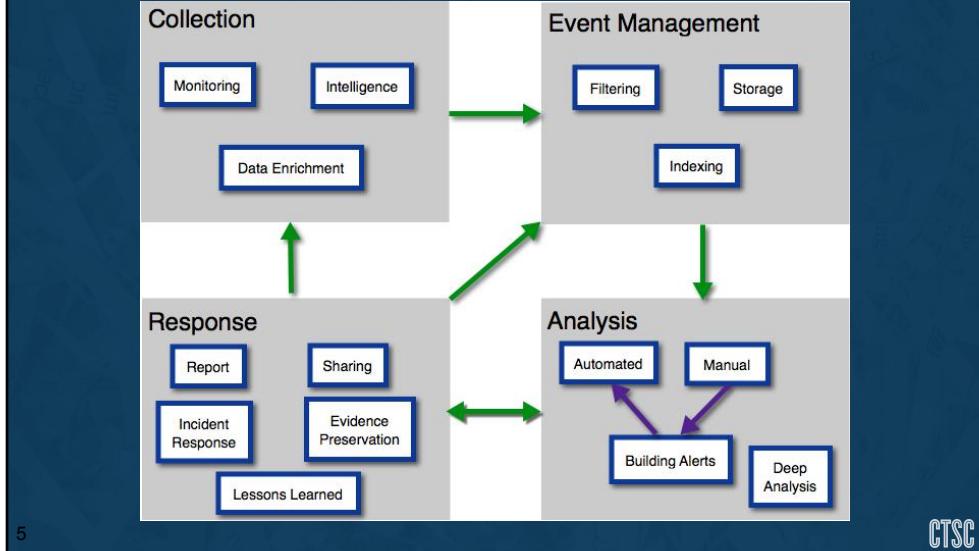
Time at the end for general Q&A with the experts (Warren & Vlad) on anything security related

image source:

<http://www.pdclipart.org/displayimage.php?album=search&cat=0&pos=4>



## Log analysis workflow



5

CTSC

[Adam]

Cycle: never ends, self reinforces (guides org. today)

Define Log: Any system or device generated record (includes intel feeds and enrichment like inventory DBs)

Filters & indexers make manageable (size can be huge to store & search)

No “right” way to organize, but should scale

Analysis (manual to auto)-> feeds into what to investigate

Response (investigation touches all cycles)



## Overview

- ***Collection***
  - Log sources
  - Intelligence
  - Data Enrichment
- Event Management
- Analysis
- Response



CTSC

6

[Warren]

Let's start with 'Collection'. We'll go over our various sources of data which include logs, intelligence and additional information that we'll call data enrichment which I'll explain more about when we get there.

# Collection Sources



- Logs

- System: Syslogs, Windows events, Host-based IDS, etc.
- Network: netflows, Bro, Snort, etc.
- Application: Apache, VPN, OAuth, etc.

- Intelligence

- REN-ISAC SES, Critical Stack, Team Cymru MHR, ...

- Data Enrichment

- Additional information to complement logs
- LDAP, DHCP, Inventories



7

CTSC

[Warren]

Let's talk about our various sources. Can you name a few different log sources that you deal with?

We can usually think of logs in one of 3 general categories. System level logs that indicate operating system issues like kernel logs, security audit logs and so on. Network logs that record network traffic metadata, maybe even full packet information. Also as Application level logs like Apache, vpn, Shibboleth, etc logs.

Intelligence is information that can help identify malicious actors and can be from internal or external sources.... These are things like IP addresses of scanners, urls of malicious sites, file hash information for malicious files.

Data enrichment helps to complement data points in your logs. For example you could have a table that associates IP addresses or MAC addresses with specific end-users or groups in your organization. We'll discuss this more in a few slides.

image:

<https://pixabay.com/en/pokemon-ball-pokemon-go-video-games-1530315/>



## Collection - Logs

- Syslog
- Windows Event Logs
- HIDS
  - OSSEC
  - Auditd
  - Keystroke logs
- Netflows
- Bro/NIDS
- DNS
- Scans



CTSC

8

[Warren]

There are various forms of log input that you can use for analysis. The more you have available the more complete of a picture you can have of an incident. The trade off is also that you can have too much information that it can dilute your investigation causing you to take too much time for analysis or hiding your evidence in plain site.

Let's look at a few of these various log sources.

image: <https://pixabay.com/en/wood-logs-lumber-woodpile-firewood-1209632/>



## Collection - Logs - Syslog

- De Facto framework on linux systems
- System & any software can log to it
- Can log centrally; default unreliable UDP
- Components
  - Priority
  - Timestamp
  - Hostname
  - Process
  - Message



9

CTSC

[Warren]

Syslog is a standard message logging system in essentially all unix and unix like systems, networking gear and a myriad of other devices. It's quite simplistic and will likely be one of your largest sources of data, especially if you primarily are a Unix shop.

Most system admins use syslog for debugging applications or identifying what happened on their systems. At the very minimum syslogs are initially created with facility code, a severity, a timestamp, and a message. The facility code identifies the type of program the message is coming from like authentication subsystem, kernel messages, mail system, etc. The severity identifies the classification of the message such as error, debug, warning, critical, alert and so on. Finally the message which includes the timestamp at the beginning complete a single syslog line.

The syslog timestamp can vary from system to system. It can be in the format of yyyy-mm-dd hh:mm:ss. It could include timezone information or not. It could be written out as month and day of the week names or it can be completely numeric. This is important to know because if you collect syslogs from multiple systems it's critical to identify the timezone of the system the log came from and to have some kind of consistency in format to streamline analysis.

Syslogs by default are stored locally and usually have to be configured to be

sent to a remote system. It's best to collect syslog from across the enterprise to a centralized log collection aggregator. This can be a single system, or a series of relays to a single, or multiple collectors.

Be aware, however, that if you don't design your collection infrastructure correctly, you could inadvertently crash systems due to network failures.

For example, if you have your system set up to forward syslogs to another host and that remote host becomes unresponsive and is unable to receive more logs. The sending system will queue these logs until it runs out of local resources and crashes. The flip side is that if you're sending logs over UDP, you can lose logs.

image: <https://pixabay.com/en/bash-terminal-linux-unix-computer-161382/>

## Collection - Logs - Syslog



Priority (Facility + Severity): e.g. <38>, <86>

FACILITY		SEVERITY	
0	kern	kernel messages	0 emerg System is unusable
1	user	user-level messages	1 alert
2	mail	mail system	2 crit Critical conditions
3	daemon	system daemons	3 err Error conditions
4	auth	security/authorization messages	4 warning
5	syslog	messages generated internally by syslogd	5 notice Events that are unusual, but not error conditions.
6	lpr	line printer subsystem	6 info Normal operational messages that require no action.
7	news	network news subsystem	7 debug Information useful to developers for debugging the application.
8	uucp	UUCP subsystem	
9		clock daemon	
10		authpriv/security/authorization messages	
11	ftp	FTP daemon	
12	-	NTP subsystem	
13	-	log audit	
14	-	log alert	
15	cron	scheduling daemon	
16	local0	local use 0 (local0)	
17	local1	local use 1 (local1)	
18	local2	local use 2 (local2)	
19	local3	local use 3 (local3)	
20	local4	local use 4 (local4)	
21	local5	local use 5 (local5)	
22	local6	local use 6 (local6)	
23	local7	local use 7 (local7)	



CTSC

11

[Warren]

Here's a breakdown of the syslog priority. you may see this priority as a number at the beginning of syslog line. It's a combination of the facility and severity code. You multiply the facility by 8 and add the severity value.

image: <https://pixabay.com/en/bash-terminal-linux-unix-computer-161382/>



## Syslog Examples - SSH

```
<38>Aug  1 09:13:58 root sshd[19468]: Accepted publickey  
for wraquel from 10.12.23.15 port 49474 ssh2: RSA  
2b:cb:82:f0:22:d7:8a:f6:cd:70:43:b3:de:cf:5d:ee  
  
<86>2016-08-01T09:13:48.764820-05:00 bastion sshd[2193]:  
Accepted keyboard-interactive/pam for wraquel from  
10.12.23.15 port 49458 ssh2  
  
<38>Aug  1 14:05:17 dev2 sshd[31622]: Failed password for  
root from 10.11.128.16 port 48593 ssh2  
  
<38>Aug  1 09:37:20 honeypot sshd[9256]: Failed password  
for invalid user pi from 192.168.58.61 port 59699 ssh2
```

[Warren]

Here are some sample syslog lines. A common log you're interested in is for ssh logins.

Red - Timestamp

Yellow - Hostname

Cyan - is the process and the process ID

White - the message

38 = 4 - auth/6 - info

86 = 10 - authpriv/6 - info

## Syslog issues

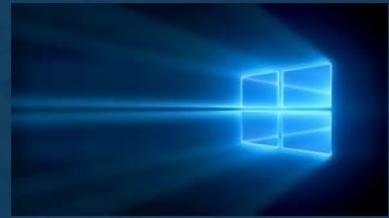
---

- Only so much can be conveyed in text
- Unicode can sometimes cause issues
- Relays may include additional relay hostnames
- Timestamps may be off

## Collection - Logs - Windows Event Logs



- Stored locally; Overwrites itself
- Can be centrally collected
- Unicode
- Very Verbose
- Central Storage a.k.a. Subscriptions



14

CTSC

[Warren]

Syslog is the standard for the unix world, but for Windows, the Windows Event Logging system is the standard. Windows event logs are much more complex. Like syslog, it's possible to centralize windows log collection.

Binary XML format

Windows logs, at least now, are stored in an xml format. It's a wrapping log meaning that it has a predefined log size and when that log limit is reached it wraps back around to the beginning of the log and overwrites the initial logs. The last I checked the default log size was 20MB with a max size at 2TB.

Subscriptions

Event Log Format

Log wrapping -> event log limit size.

Types of Event Logs

Unicode

image source:

[http://blogs-images.forbes.com/patrickmoorhead/files/2015/08/Windows\\_10\\_Hero.png](http://blogs-images.forbes.com/patrickmoorhead/files/2015/08/Windows_10_Hero.png)

<https://www.ultimatewindowssecurity.com/securitylog/encyclopedia/Default.aspx>  
(list of event IDs)

# Collection - Logs - Windows Event Logs



The screenshot shows a Windows Event Viewer window for Event 4634. The event details are as follows:

Subject:	Security ID:	D4N6\wraquel
Account Name:	wraquel	
Account Domain:	D4N6	
Logon ID:	0x688A5B	
Logon Type:	10	
This event is generated when a logon session is destroyed. It may be positively correlated with a logon event using the Logon ID value. Logon IDs are only unique between reboots on the same computer.		
Log Name:	Security	
Source:	Microsoft Windows security	
Event ID:	4634	
Level:	Information	
User:	N/A	
OpCode:	Info	
More Information:	<a href="#">Event Log Online Help</a>	

16

CTSC

[Warren]

Here is an example of a windows event log.

This is viewed through the default Event Viewer in windows.

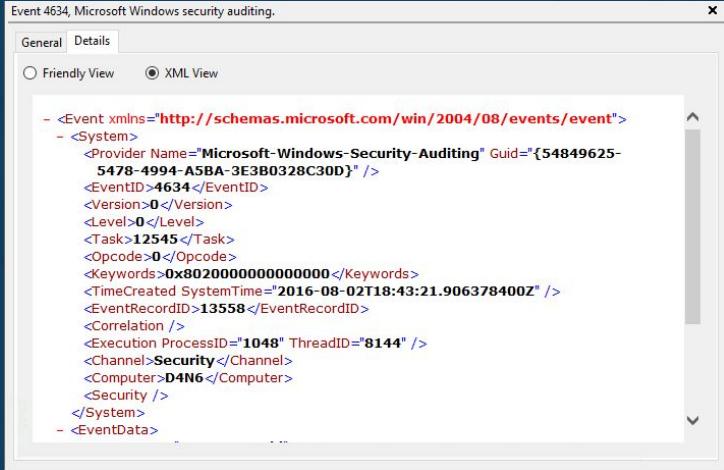
We have the Security ID, in this case it's a local account, wraquel on my forensics server D4N6. This is reflected in the security id, account name and domain name.

There is the Logon ID which is supposed to be a unique number that is reset on system reboots.

There's a logon type, or HOW I logged in. Type 10 is remotely via remote desktop.

You can see this is specifically a security log event, when it was logged (etc).

## Collection - Logs - Windows Event Logs



A screenshot of the Windows Event Log viewer window titled "Event 4634, Microsoft Windows security auditing.". The window has tabs for "General" and "Details", with "XML View" selected. The content area displays the XML structure of the event log entry. The XML code is as follows:

```
<Event xmlns="http://schemas.microsoft.com/win/2004/08/events/event">
  <System>
    <Provider Name="Microsoft-Windows-Security-Auditing" Guid="{54849625-5478-4994-A5BA-3E3B0328C30D}" />
    <EventID>4634</EventID>
    <Version>0</Version>
    <Level>0</Level>
    <Task>12545</Task>
    <Opcode>0</Opcode>
    <Keywords>0x8020000000000000</Keywords>
    <TimeCreated SystemTime="2016-08-02T18:43:21.906378400Z" />
    <EventRecordID>13558</EventRecordID>
    <Correlation />
    <Execution ProcessID="1048" ThreadID="8144" />
    <Channel>Security</Channel>
    <Computer>D4N6</Computer>
    <Security />
  </System>
  <EventData>
```

17

CTSC

[Warren]

Here we can look specifically at details in an XML format. The Friendly View is just like the general tab, nicely formatted for easier reading. The log file itself is in a Binary XML format and this is what the xml record looks like.



## Collection - Logs - Windows Event Logs

Aug 2 13:43:21 D4N6 MSWinEventLog#0111#011Security#0 Aug 02 13:43:21 2016#0114634#011Microsoft-Windows-Security-Auditing#011N /A#011N/A#011Success Audit#011D4N6#011Logoff#011#011An account was logged off. Subject: Security ID: S-1-5-21-3934507682-2419030825-887421081-1004 Account Name: wraquel Account Domain: D4N6 Logon ID: 0x688A5B Logon Type: 10 This event is generated when a logon session is destroyed. It may be positively correlated with a logon event using the Logon ID value. Logon IDs are only unique between reboots on the same computer.#01113558#015

[Warren]

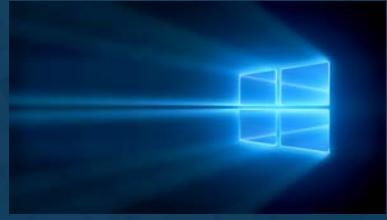
Now lets look at what happens when you convert windows event logs into syslog format. You may want to do this for specific reasons like having all your logs in syslog format. You can see here it's kind of ugly. The the #011 characters are the ascii code for, in this case, tab. This is through nxlog, software that can convert windows logs to syslog format. You can see that it's not that easily readable. #015 is the the newline character.

Talk about benefits, caveats. Lose some interpretations, hard to parse. Must be able to parse correctly on collector.

## Collection - Logs - Windows Event Logs



- Working in mixed environments
  - Subscription service to collect centrally
  - Can be converted to Syslog
  - Can be sent directly to some information managers



19

CTSC

[Warren]  
rsyslog/nxlog -> syslog collector

Windows Subscription service to centralize logs. (from there sent to Splunk, syslog, etc)

Need to configure a source, and a collector. Once they have been configured you create a subscription to indicate what logs you want to collect centrally. The subscription is configured on the collector and events that match this subscription are forwarded by the individual sources.



## Collection - Logs - HIDS

- Usually provides
  - File Integrity
  - Rootkit detection
  - Log Monitoring
  - Process Monitoring



20

CTSC

[Warren]

Host based intrusion detection software can monitor activity on your host. This can include things like file integrity monitoring as was familiar with a tool called 'Tripwire'. HIDS can perform checks to determine if files may be rootkits. They can monitor processes and sometimes even record keystrokes.

OSSEC is a good example of an all around HIDS tool.

OSSEC stands for Open Source HIDS SECurity. OSSEC can provide file integrity monitoring indicating when files change. It can provide rootkit detection. It can follow logs and produce alerts when certain conditions are met. It can monitor active processes.

auditd is another useful tool in the Linux world that can watch for interesting system calls.

In some situations you could also be enabling keystroke logging. At NCSA we leverage a tool called Instrumented SSH that can perform keystroke and session monitoring of ssh sessions. This is done through a patch applied to openssh and only applied on nodes that users jump through, like specific login nodes or bastion hosts for management. It should be noted that any activity that performs keystroke logging should be approved through the appropriate parties.

HIDS monitored activity is very important for identifying activity around a compromise and collecting these logs can be very critical.

image: <https://pixabay.com/en/server-computer-network-database-23315/>



## Collection – Scan Logs

- Sources
  - NMAP
  - Qualys
  - Masscan



CTSC

22

[Warren]

Another useful source of information that should be collected is scan reports, pen testing reports, service audit reports, etc.

These scan results tell you a number of things. What open services there are on your network. What active hosts there are. You can use this information for multiple things (examples).

These can give you a snapshot of your network or even an overview over a period of time if these scans are regular. They can be network based scans, local scans identifying users or software versions but you'll want to know this information. (why)

image: <https://pixabay.com/en/scanner-handheld-barcode-scanning-36385/>



## Overview

- Collection
  - Log sources
  - ***Intelligence***
  - Data Enrichment
- Event Management
- Analysis
- Response



23

CTSC

[Warren]

This morning we will walk through identifying and understanding our various log sources which we will call 'Collection'.

We will then talk about taking those collection sources and methods of storing them for long term archival and short term analysis.

Finally we will have a brief intro into analysis that will carry into a hands on example after lunch.

## Collection - Intelligence



- Collective Intelligence Framework (CIF)
- MISP
- OSINT (Open Source Intelligence)
- Critical Stack
- whitelisting/blacklisting



24

CTSC

[Warren]

Intelligence data helps to quickly identify events associated with known good and bad activity.

If you're not familiar with IoC, it stands for Indicators of Compromise. They are indications of a known compromise and can be as simple as a single IP address or as complex as the MD5 sum of a malicious file paired with IPs or DNS names for command and control hosts.

There are also indicators that you can use for known-good events, like traffic for system updates, traffic from a trusted remote site to your local host, traffic from your scanning hosts and so on.

There are services available, both free and paid, or course, that can provide you with this kind of intelligence. A few examples include ShadowServer, Phishtank, Team Cymru, Spamhaus, and so on.

You can pair this intelligence with your monitoring tools to identify potentially malicious activity and respond accordingly. This could be for automatic blocks or just alerts to investigate.

There are a few tools that can be leveraged to collect multiple sources of intelligence in order to automate integration with your infrastructure.

image: <http://yahooentertainment.tumblr.com/image/88019669538>



## Overview

- Collection
  - Log sources
  - Intelligence
  - ***Data Enrichment***
- Event Management
- Analysis
- Response



CTSC

26

[Warren]

This morning we will walk through identifying and understanding our various log sources which we will call 'Collection'.

We will then talk about taking those collection sources and methods of storing them for long term archival and short term analysis.

Finally we will have a brief intro into analysis that will carry into a hands on example after lunch.

## Collection - Data Enrichment



- Helps define non-descript data
- Types
  - hostnames
  - groups
  - users
- Sources



27

CTSC

[Warren]

Incoming logs are usually very minimal. An IP address or a hostname, maybe a username or a user id are included in the log line. Data Enrichment sources help complement your data in order to better “enrich” your analysis.

A few good examples include things like:

DHCP logs. They can provide a mac address and sometime a hostname. If you log enough you can track an IP address all the way back to a port. If you don't know who a system belongs to sometimes the system name that's presented in DHCP logs (\*example\*) can provide a clue to the owner. A MAC address ran through the IEEE OUI (Organizationally Unique Identifier) registry can sometimes tell you the manufacturer of a system.

Configuration management databases can help identify the owner of a system by IP address. Services that should be running on the system.

LDAP can tell you what department or groups a person belongs to or contact information.

There are external sources like the Team Cyrum ASN lookup tool that can tell you information about an IP address.

GeoIP databases can tell you the approximate location on earth that an IP address is located.

Data enrichment can come from the logs you are already collecting. They are usually the type of logs that don't provide "event" type information, but expand the information about the resources identified in our logs.

image: <https://pixabay.com/en/hammer-wrench-repair-work-industry-28636/>

## Collection - Data Enrichment



- Helps define non-descript data
- Types
- Sources
  - DHCP
  - Lifecycle Management Tools  
(Foreman, puppet)
  - CMDB
  - Inventory



[Warren]

image: <https://pixabay.com/en/hammer-wrench-repair-work-industry-28636/>



## Collection - Authorization

- Legal Issues
- Buy In
- Who owns the data?



30

CTSC

[Warren]

Now that we've covered a few different types of data that we could collect let's look at a few issues that you'll run into trying to get these logs.

There are a plethora of systems on Science networks. They can include a central IT, individual research groups, network infrastructure, affiliates, external customers and so on. All types of data could be present in those logs so you can't just blindly tell people, send me your logs. You run into some problems like:

- What kind of information are you sending me?
- What format is it in?
- Are there compliance issues like being sent from systems that interact with sensitive data, will that data somehow end up in your logs knowingly or unknowingly
- Do you have a legal obligation to meet.

You'll need to review your incoming data to determine legal obligations.

The other problem is getting buy in from your community in sending you their logs. Some blockers that you'll run into include:

- Security concerns sending you their information
- Lack of knowledge on how to do this
- Lack of desire to send info.

The issue that is usually most prevalent is that the people that have the data you want to collect has no buy in. If you have a security awareness program for your users, that's one thing, you also need to consider a security awareness program for your system and network administrators. Easily the biggest blocker is how much effort is required for admins to send you their data. You can help minimize this by providing guidance on how to send info. For example a page that includes configuration scripts and step-by-step notes on how to send you their logs.

Another issue is that people don't want to give you data that they don't have to. Being hassled to provide data they are not required to provide can be problematic. You can address this a number of ways:

- Trust must be earned
- Make it mandatory – baked into grant requirements or if they're paying for security services have it part of the agreement. This may involve higher level AUP/SLA/MOU changes.

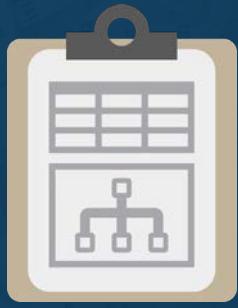
Special exemptions  
IRB data  
Different restriction policies

image: <https://pixabay.com/en/meeting-relationship-business-1019771/>



## Overview

- Collection
- ***Event Management***
  - Filtering
  - Storage
  - Indexing
  - Accessibility
- Analysis
- Response



[Warren]

image: <https://pixabay.com/en/clipboard-data-science-chart-908886/>



## IEM – Considerations

- How do we send logs to a central location?
- What format are they in?
- Can this format be converted?
- How will we store this information for later retrieval?
- What rate is all this data coming in at?
- Do I want to store all this data or some of it?



33

CTSC

[Warren]

Earlier we went over identifying your various information inputs. Now we need to determine how to store them for later analysis.

Some questions we need to ask are:

- How do we send logs to a central location
- What format are they in?
- Can this format be converted?
- How will we store this information for later retrieval?
- What rate is all this data coming in at?
- Do I want to store all this data or some of it?

image: <https://pixabay.com/en/question-mark-pile-question-mark-1495858/>



## Overview

- Collection
- Event Management
  - *Filtering*
  - Storage
  - Indexing
  - Accessibility
- Analysis
- Response



CTSC

[Warren]

image: <https://pixabay.com/en/clipboard-data-science-chart-908886/>



## Filtering

- Reducing amount of incoming data
  - Can we drop useless messages?
  - Do we need all these sources?
- Sorting on the fly
  - Timestamps
  - Message Types
  - Categorization

[Warren]

image: <https://pixabay.com/en/clipboard-data-science-chart-908886/>



## Overview

- Collection
- Event Management
  - Filtering
  - ***Storage***
  - Indexing
  - Accessibility
- Analysis
- Response



[Warren]

image: <https://pixabay.com/en/clipboard-data-science-chart-908886/>



## IEM – Transfer Methods

- Scheduled copies via scp, ftp
- Live stream
- Pub/Sub Message Brokers
- Scaling infrastructure
- Encryption



37

CTSC

[Warren]

image: <https://pixabay.com/en/network-cables-network-connector-494648/>



## IEM - Timestamps

- Properly sync logs
- Consistent timezones
- Delayed logs placed in incorrect locations
- Lack of timezone info
- Skew



CTSC

38

[Warren]

image: <https://pixabay.com/en/folder-synchronize-sync-computer-26706/>

## Storage Goals – Information and Event Management



- Centralized access to information
- Consistent format
- Ease of access
  - Machine parseable
  - Human Readable
  - Categorized



39

CTSC

[Warren]

Once we've identified our various data inputs we need to manage them. Let's identify some requirements for being able to manage logs in a scalable way.

What is our goal here? We want to make accessing logs easy. We want it accessible for analysis in future steps. We want to be able to find what we're looking for quickly, or at least know exactly where we will find specific data. We need to try to correlate various different logs sources.

The best way to do this is to centralize log collection. Sending all your data to a central system has multiple considerations.

We obviously understand that we can't ideally place all our logs in one place, like everything in syslog format. Some log formats are not easily translatable in that way, like windows event logs. We do, however, want to minimize the number of locations we need to access in order to get the information we need.

We need to make sure that our data is stored in a consistent and accessible format. That could be flat files like syslog or databases storing scan results.

These different storage have to have a method of getting logs to them. e.g. syslog via TCP, s/ftp/scp locations.

Data needs to be presentable in a human readable form, although it may not be stored that way. It should at the very least be stored in a machine parseable way.

Incoming Log format vs storage

Transfer methods

Accessibility

Permissions

Data separate issues (FERPA/HIPPA/PII)

Proper Categorization

Storage space

I/O Limits

Long term vs short term storage

image: <https://pixabay.com/en/horizontal-old-weights-old-scale-930716/>



## IEM – Storage Concerns

- Scalability
- I/O Limitations
- Long Term storage
  - FOIA Requests
  - Archivists
- Categorization
- Timestamp issues
- Raw vs Parsed



41

CTSC

[Warren]

Will we be able to deal with bursts, long term storage?

Can our hardware handle the influx of data? Will data be dropped?

Do we have constraints to adhere to? Are we required to keep logs for a certain amount of time?

We need to store data in a way that is logical and easily accessible.

Are logs stored in the correct timestamp, come from the correct time zones? Are they processed in bulk or on the fly?

Do we want to store information raw with indexes or as parsed information or both?

image: <https://pixabay.com/en/wine-cellar-cave-bottles-old-1329061/>



## Overview

- Collection
- Event Management
  - Filtering
  - Storage
  - ***Indexing***
  - Accessibility
- Analysis
- Response



[Warren]

image: <https://pixabay.com/en/clipboard-data-science-chart-908886/>



## Indexing

- Quicker access to raw data
- Keywords
- Timestamps/Frames
- Summaries



CTSC

[Warren]

image: <https://pixabay.com/en/clipboard-data-science-chart-908886/>



## Overview

- Collection
- Event Management
  - Filtering
  - Storage
  - Indexing
  - ***Accessibility***
- Analysis
- Response



[Warren]

image: <https://pixabay.com/en/clipboard-data-science-chart-908886/>



## IEM - Accessibility & Interfaces

- Dashboard Access
- Direct log access
- Permissions
- Data Sensitivity



45

CTSC

[Warren]

Who needs access to these logs apart from your analysis?

Why duplicate storage for other groups?

Is there sensitive data?

Do you have the ability to limit access to the right data for users?

image: <https://pixabay.com/en/key-metal-metallic-security-steel-575681/>



## nInfo example

<https://github.com/JustinAzoff/ninfo>



[Warren]

image: <https://pixabay.com/en/fishing-nets-fishing-rope-fisherman-1179533/>



## IEM – Log Correlations

- Consider how logs help complete a story
- How do they relate
- Automating Correlation



CTSC

[Warren]

image: <https://pixabay.com/en/puzzle-last-particles-piece-654957/>



## IEM – Log Management Tools

- Splunk
- ELK



48

CTSC

[Warren]

note: ELK renaming to Elastic Stack

There do exist a number of log management tools that already exist. A few notable examples include suites like Splunk and ELK. Many of them apply this overall approach of “Security” Information and Event Management. They vary in cost but invariably the largest factor in considering these tools is the fine tuning.

image:

<http://opentica.com/wp-content/uploads/2016/02/Screen-Shot-2014-02-25-at-4.4.2.52-PM-1024x557-830x451.png>

image: <http://www.splunk.com/content/dam/splunk/img/grafh.png>

image: <http://www.rsyslog.com/files/2015/06/Missing-tabs.png>

## IEM – Log Management Tools



- May take years to tune
- Specialized Training
- Black box



49

CTSC

[Warren]

Even between pay solutions like Splunk and Free solutions like the Elasticsearch Stack you have to consider a number of things. First these tools are designed to be configured for your environment. They often don't come ready for diverse environments seen in the Science industry. If you have the money they can manage it for you. You can train people to run these tools. Money can solve everything, usually.

For corporate environments that are extremely strict and there are very specific aspects to host systems and servers operate, it can be fairly straightforward.

Another issue is the black box nature of many applications. Without understanding how the underlying logic of the application works it could be difficult integrating your own logs and analysis into this system.

image:

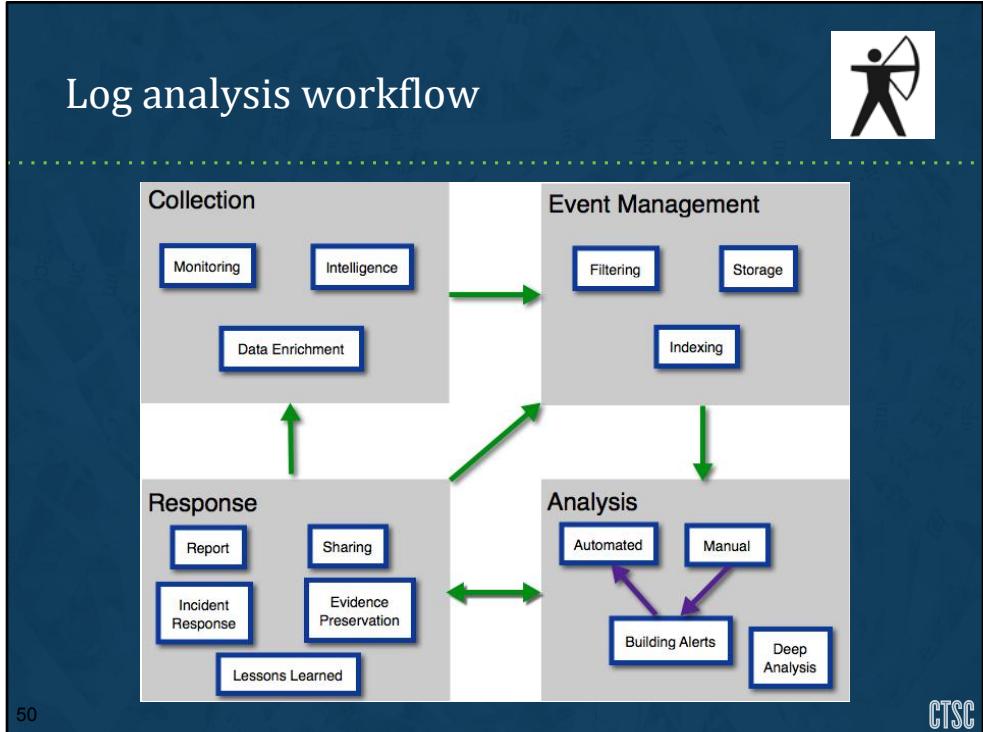
<http://opentica.com/wp-content/uploads/2016/02/Screen-Shot-2014-02-25-at-4.4.2.52-PM-1024x557-830x451.png>

image: <http://www.splunk.com/content/dam/splunk/img/graph.png>

image: <http://www.rsyslog.com/files/2015/06/Missing-tabs.png>



## Log analysis workflow



[Adam]

Cycle: never ends, self reinforces (guides org. today)

Define Log: Any system or device generated record (includes intel feeds and enrichment like inventory DBs)

Filters & indexers make manageable (size can be huge to store & search)

No “right” way to organize, but should scale

Analysis (manual to auto)-> feeds into what to investigate

Response (investigation touches all cycles)



## Analysis

- Collection
- Event Management
- ***Analysis***
- Response



[Warren]->[Vlad]

image: <https://pixabay.com/en/arrows-growth-hacking-marketing-1229855/>



## Analysis

- Manual
- Alerts
- Automated
- Deep Dives



[Warren]->[Vlad]

image: <https://pixabay.com/en/arrows-growth-hacking-marketing-1229855/>



## Analysis

- ***Manual***
  - Requires domain knowledge
  - Tool proficiency
- Alerts
- Automated
- Deep Dives



[Warren]->[Vlad]

image: <https://pixabay.com/en/arrows-growth-hacking-marketing-1229855/>



## Analysis

- Manual
- ***Alerts***
  - Creates reports
  - Can act (block, notify)
  - Lots of tweaking
- Automated
- Deep Dives



[Warren]->[Vlad]

image: <https://pixabay.com/en/arrows-growth-hacking-marketing-1229855/>



## Analysis

- Manual
- Alerts
- **Automated**
  - Watch for false positives
  - Track actions
  - Review often
- Deep Dives



[Warren]->[Vlad]

image: <https://pixabay.com/en/arrows-growth-hacking-marketing-1229855/>



## Analysis

- Manual
- Alerts
- Automated
- ***Deep Dives***
  - One off analysis
  - May require access to data you don't have access to



[Warren]->[Vlad]

image: <https://pixabay.com/en/arrows-growth-hacking-marketing-1229855/>



## Analysis

- What about experience?
- Just act like a scientist.



[Warren]->[Vlad]

image: <https://pixabay.com/en/arrows-growth-hacking-marketing-1229855/>



## Analysis - Scientific Method

- Make an observation
- Ask questions
- Develop testable predictions
- Test, test, test (refine)
- Develop a conclusion (was I correct?)



Strange url request from a server  
What is the purpose of the URL?  
Should the server be doing this?

image: <https://pixabay.com/en/arrows-growth-hacking-marketing-1229855/>

# Introduction to Bro Logs

# What is Bro?

- "Network Security Monitor"
- Inspects all network traffic
- Generates forensically sound logs
- Has a scripting language, making it very extensible (automation, custom functionality, etc.)

# Bro Logs

- Goal: An analyst can review the Bro logs and determine what occurred on the network
- Logs are small (many installations keep years of logs; some perpetual)
- Logs are plain-text, tab-delimited CSV
- Designed for grep, sed, awk

# Bro Log Limitations

- Logs are an abstraction
  - 100s of Gbps of input, a few GB/hour of output requires loss of data
- Only some protocols can be parsed
- Rising use of encryption
- Limited by network visibility

## Example 1: IRC

141.142.11.2:4766 -> 164.32.77.23:6667

JOIN #foobar

irc.log	Value
id.orig_h	141.142.11.2
id.orig_p	4766
id.resp_h	164.32.77.23
id.resp_p	6667
nick	USA 74634
user	[urX]-700159
command	JOIN
value	#foobar

## Example 2: HTTP

```
> GET /phpBB HTTP/1.1  
< HTTP/1.1 404 Not Found
```

	Value
http.log	
id.orig_h	192.168.1.20
id.orig_p	7182
id.resp_h	141.142.192.147
id.resp_p	80
method	GET
host	<a href="http://www.ncsa.edu">www.ncsa.edu</a>
uri	/phpBB
status_code	404

# Types of Bro Logs: Network Protocols

- conn
- dhcp
- dns
- ftp
- http
- irc
- kerberos
- mysql
- radius
- rdp
- sip
- smtp
- snmp
- socks
- ssh
- ssl
- syslog
- tunnel

# Types of Bro Logs: Files

- files
- pe (Portable Executable)
- x509 (Certificate information)

# Types of Bro Logs: Detection

- intel (Intelligence data hits)
- notice (Alerts from Bro scripts)
- signatures (Traffic signature hits)

# Types of Bro Logs: Network Observations

- known\_certs (SSL certs observed)
- known\_devices (MAC addresses)
- known\_hosts (IPs w/ established TCP)
- known\_services (Open ports)
- software (Software w/ version info)

# Viewing Logs

- less
- grep
- bro-cut
- [try.bro.org](http://try.bro.org)
- ELK
- Splunk

# Finding Compromised Systems

# Phases of Compromise

1. Reconnaissance
2. Exploitation
3. Reinforcement
4. Consolidation
5. Pillage

13

1. Reconnaissance. Identify target assets and vulnerabilities, indirectly or directly.
2. Exploitation. Abuse, subvert, or break a system by attacking vulnerabilities or exposures. If the intruder does not seek to maintain persistence, then this could be the end of the compromise.
3. Reinforcement. The intruder deploys his persistence and stealth techniques to the target.
4. Consolidation. The intruder ensures continued access to the target by establishing remote command-and-control.
5. Pillage. The intruder executes his mission. Here we assume data theft and persistence are the goals.

# Phases of Compromise

- 1. Reconnaissance**
2. Exploitation
3. Reinforcement
4. Consolidation
5. Pillage

14

Low-hanging fruit: won't stop a determined attacker.

## Recon: HTTP Not Found

Which IP had the most HTTP 404 Not Found errors?

```
$ cat http.log |  
bro-cut id.orig_h status_code |  
grep 404 |  
sort | uniq -c | sort -n |  
tail -n 1
```

```
165 64.39.106.131 404
```

```
$ dig +short -x 64.39.106.131
```

```
sn031.s01.sea01.qualys.com.
```

# Phases of Compromise

1. Reconnaissance
- 2. Exploitation**
3. Reinforcement
4. Consolidation
5. Pillage

# Compromise 1: http.log

id.resp_h	91.134.161.42
id.resp_p	80
method	GET
host	top4download.org
uri	/
referrer	-
user_agent	Mozilla/5.0 (Windows NT 6.1; WOW64; Trident/7.0; rv:11.0) like Gecko
status_code	200
resp_mime_types	text/html

```
> GET top4download.org/ (IE 11)
< 200 OK (text/html)
```

## Compromise 1: http.log

id.resp_h	91.134.161.60
id.resp_p	80
method	GET
host	roseindia.vip
uri	/?644v0o1fsfarlf06=24&4e0lef...
referrer	http://top4download.org/
user_agent	(IE 11)
status_code	200
resp_mime_types	text/html

## Compromise 1: http.log

id.resp_h	91.134.161.60
id.resp_p	80
method	GET
host	roseindia.vip
uri	/?644v0olfsfarlf06=24&4e0lef...
<b>referrer</b>	<b>http://top4download.org/</b>
user_agent	(IE 11)
status_code	200
resp_mime_types	text/html

# Referrer Chain

```
1.> GET top4download.org/ (IE 11)
   < text/html

2.> GET roseindia.vip/?
    644v0o1fsfarflf06=24&4e0flef186v43re1bv=1280
    &37fj4d7g94969r=720 (IE 11)
   < text/html

3.> GET 18a43zad864bo96.armlay.gdn/
   < text/html

4.> GET 18a43zad864bo96.armlay.gdn/
    712g40rdf7k06 (IE 11)
   < application/x-shockwave-flash
```

# Overview for IP

91.134.161.42	GET	top4download.org	text/html
91.134.161.60	GET	roseindia.vip	text/html
62.138.5.199	GET	18a43z96.armlay.gdn	text/html
62.138.5.199	GET	18a43z96.armlay.gdn	application/x-shockwave-flash
62.138.5.199	GET	18a43z96.armlay.gdn	text/html
62.138.5.199	GET	18a43z96.armlay.gdn	application/x-shockwave-flash
62.138.5.199	GET	62.138.5.199	-
62.138.5.199	GET	62.138.5.199	application/x-dosexec
62.138.5.199	GET	62.138.5.199	application/x-dosexec
52.3.78.30	GET	ipinfo.io	text/json

The image shows a screenshot of the malwr analysis interface. The main header features the word "malwr" in a large, bold, black font, with a small black beetle icon integrated into the letter "r". Below the header is a navigation menu with several options: "Quick Overview", "Static Analysis", "Behavioral Analysis", "Network Analysis" (which is currently selected, indicated by a blue border), "Dropped Files", and "Comment Board (0)". To the right of the menu is a large button labeled "Download PCAP" with a circular arrow icon. Below the menu, there are links to "Domains (1)", "Hosts (2049)", "HTTP (1)" (which is highlighted in blue), "IRC (0)", and "SMTP (0)". The main content area is titled "HTTP Requests" and contains a table with two columns: "URI" and "DATA". A single row is shown, corresponding to the highlighted HTTP request: `http://ipinfo.io/json`. The "DATA" column displays the raw HTTP request: `GET /json HTTP/1.1`  
Host: ipinfo.io

URI	DATA
<code>http://ipinfo.io/json</code>	<code>GET /json HTTP/1.1</code> Host: ipinfo.io

## Overview for IP

1. Redirect chain with random-looking domain names, and suspicious TLDs (.vip, .gdn)
2. Shockwave Flash followed by Windows executable download
3. Queried for IP address information

## Compromise 2: http.log

Jun 17 23:00:10	CcMeer3amA5aZ9nrx	107.160.46.226	4908
Jun 18 02:10:21	CFVSv31q8HACwAJS0c	107.160.46.226	4534
Jun 18 02:10:21	CQMaBW2KP1XCGMVNlb	107.160.46.226	4533
Jun 18 02:34:35	CqA2Xg3qh9LrpI6IEj	107.160.46.226	2516
Jun 18 02:34:35	CTAMVF3Rv4jhcgBRAc	107.160.46.226	2517
Jun 18 02:34:35	CTAMVF3Rv4jhcgBRAc	107.160.46.226	2517
Jun 18 02:35:02	CaBfuW2tjnMvk7FnIl	107.160.46.226	3747
Jun 18 02:35:02	CSI7QrHUkubbD8nU1	107.160.46.226	3750
Jun 18 02:35:02	CSI7QrHUkubbD8nU1	107.160.46.226	3750
Jun 18 02:35:21	CNKTTv3nBgLDfPfs8h	107.160.46.226	4118
Jun 18 02:35:21	CNKTTv3nBgLDfPfs8h	107.160.46.226	4118
Jun 18 02:35:21	CvIFb23hBFppvF8Rvc	107.160.46.226	4117
Jun 18 02:35:22	Cdps1P3VQgyPytnHmk	141.142.234.27	47772

# Phases of Compromise

1. Reconnaissance
2. Exploitation
- 3. Reinforcement**
4. Consolidation
5. Pillage

26

1. Reconnaissance. Identify target assets and vulnerabilities, indirectly or directly.
2. Exploitation. Abuse, subvert, or break a system by attacking vulnerabilities or exposures. If the intruder does not seek to maintain persistence, then this could be the end of the compromise.
3. Reinforcement. The intruder deploys his persistence and stealth techniques to the target.
4. Consolidation. The intruder ensures continued access to the target by establishing remote command-and-control.
5. Pillage. The intruder executes his mission. Here we assume data theft and persistence are the goals.

# Compromise 3: Reinforcement

```
Jul 27 19:32:19 141.142.227.45 22 SSH::SERVER OpenSSH_6.6.1p1  
Jul 27 20:29:39 141.142.227.45 22 SSH::SERVER OpenSSH_6.6.1p1  
Jul 27 22:27:53 141.142.227.45 22 SSH::SERVER OpenSSH_6.6.1p1  
Jul 27 23:30:34 141.142.227.45 22 SSH::SERVER OpenSSH_6.5.1p1
```

# Compromise 3: Reinforcement

```
Jul 27 19:32:19 141.142.227.45 22 SSH::SERVER OpenSSH_6.6.1p1  
Jul 27 20:29:39 141.142.227.45 22 SSH::SERVER OpenSSH_6.6.1p1  
Jul 27 22:27:53 141.142.227.45 22 SSH::SERVER OpenSSH_6.6.1p1  
Jul 27 23:30:34 141.142.227.45 22 SSH::SERVER OpenSSH_6.5.1p1
```

# Finding Compromised Accounts

# Phases of Compromise

1. Reconnaissance
2. Exploitation
- 3. Reinforcement**
4. Consolidation
5. Pillage

30

1. Reconnaissance. Identify target assets and vulnerabilities, indirectly or directly.
2. Exploitation. Abuse, subvert, or break a system by attacking vulnerabilities or exposures. If the intruder does not seek to maintain persistence, then this could be the end of the compromise.
3. Reinforcement. The intruder deploys his persistence and stealth techniques to the target.
4. Consolidation. The intruder ensures continued access to the target by establishing remote command-and-control.
5. Pillage. The intruder executes his mission. Here we assume data theft and persistence are the goals.

# Bro Logs

- mysql
- http
- radius
- kerberos
- conn
- software

# Authentication Logs

1. Successful login by "deprovisioned" user
2. Logins at strange hours (or during vacation)
3. Obvious scanning activity
4. Multiple failures followed by a success
5. GeoIP data (where is the user coming from?)

# MySQL Logs

- ts: Timestamp for the event
- uid: Unique ID for the connection
- id: orig\_h, orig\_p, resp\_h, resp\_p
- cmd: The command that was issued
- arg: The argument issued to the cmd
- success: Did the command succeed?
- rows: The number of affected rows
- response: Server message

# MySQL Logs

1. Detect vulnerable web servers (SQLi)
2. Look for large number of rows being returned.
3. 

```
select unhex('7F454C40000E9...00')
into dumpfile '/usr/lib64/mysql/
plugin/unknown/xiaojiji64.so'
```

# HTTP Logs

id.resp_h	91.134.161.42
id.resp_p	80
method	GET
host	<a href="http://www.ncsa.edu">www.ncsa.edu</a>
uri	/
referrer	<a href="http://shibboleth.ncsa.eu/idp/profile/SAML2">shibboleth.ncsa.eu/idp/profile/SAML2</a>
user_agent	Mozilla/5.0 (Windows NT 6.1; WOW64; Trident/7.0; rv:11.0) like Gecko
status_code	200
resp_mime_types	text/html

# HTTP Logs

- Referrer mismatch
- File type mismatch (a.jpg is x-doseexec)
- Header abuse (HTTPoxy)

# RADIUS Logs

- Extended attributes:
  - OS of the client
  - List of interfaces on the client
  - Physical location (wireless AP)
- Often better than RADIUS accounting

# Phases of Compromise

1. Reconnaissance
2. Exploitation
3. Reinforcement
- 4. Consolidation**
5. Pillage

38

1. Reconnaissance. Identify target assets and vulnerabilities, indirectly or directly.
2. Exploitation. Abuse, subvert, or break a system by attacking vulnerabilities or exposures. If the intruder does not seek to maintain persistence, then this could be the end of the compromise.
3. Reinforcement. The intruder deploys his persistence and stealth techniques to the target.
4. Consolidation. The intruder ensures continued access to the target by establishing remote command-and-control.
5. Pillage. The intruder executes his mission. Here we assume data theft and persistence are the goals.

# Kerberos Logs

- Mimikatz: Service ticket request without an authentication ticket?
- Weak crypto algorithms
- See password changes - report for accounts with no expiration

# Phases of Compromise

1. Reconnaissance
2. Exploitation
3. Reinforcement
4. Consolidation
- 5. Pillage**

40

1. Reconnaissance. Identify target assets and vulnerabilities, indirectly or directly.
2. Exploitation. Abuse, subvert, or break a system by attacking vulnerabilities or exposures. If the intruder does not seek to maintain persistence, then this could be the end of the compromise.
3. Reinforcement. The intruder deploys his persistence and stealth techniques to the target.
4. Consolidation. The intruder ensures continued access to the target by establishing remote command-and-control.
5. Pillage. The intruder executes his mission. Here we assume data theft and persistence are the goals.

# Conn Logs

- Exfil - large flows
- Protocol mismatches (ssl over tcp/80)
- Missing protocols (udp/53, but not DNS)
- Entropy analysis

# Putting It All Together: Intelligence

# Why NSM and not IDS?

- Crucial difference:
  - IDS is logging "bad" things
  - NSM is logging everything
  - What if you didn't know that something was bad?

# Post-Mortem Analysis

A system or account was compromised.  
Search \*.log and generate indicators of  
compromise:

```
zcat 2016-08-*/*conn*.log.gz |  
awk '$3 == "64.39.106.131"'
```

# Post-Mortem Analysis

A system or account was compromised.  
Search \*.log and generate indicators of  
compromise:

```
zcat 2016-08-*/conn*.log.gz |  
awk '$3 == "64.39.106.131" || ($1 ~ /^#/)'
```

# Post-Mortem Analysis

A system or account was compromised.  
Search \*.log and generate indicators of  
compromise:

IP addresses involved:

64.39.106.131 is doing recon scans

# Post-Mortem Analysis

A system or account was compromised.  
Search \*.log and generate indicators of  
compromise:

URLs:

<http://18a43zad864bo96.armlay.gdn/> is  
hosting malware

# Post-Mortem Analysis

A system or account was compromised.  
Search \*.log and generate indicators of  
compromise:

Software versions:

OpenSSH\_6.5.1p1 is outdated and might  
be trojaned

# Post-Mortem Analysis

A system or account was compromised.  
Search \*.log and generate indicators of  
compromise:

E-mail addresses:

vladg@illinois.edu is sending malicious  
attachments

# Post-Mortem Analysis

A system or account was compromised.  
Search \*.log and generate indicators of  
compromise:

Domain names:

ncsa.eu is a phishing site

# Post-Mortem Analysis

A system or account was compromised.  
Search \*.log and generate indicators of  
compromise:

File hashes:

3aac91181c3b7eb34fb7d2b6dd673f4827fcf07  
is a Flash exploit

# Post-Mortem Analysis

A system or account was compromised.  
Search \*.log and generate indicators of  
compromise:

File names:

r00tkit.tgz is a legitimate Linux ISO

# Post-Mortem Analysis

A system or account was compromised.  
Search \*.log and generate indicators of  
compromise:

SSL certificate hashes:

EC:50:1C:4A:...:A2:4C:C6:60:CC:49:03:86  
is the default Ubuntu snake oil key  
with a well known private key

# Post-Mortem Analysis

A system or account was compromised.  
Search \*.log and generate indicators of  
compromise:

Pub key hashes:

a1:73:d1:e1:25:72:79:...:ed:81:bf:67:98  
is a trojaned sshd

# Post-Mortem Analysis

Once you have indicators of compromise, you can search your historical logs to find other compromised accounts or systems. Some might have different TTPs, which generate more indicators.

# Pre-Mortem Analysis

Perhaps more importantly, you can use Bro's Intel framework to load your indicators and be alerted when they're seen in the future.

Indicator	Type	Comment	Notice?
64.39.106.131	ADDR	Recon scan	F
<a href="http://18a43zad864bo96.armlay.gdn/">http://18a43zad864bo96.armlay.gdn/</a>	URL	Hosting malware	T
<a href="http://ncsa.eu">ncsa.eu</a>	DOMAIN	Phishing site	T
r0otkit.tgz	FILE_NAME	Malware	T

# Where Are You Looking?

- Domain
  - DNS::IN\_REQUEST
  - HTTP::IN\_HOST\_HEADER
  - SSL::IN\_SERVER\_NAME
  - X509::IN\_CERT

## Overview



- Collection
- Event Management
- Analysis
- **Response**
  - Reports
  - Incident Response
  - Lessons Learned
  - Intelligence Sharing
  - Preservation



CTSC

59

[Warren]

The final section of the log analysis lifecycle is the reports section. We first identified all the data we were collecting, we then looked at how to store that data. Next we walked through doing actual analysis against these logs. Now we're looking at what to do with all the information we pull out of our analysis.

The output of our analysis section results in a number of things.

Reports

Actionable items for incident response

Issues that need to be improved

Identification of known threats which can be shared with partner organizations.

image: <https://pixabay.com/en/halloween-party-panel-celebration-1013907/>



## Response - Reports

- Automated Reports
  - Login counts
  - Log volume counts
  - Top Talkers
- Metrics
  - How much data you are consuming
  - Number of events
  - Response times



60

CTSC

[Warren]

Usually it's a good idea to have a number of automatic reports that you can review to determine if there are issues in your organization.

(Who do we make reports for)

Some examples include:

- Login counts from unfamiliar hosts
- Valid/Invalid login counts (good for spotting brute force attempts or breakdowns in central authentication)
- Volume of incoming logs
- Top talkers
- Darknet activity

These reports can be generated and automatically acted upon during your automated analysis stage. You can have specific thresholds that can detect network scans, brute force attacks, web spider activity and such.

A very important report that you should be generating is a metrics report. IT Security usually has a difficult time justifying its existence because when we're actually doing a good job, we're practically invisible. Having numbers to present to C-level can help validate your role.

- Number of logs that you scan

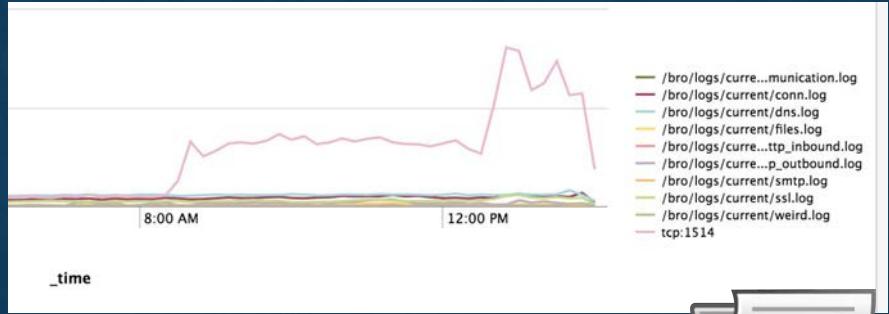
- Number of attacks
- Number of blocks
- Number of systems monitored through different methods like syslog, passive network monitoring,
- Network activity that's being analyzed
- Response times to suspicious events

Not only are these good for justifying staff and resources, but they can help you identify gaps in your security coverage.

\*\*Maybe find more examples than the one below\*\*

image: <https://pixabay.com/en/folder-files-paper-office-document-303891/>

## Reports - Example



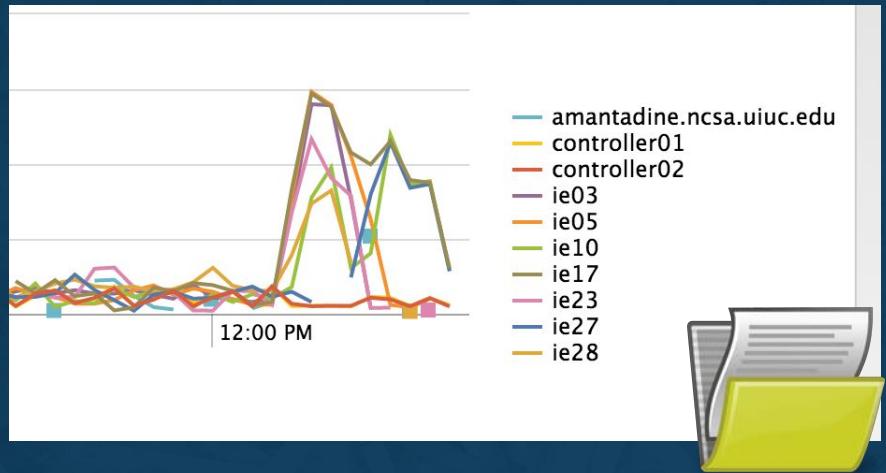
CTSC

62

[Warren]  
syslog graph

image: <https://pixabay.com/en/folder-files-paper-office-document-303891/>

## Reports - Example



CTSC

63

syslog graph (continuation)

image: <https://pixabay.com/en/folder-files-paper-office-document-303891/>



## Overview

- Collection
- Event Management
- Analysis
- Response
  - Reports
  - ***Incident Response***
  - Lessons Learned
  - Intelligence Sharing
  - Preservation



CTSC

image:

<http://www.johnwatsonblog.co.uk/staticarchive/f76f7d1887f444a697624d664f0110a453d397dc.jpg>

## Incident Response

- Provide all logs from time frame
- May expand in scope
- May require specialized knowledge
- Remember time skews
- May identify new sources



CTSC

image:

<http://www.johnwatsonblog.co.uk/staticarchive/f76f7d1887f444a697624d664f0110a453d397dc.jpg>



## Overview

- Collection
- Event Management
- Analysis
- Response
  - Reports
  - Incident Response
  - **Lessons Learned**
  - Intelligence Sharing
  - Preservation



image: <http://fcapaprofessor.com/wp-content/uploads/2016/05/lessonslearned.jpg>

## Response – Lessons Learned



- Are there false positives that need to be fixed?
- Do metrics indicate anything?
- What are the gaps in our analysis and monitoring?



67

CTSC

Inevitably you'll have issues with your framework. Maybe you're blocking legitimate hosts automatically, losing syslogs somewhere, or consistently having to wade through a sea of false positives. You'll need to make sure you have a Review or Lessons Learned process in your framework to identify problems and start addressing them. This review should include examining FP, FN, TN, and even TP to ensure your detection is working correctly, examining your metrics collection and identifying places where you have gaps. For example, the collection of certain types of logs take significantly longer than other logs. Identifying efficiencies should be a goal.

As mentioned earlier, metrics are a critical aspect, which is often overlooked. Metrics will help you provide a face to an otherwise invisible operation.

image: <http://fcpaprofessor.com/wp-content/uploads/2016/05/lessonslearned.jpg>



## Overview

- Collection
- Event Management
- Analysis
- Response
  - Reports
  - Incident Response
  - Lessons Learned
  - ***Intelligence Sharing***
  - Preservation



68

CTSC

The final section of the log analysis lifecycle is the reports section. We first identified all the data we were collecting, we then looked at how to store that data. Next we walked through doing actual analysis against these logs. Now we're looking at what to do with all the information we pull out of our analysis.

The output of our analysis section results in a number of things.

Reports

Actionable items for incident response

Issues that need to be improved

Identification of known threats which can be shared with partner organizations.

image: <https://pixabay.com/en/halloween-party-panel-celebration-1013907/>

## Response – Intelligence Sharing



- Providing Blocklists
  - To partner institutions
  - To individual hosts/groups
- Intelligence Communities
  - ISACs
  - OSINT Contributors



69

CTSC

Collaboration is a very important aspect of any scientific endeavor. Information Security is no exception. Collecting intel from partner sites and even sharing with them helps to provide a better security. For example, say that you have compromised hosts that are scanning the entire internet for ssh servers. Two other institutions that you share intel with have seen scans from this host but you haven't. Implementing intel from these sites can help you prepare for those by blocking that scanner.

Say that you have a compromised host on your network and you have the information of where that attack came from, files they've downloaded and other useful compromise indicators. This info can potentially help you identify other compromised hosts or block compromise vectors. Sharing this with your partner institutions can help them prevent or even identify compromises on their networks and vice versa.

You can also provide this intel for admins within your site to be implemented at the host level. Maybe they won't willingly provide access to their systems but they may take advantage of lists systems to block.

\*Discuss CIF and MISP here if not touched on earlier in Collection portion\*

image: <https://pixabay.com/en/elephant-africa-namibia-nature-dry-1170108/>



## Overview

- Collection
- Event Management
- Analysis
- Response
  - Reports
  - Incident Response
  - Lessons Learned
  - Intelligence Sharing
  - ***Preservation***



70

CTSC

[Adam]

The final section of the log analysis lifecycle is the reports section. We first identified all the data we were collecting, we then looked at how to store that data. Next we walked through doing actual analysis against these logs. Now we're looking at what to do with all the information we pull out of our analysis.

The output of our analysis section results in a number of things.

Reports

Actionable items for incident response

Issues that need to be improved

Identification of known threats which can be shared with partner organizations.

image:<http://www.dcpputnamconsulting.com/wp-content/uploads/2014/07/Hoarding.jpg>

## Reports - Preservation & Sharing



- Have a documented retention policy w/ defaults
  - Consistency & standards of practice protect you
  - Document exceptions and processes!
- When you might store more briefly
  - Special sensitive data
- When you might store longer
  - FOIA, e-discovery, research,...
  - Provenance & chain of custody may be critical
  - ASCII is your friend
  - Keep originals



71

CTSC

[Adam]

Tension between hoarding & exposure.

image: <https://pixabay.com/en/files-folders-papers-objects-313733/>

## Reports - Preservation & Sharing



- Who do you share with?
  - Institutional policies, regulations, your own privacy policy
  - IRBs, NDAs, etc.
- How do you protect it?
  - Securing transfers
  - Sanitization
  - Limited queries
  - Bring researcher to the data

[Adam]

image: <https://pixabay.com/en/files-folders-papers-objects-313733/>



## Response – Active Response

- Vlad demo'ing NCSA's Black-Hole router



CTSC

image: <https://pixabay.com/en/halloween-party-panel-celebration-1013907/>



CENTER FOR TRUSTWORTHY  
SCIENTIFIC CYBERINFRASTRUCTURE  
The NSF Cybersecurity Center of Excellence

## CTSC webinar series: [trustedci.org/webinars](http://trustedci.org/webinars)

---

We thank the National Science Foundation (grant 1547272) for supporting our work.

The views and conclusions contained herein are those of the author and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the NSF.



CENTER FOR TRUSTWORTHY  
SCIENTIFIC CYBERINFRASTRUCTURE  
The NSF Cybersecurity Center of Excellence

## Thank You

CTSC:  
[trustedci.org](http://trustedci.org)  
 @TrustedCI

Bro:  
[bro.org](http://bro.org)  
 @Bro\_IDS

NCSA:  
[ncsa.illinois.edu](http://ncsa.illinois.edu)  
 @NCSAAatIllinois

---

We thank the National Science Foundation (grant 1547272) for supporting our work.

The views and conclusions contained herein are those of the author and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the NSF.