Welcome to

5. Baseline Your Data

KEA Kompetence SIEM and Log Analysis

Henrik Kramselund he/him han/ham hlk@zencurity.com @kramse

Slides are available as PDF, kramse@Codeberg

5-baseline-siem.tex in the repo security-courses

Goals for today

Todays goals:

- How would you design a minimal production setup
- Selecting technology – some recommendations
- Alerting and reporting – what is available now in your systems

Photo by Thomas Galler on unsplash

Plan for today

Subjects

- Real systems, how to design
- Technology options
- A little Veris, and regular/yearly state-of-reports

Exercise theme: Moving into a real deployment

- Mostly discussion about options for deploying SIEM tools and technologies
- Alerting and reporting

Follow up from last, Dashboard example designed for Kibana – I don't use it myself YMMV
https://github.com/devdjdjdj/kibana-presenter

Reading Summary

DDS 7. Learning from Security Breaches VERIS

DDS 12. Moving Toward Data-Driven Security

IDIR 1. Introduction

IDIR 2. Basics of Intelligence

Reading Summary, continued

Source: DDS 7. Learning from Security Breaches VERIS

- We all see the same attacks, make it easier to communicate between applications and organizations
- Learn from others
- Another example MITRE ATT&CK® https://attack.mitre.org/

We will now browse the chapter, and discuss your experiences

Reading Summary, continued

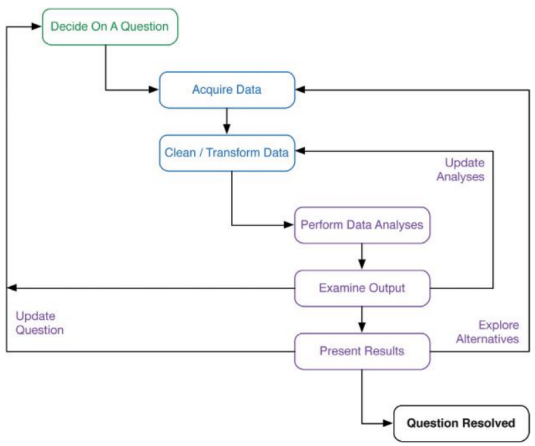


FIGURE 12-2 *The data science workflow*

- Find and Collect Relevant Data
- Learn through Iteration
- Find Statistics

Source: DDS 12. Moving Toward Data-Driven Security

Reading Summary, continued

Building a Real-Life Security Data Science Team

... a clear goal: Given an IP address (or IP/Port combination), **search across all our perimeter devices in less than five minutes.**

Three core principles focused the team.

- First, explore the open source versions of tools before engaging vendors. If you don't know how the sausage is being made, you really have no idea what's being done, and this is vital when working with real data.
- Second, follow the mantra of "no single tool; no single database; and, no single approach to solving a problem." Do not put blinders on because you are either comfortable with certain technologies or have an affinity for a certain tool.
- Third, failure is expected, but you must learn from each journey down the wrong path. Continuous adaptation and adjustment is the name of the game.

Source: DDS 12. Moving Toward Data-Driven Security

Reading Summary, continued

> When you begin implementing intelligence-driven incident response, it is important to have a solid understanding of both intelligence and incident-response processes. Part 1 provides an introduction to cyber-threat intelligence, the intelligence process, the incident-response process, and how they all work together.

Source: IDIR 1. Introduction

- Obviously a book about incident response, but a modern one

Reading Summary, continued

> There was a time when many people considered indicators of compromise, or IOCs, to be synonymous with threat intelligence. IOCs, which we will reference a lot and cover in depth later in the book, are things to look for on a system or in network logs that may indicate that a compromise has taken place. This includes IP addresses and domains associated with command-and-control servers or malware downloads, hashes of malicious files, and other network- or host-based artifacts that may indicate an intrusion. As we will discuss throughout this book, however, there is far more to threat intelligence than IOCs, although IOCs still remain one of the most common types of technical intelligence around intrusions.

Source: IDIR 2. Basics of Intelligence

- List sources of intelligence, HUMINT, OSINT, SIGINT etc.
- Introduces the OODA loop, “observe, orient, decide, act.” and the Intelligence Cycle
- OODA loop also used in the SOC book

Data is a piece of information, a fact, or a statistic. Data is something that describes something that is. In our previous example about the weather report, the temperature is a piece data. It is a fact, something that has been measured using a **proven and repeatable process**. Knowing the temperature is important, but in order to be useful for decision making, it must be analyzed in the context of what else is going on that day. **In information security, an IP address or domain are data.** Without any addi- tional analysis to provide context, they are simply facts. When various **data points are gathered and analyzed** to provide **insight** around a particular requirement, it **becomes intelligence**.

Source: IDIR 2. Basics of Intelligence

- Without data we cannot do much
- If you have an incident, and there will be incidents, you need it!
- Better gather some basic data now, even if you don't use it currently

Intelligence is derived from a process of collecting, processing, and analyzing data. Once it has been analyzed, it **must be disseminated** in order to be useful. Intelligence that does not get to the right audience is wasted intelligence. Wilhelm Agrell, a Swedish writer and historian who studied peace and conflict, once famously said, "Intelligence analysis combines the dynamics of journalism with the problem solving of science."

Source: IDIR 2. Basics of Intelligence

- Sharing data helps us and others
- We can use many sources of data to enable quicker response

Indicators of Compromise

> There was a time when many people considered indicators of compromise, or IOCs, to be synonymous with threat intelligence. IOCs, which we will reference a lot and cover in depth later in the book, are **things to look for** on a system or in **network logs** that may **indicate that a compromise has taken place**. This includes IP addresses and domains associated with command-and-control servers or malware downloads, hashes of malicious files, and other network- or host-based artifacts that may indicate an intrusion.

Source: *Intelligence-Driven Incident Response* (IDIR)
Scott Roberts. Rebekah Brown

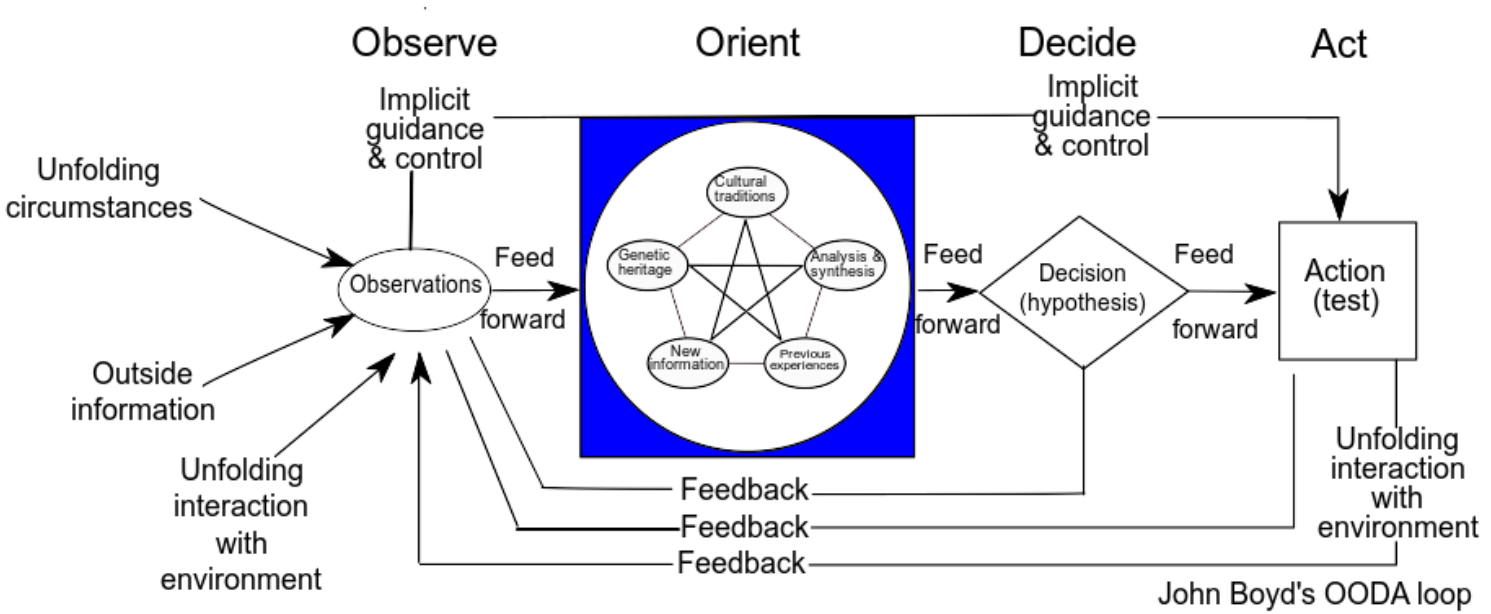
Indicators of Compromise and Signatures

An IOC is any piece of information that can be used to objectively describe a network intrusion, expressed in a platform-independent manner. This could include a simple indicator such as the IP address of a command and control (C2) server or a complex set of behaviors that indicate that a mail server is being used as a malicious SMTP relay.

When an IOC is taken and used in a platform-specific language or format, such as a Snort Rule or a Zeek-formatted file, it becomes part of a signature. A signature can contain one or more IOCs.

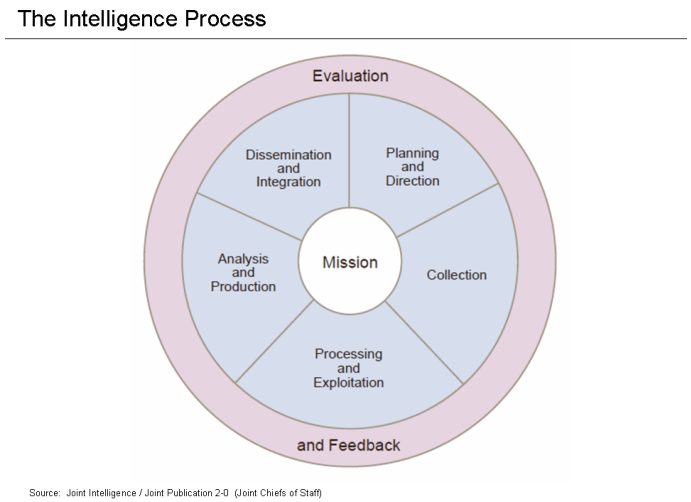
Source: Applied Network Security Monitoring Collection, Detection, and Analysis, 2014 Chris Sanders

OODA Loop by John Boyd



Source: Patrick Edwin Moran - Wikipedia https://en.wikipedia.org/wiki/OODA_loop

Intelligence Cycle or Intelligence Process



Source: https://en.wikipedia.org/wiki/Intelligence_cycle

- I decided to take the more original Intelligence Process picture, which has a bit more details

Processing

Let's look at some processing

- Processing includes normalizing collected data into uniform formats for analysis
- Indexing – Large volumes of data need to be made searchable
- Translation – for our course we might get multiple input formats but need JSON or XML
- Enrichment – Providing additional metadata for a piece of information is important. For example, domain addresses need to be resolved to IP addresses, and **WHOIS registration data fetched**
- Filtering – Not all data provides equal value, and analysts can be overwhelmed when presen- ted with endless streams of irrelevant data
- Prioritization – The data that has been collected may need to be ranked so that analysts can allo- cate resources to the most important items
  Note: this relates to a *baseline*, what errors are normal in your environment
- Visualization – Data visualization has advanced significantly and the human eye and brain can often see patterns

Metadata – enrichment

Metadata	Metacategory: Data
The DNS lookup occurred at a certain time.	Timestamp: 278621182
The internal host sent a DNS PTR request.	Network protocol: DNS PTR
The internal host had a hostname.	Location: Desktop subnet
	Source IP Address: 1.1.1.2
	Hostname: windowspc22.company.com
The internal host resolved an external host.	Location: External
	Destination IP Address: 255.123.215.3
	Hostname: dgf7adfnkjhh.com
The external host was hosted by a dynamic DNS provider.	Network: Shady DDnS Provider Inc.
	ASN: SHADY232
	Reputation: Historically risky network
The remote hostname appeared randomly generated.	Hostname: dgf7adfnkjhh.com
	Category: Unusual, nonlinguistic

Source: picture from Crafting the InfoSec Playbook, CIP

Metadata + Context

Drill down process

We have seen Kibana multiple times, but how do you use it? I recommend the following iterative process

1. Get an overview
2. Research top talkers,
3. When identified and handled, remove with filter `not host 10.1.2.3`
4. Look at the next ones

Look into details, lookup hostnames – hopefully your tool allows some help

How to get started

How to get started searching for security events?

Collect basic data from your devices and networks

- Netflow data from routers
- Session data from firewalls
- Logging from applications: email, web, proxy systems

Centralize!

Process data

- Top 10: interesting due to high frequency, occurs often, brute-force attacks
- *ignore*
- Bottom 10: least-frequent messages are interesting

Logstash pipeline

> Logstash is an open source, server-side data processing pipeline that ingests data from a multitude of sources simultaneously, transforms it, and then sends it to your favorite "stash." (Ours is Elasticsearch, naturally.)
> https://www.elastic.co/products/logstash

```
input { stdin { } }
output {
  elasticsearch { host => localhost }
  stdout { codec => rubydebug }
}
```

- Logstash receives via **input**
- Processes with **filters** - grok
- Forward events with **output**

Logstash as SNMPtrap and syslog server

```
input {
  snmptrap {
    host => "0.0.0.0"
    type => "snmptrap"
    port => 1062
    community => "xxxxx"   }
  tcp {
    port => 5000
    type => syslog  }
  udp {
    port => 5000
    type => syslog  }
}
```

- We run logstash on port 5000 - but use IPtables port forwarding
- Have you even configured SNMP traps?
- Maybe you have a device sending SNMP traps right now ...

IPtables forwarding

```
*nat
:PREROUTING ACCEPT [0:0]
# redirect all incoming requests on port 514 to port 5000
-A PREROUTING -p tcp --dport 514 -j REDIRECT --to-port 5000
-A PREROUTING -p udp --dport 514 -j REDIRE}}CT --to-port 5000
-A PREROUTING -p udp --dport 162 -j REDIRECT --to-port 1062
COMMIT
```

Inserted near beginning of /etc/ufw/before.rules on Ubuntu

Remember defense in depth, dont run a privileged Java VM process as root ☺

Grok expresssions

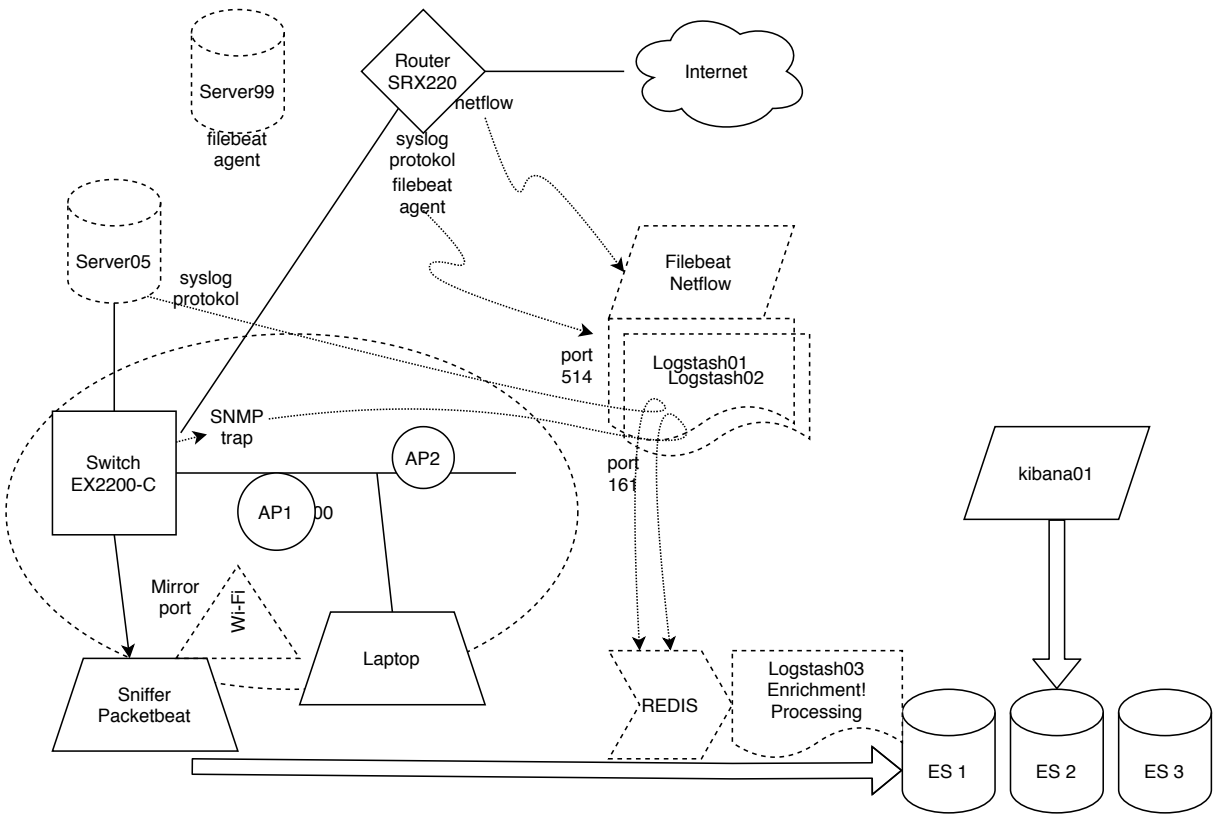
```
filter {
  if [type] == "syslog" {
    grok {
      match => { "message" => "%{SYSLOGTIMESTAMP:syslog_timestamp}
      %{SYSLOGHOST:syslog_hostname} %{DATA:syslog_program}
      (?:\[%{POSINT:syslog_pid}\])?: %{GREEDYDATA:syslog_message}" }
      add_field => [ "received_at", "%{@timestamp}" ]
      add_field => [ "received_from", "%{host}" ]
    }
    syslog_pri { }
    date {
      match => [ "syslog_timestamp", "MMM  d HH:mm:ss", "MMM dd HH:mm:ss" ]
    }
  }
}
```

- Logstash filter expressions grok can normalize and split data into fields

Source: Config snippet from

http://logstash.net/docs/1.4.1/tutorials/getting-started-with-logstash

Lets design a SIEM Infrastructure Proof of Concept



Deploying security

> **Security is a process, not a product.** Products provide some protection, but the only way to effectively do business in an insecure world is to put processes in place that recognize the inherent insecurity in the products. The trick is to reduce your risk of exposure regardless of the products or patches.

Source: https://www.schneier.com/essays/archives/2000/04/the_process_of_secur.html

Today, we will consider the deployment plan being:

- People – make sure management is on board, Sources – which data to gather,
- Technology – select SIEM, architecture, tools and products
- Dashboards – we have done parts of this, refer to SOC book also
- Procedures – left as a home exercise today

People: Data Analysis Skills

Although we could spend an entire book creating an exhaustive list of skills needed to be a good security data scientist, this chapter covers the following skills/domains that a data scientist will benefit from knowing within information security:

- Domain expertise—Setting and maintaining a purpose to the analysis
- Data management—Being able to prepare, store, and maintain data
- Programming—The glue that connects data to analysis
- Statistics—To learn from the data
- Visualization—Communicating the results effectively

It might be easy to label any one of these skills as the most important, but in reality, the whole is greater than the sum of its parts. Each of these contributes a significant and important piece to the workings of security data science.

Source: *Data-Driven Security: Analysis, Visualization and Dashboards* Jay Jacobs, Bob Rudis
ISBN: 978-1-118-79372-5 February 2014 https://datadrivensecurity.info/ - short DDS

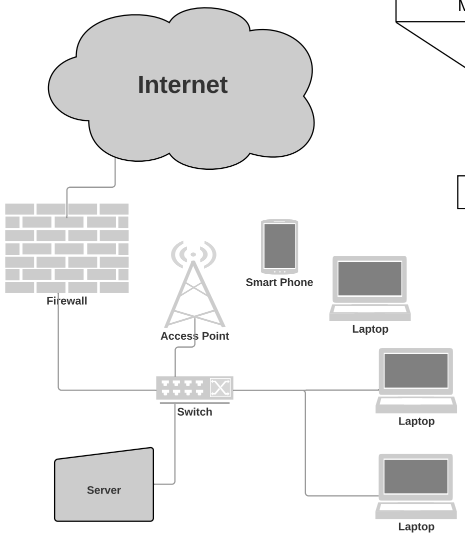
People: Get management buy-in

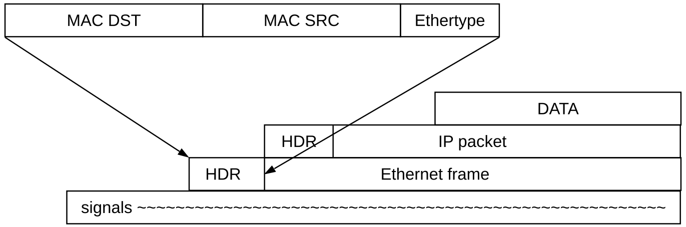
You will probably need help from:

- Buy-in from management, for requesting resources
- Network and security departments – getting data, opening ports
- Application developers, web site programmers

Lifeguard training photo by Margarida CSilva on Unsplash

Sources: Network overview without SIEM





Ethernet frames contain packets

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|Version|  IHL  |Type of Service|          Total Length         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|         Identification        |Flags|      Fragment Offset    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  Time to Live |    Protocol   |         Header Checksum       |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                       Source Address                          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                    Destination Address                        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                    Options                    |    Padding    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Example Internet IPv4 Packet Header

- Internet, routers, firewalls, switches, clients and servers (Wi-Fi not shown)

Sources: Strategy for implementing identification and detection

We recommend that the following strategy is used for implementing identification and detection – logging:

- ☐ Enable system logging from servers
- ☐ Enable system logging from network devices
- ☐ Enable logging from client devices
- ☐ Centralize logging
- ☐ Add search facilities and dashboards
- ☐ Perform system audits manually or automatically
- ☐ Setup alerting and notification with procedures

Extended Sources

When a basic logging infrastructure is setup, it can be expanded to increase coverage, by adding more sources:

- DNS query logging – will enable multiple cases to be resolved, example malware identification and tracing, when was a malware domain queried, when was the first infection
- Web proxy logging – which web pages did which client access
- Session data from Firewalls, Netflow – traffic patterns can be investigated and both attacks and cases like exfiltration can likely be seen

Hint: Take the sources available first, make a proof-of-concept, expand coverage

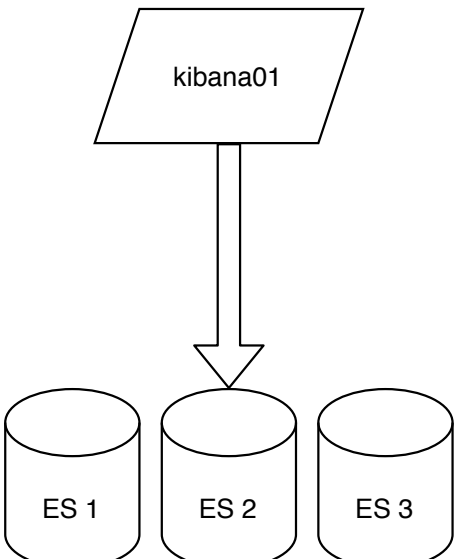
Architecture: Tools

We will use the tools presented during the course:

- Elastic stack: Elasticsearch, Logstash, Kibana, Filebeat, Packetbeat
- Zeek and Suricata can easily be added at a later stage
- Likewise DNS and web proxy logging could be added

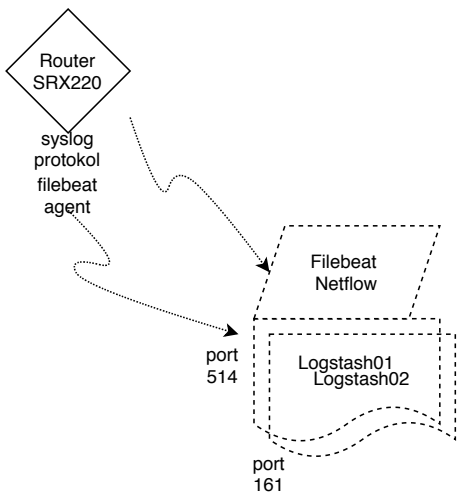
The setup discussed here would be a good proof of conccept, and be valuable almost immediately

Elasticsearch



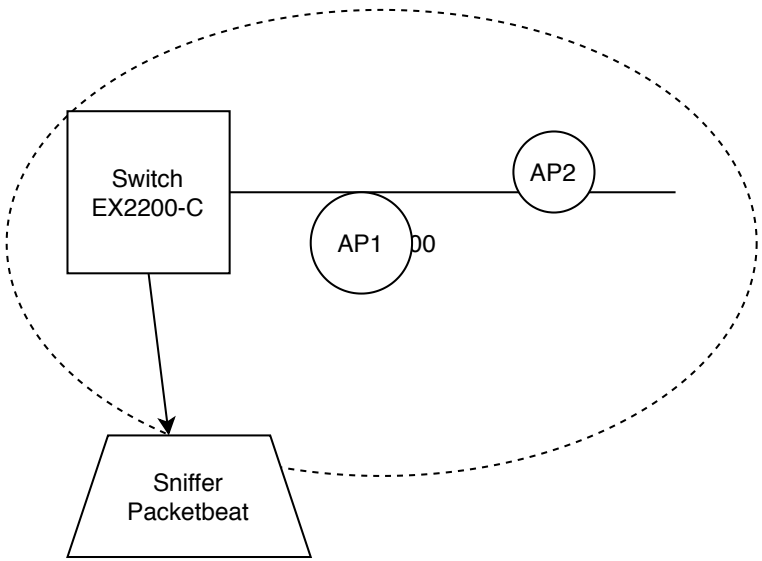
- We plan to build a basic cluster with Elasticsearch, latest stable
- Multiple ES nodes for easier upgrade, redundancy and performance
- Each have 200Gb disk and 16Gb memory allocated

Logstash – syslog and SNMP trap receiver



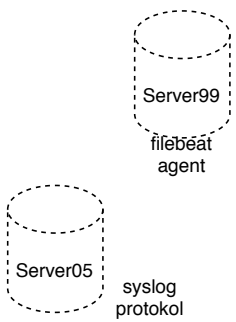
- We have network devices which can only send syslog and SNMP trap – *push events from the network*
- So enable inputs: snmptrap, tcp, udp and use UFW to redirect ports
- We have made two servers, which use VRRP to have a common address

Packetbeat



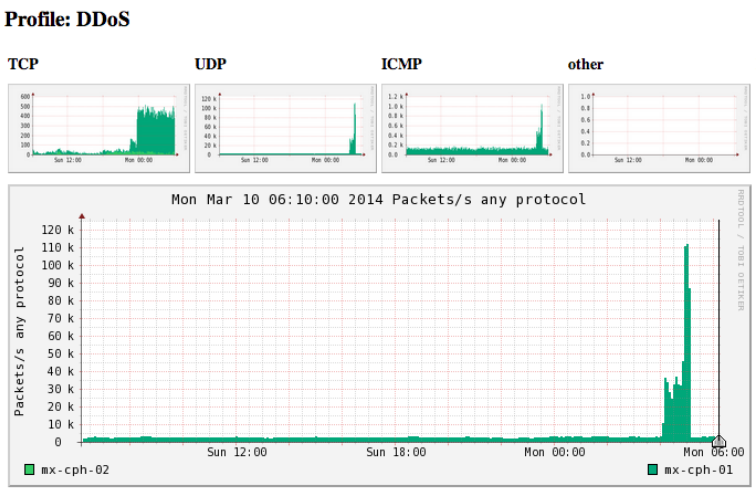
- By installing packetbeat and doing network mirroring from the network switch, we can gather a lot of information
- Packetbeat supports Elastic Common Schema (ECS) https://www.elastic.co/beats/packetbeat
- ICMP (v4 and v6) DHCP (v4) DNS HTTP AMQP 0.9.1 Cassandra Mysql PostgreSQL Redis Thrift-RPC MongoDB Memcache NFS TLS SIP/SDP (beta)

Application servers



- We told the server and application people to use Filebeat and Syslog
- The Linux people decided to use syslog
- Windows servers use Filebeat https://www.elastic.co/beats/filebeat
- All of them send to the Logstash instances

Baseline



- Picture from NFsen running a specific profile to catch attacks
- When you have a running system, it will start to gather a baseline
- Comparing data from various times become possible, and usefull
- The best baseline is from running the actual systems and services for an extended *learning* period

Alerting

We're excited to announce a new alerting framework that delivers a first-class alerting experience natively within the SIEM, Uptime, APM, and Metrics applications as part of the Kibana 7.7 release.

Alerting is a fundamental use case across the Elastic Stack, which is why we're making it part of the core experience within Kibana. Whether you are monitoring application transactions or tracking brute force login attempts, our goal is to provide a tailored experience that allows you to build powerful alerts in the normal flow of your task. The new alerting framework is built from the ground up and designed to offer more than just convenient interfaces. We understand the need to go beyond just notifying people which is why we've also incorporated the ability to trigger predefined actions that can do anything from sending an email to using brand new third-party integrations with platforms like Slack and PagerDuty.

The new alerting framework is being introduced as a beta in the 7.7 release of Kibana and is available immediately on the Elasticsearch Service on Elastic Cloud, or for download.

- https://www.elastic.co/blog/introducing-the-new-alerting-framework-for-observability-security-and-the-elastic-stack
- https://www.elastic.co/what-is/kibana-alerting
- https://www.elastic.co/blog/alerting-in-the-elastic-stack

Alerting everywhere

> Alerting everywhere: Kibana 7.7 introduces ubiquitous alerting for Elastic Observability, Elastic Security, and the Elastic Stack. Users can now create alerts directly from within the SIEM, APM, Metrics, and Uptime applications as well as for any index.

- Seems a lot has happened with alerting in the new version!
- Lets try to work with the alerting framework, note: sending email can sometimes be tricky without some configuration.

Exercise

Now lets do the exercise

Alerting in Eleastic Stack – 30min

which is number **30** in the exercise PDF.

Common task: Elasticsearch Upgrades!

Should we try upgrading to next major version of ES?
https://www.elastic.co/blog/whats-new-elastic-8-0-0

How would we proceed

- Make a backup, clone the VM
- Update repositories, ansible or manually (we only have on cluster node)
- Shut down ES, easy since only one node
- Upgrade software and start again – this will update the data on that node!

Extra steps are needed when upgrading larger clusters. You should definitely try upgrading before going into production. Better if you have multiple clusters, test/staging and production.

ES reporting

Push a button, get a report. Easy.
Kibana is a fantastic way to visualize and explore your Elasticsearch data. Its reporting features let you easily export your favorite Kibana visualizations and dashboards. Each report is print-optimized, customizable, and PDF-formatted. And the option to add your own logo will give your reports the branded, polished look that will color your team impressed.

Source: https://www.elastic.co/what-is/kibana-reporting

- Not sure I agree, but some features are available
- Discussion! Writing and presenting are two very different things, so are dashboards and reports!

Automating Report Generation

Create a POST URL

Create the POST URL that triggers a report to generate.

Source: https://www.elastic.co/guide/en/kibana/current/automating-report-generation.html

- Not sure I agree, but some features are available
- I like the automated report generation, getting data pushed from ES is a great feature.
- Correlation also added https://www.elastic.co/blog/whats-new-elastic-security-7-10-0-correlation-cloud-visibility-detectio

Automatic reporting: tcpflow

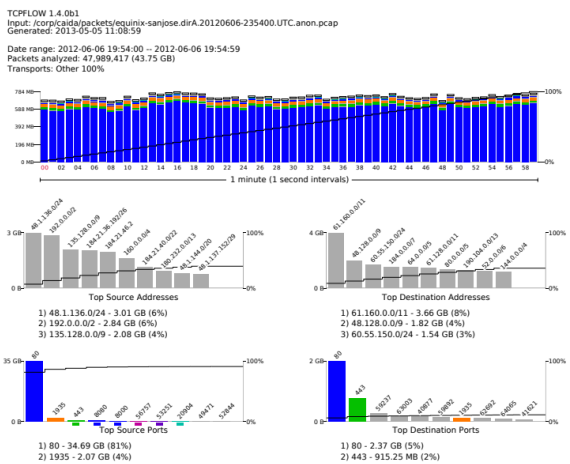


Figure 1: *tcpflow*'s one-page visualization. The color key for the timeline stacked bargraph is presented in the source and destination graphs. Each graph includes a CDF. Here we shows the CAIDA equinix-sanjose.dirA.20120606-235400.UTC.anon.pcap capture

- https://github.com/simsong/tcpflow
- *Passive TCP Reconstruction and Forensic Analysis with tcpflow*, Simson Garfinkel and Michael Shick, Naval Postgraduate School Technical Report NPS-CS-13-003, September 2013. https://calhoun.nps.edu/handle/10945/36026

Another route perhaps?

We could also look at how others prepare their processes, using their SIEM etc.

Browse chapter 1 from Practical Threat Intelligence, Valentina Palacín

Practical Threat Intelligence and Data-Driven Threat Hunting Valentina Palacín, 2021, ISBN: 978-1-83855-637-2

- Chapter 1: What Is Cyber Threat Intelligence?
- Why this book?! It is about Mitre ATT&CK framework and it has been in Humble Bundles, so if it shows up again, buy it
- This book is very much hands on with lots of links, references, tools and names
- I will now present a bit from the book, since you don't have it

Investigate links

- We cannot go through all of it, but we can get inspired
- Since this came from real actors, campaigns, threats it is mostly what a real case would be

This will lead to the later part, doing *an investigation*

Investigate links 1: OSSEM

OSSEM: To help with the heavy work of creating data dictionaries, the Rodriguez brothers created the Open Source Security Events Metadata (OSSEM) for documenting and standardizing security event logs. The project is open source and can be accessed through the project's GitHub repository https://github.com/hunters-forge/OSSEM.

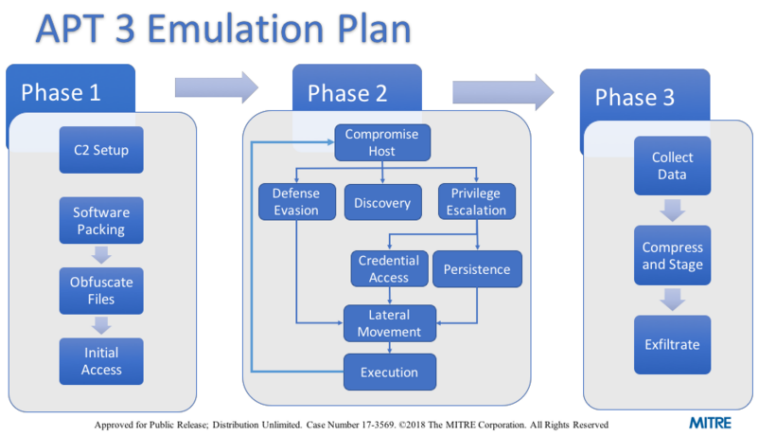
Investigate links 2: Threat Hunter Playbook

The Threat Hunter Playbook: This open source project is maintained by the Rodriguez brothers and is meant to help with the documentation project and sharing threat hunting concepts, developing certain techniques, and building the hypothesis. You can read more about it in the project's GitHub repository https://github.com/hunters-forge/ThreatHunter-Playbook

Investigate links 3: Adversary emulation

Emulating the adversary: Adversary emulation is a way for red teamers to replicate adversary behaviors in their organization's environments. In order to do that, the adversary behaviors need to be mapped and the techniques used by them need to be chained together to create an action plan. The MITRE ATT&CK™ Framework provides an example of how to create an emulation plan based on APT3
https://attack.mitre.org/resources/adversary-emulation-plans/

MITRE Adversary Emulation Plans



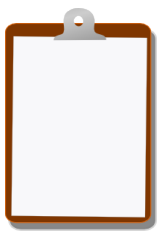
To showcase the practical use of ATT&CK for offensive operators and defenders, MITRE created Adversary Emulation Plans. These are prototype documents of what can be done with publicly available threat reports and ATT&CK.

Source: https://attack.mitre.org/resources/adversary-emulation-plans/

Investigate links 4: Mordor dataset

Mordor: For this stage of the hunt, the Rodriguez brothers created the Mordor project, which provides "pre-recorded security events generated by simulated adversarial techniques"in JSON format.
https://github.com/hunters-forge/mordor

For Next Time



Think about the subjects from this time, write down questions

Check the plan for chapters to read in the books

Visit web sites and download papers if needed

Retry the exercises to get more confident using the tools