



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@Github [🔗](#)

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, <https://github.com/devdjdj/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/>

# Reading Summary, continued

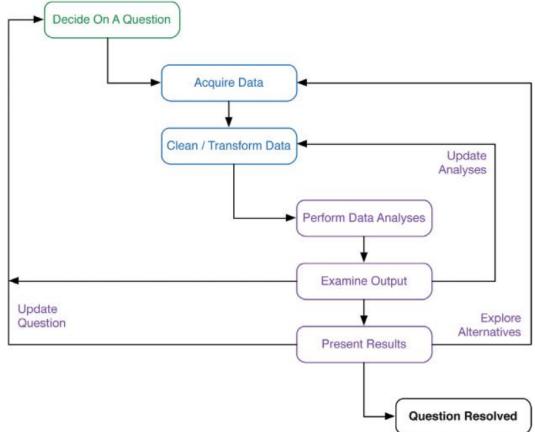


FIGURE 12-2 The data science workflow

- 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

## Side note: redundancy



- Redundancy can often be achieved in multiple ways
- Two servers - physical servers
- Two disks - RAID1 mirrors
- OR using virtualisation
- Redundancy is not an alternative to backup!
- Make sure you plan for upgrades too! Upgrading one server while the other handle messages

## 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 REDIRECT --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



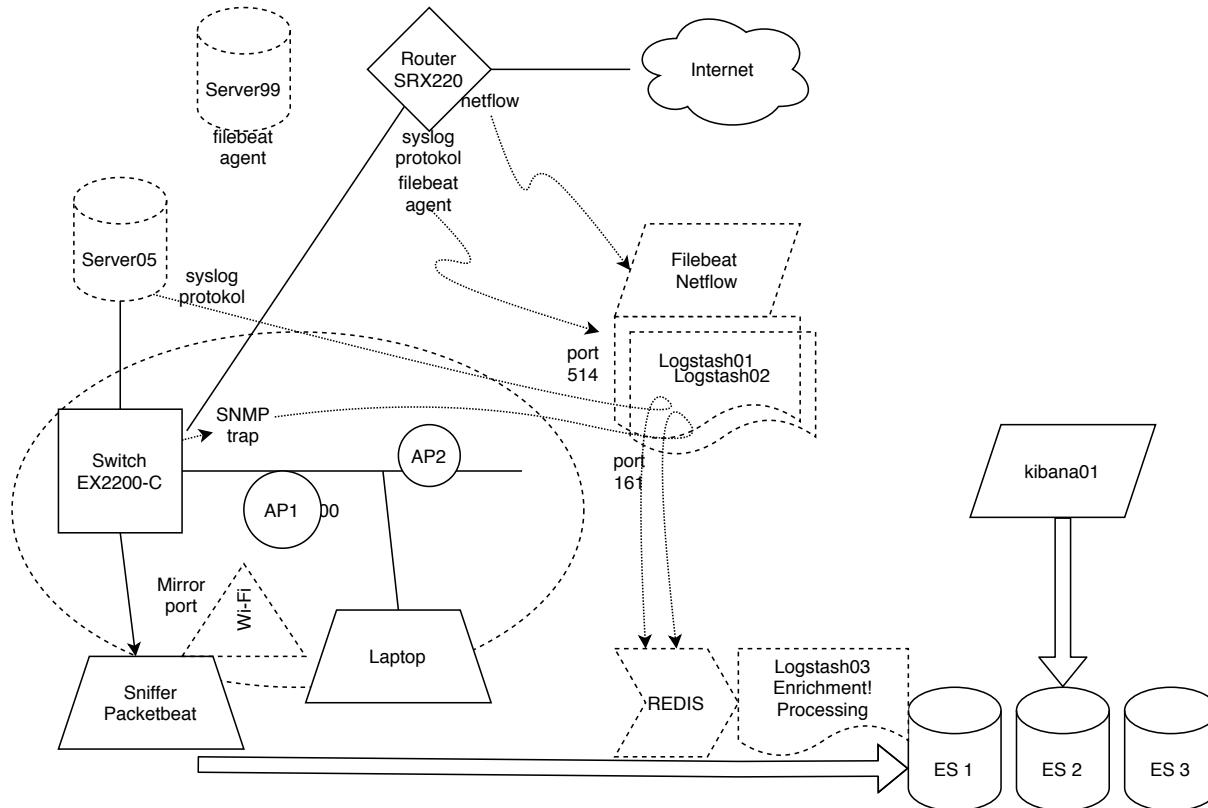
```
filter {
  if [type] == "syslog" {
    grok {
      match => { "message" => "%{SYSLOGTIMESTAMP:syslog_timestamp}
%{SYSLOGHOST:syslog_hostname} %{DATA:syslog_program}
(?:\[%{POSINT:syslog_pid}\])?: %{GREEDYDATA:syslog_message}" }
      add_field => [ "received_at", "%{@timestamp}" ]
      add_field => [ "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](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

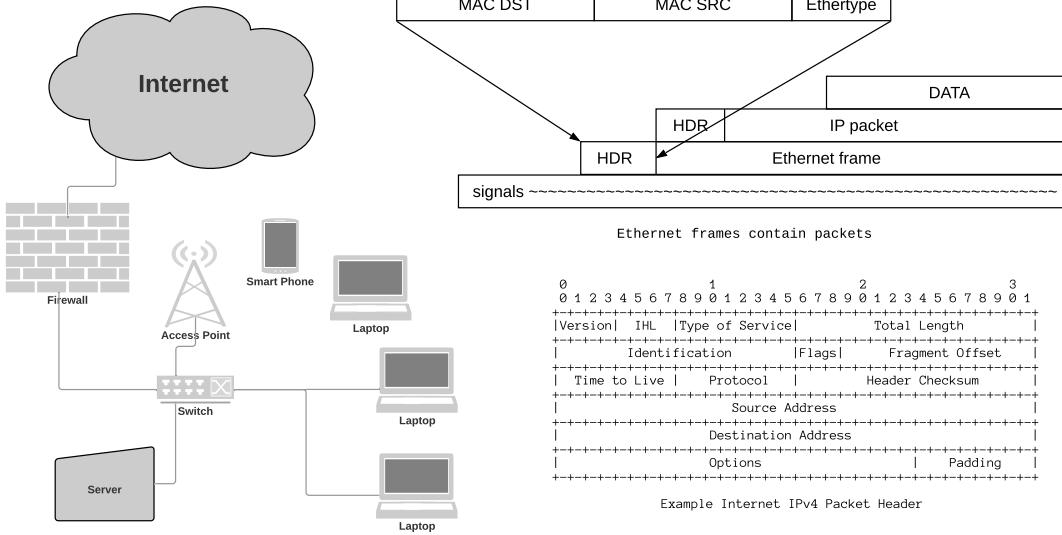


You will probably need help from:

- 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



- 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

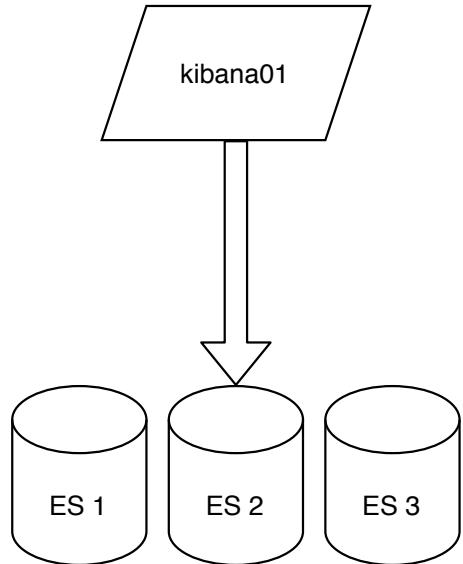


We will use the tools presented during the course:

- 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

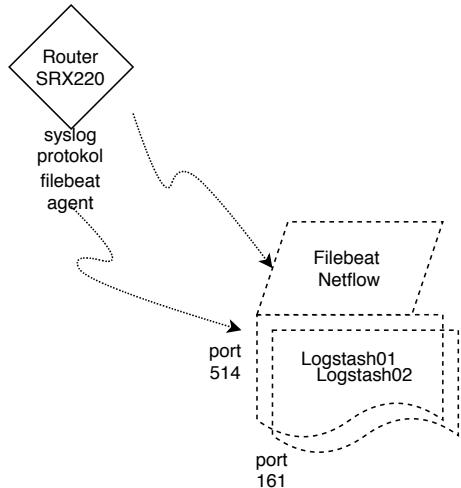
The setup discussed here would be a good proof of concept, and be valuable almost immediately

# Elasticsearch



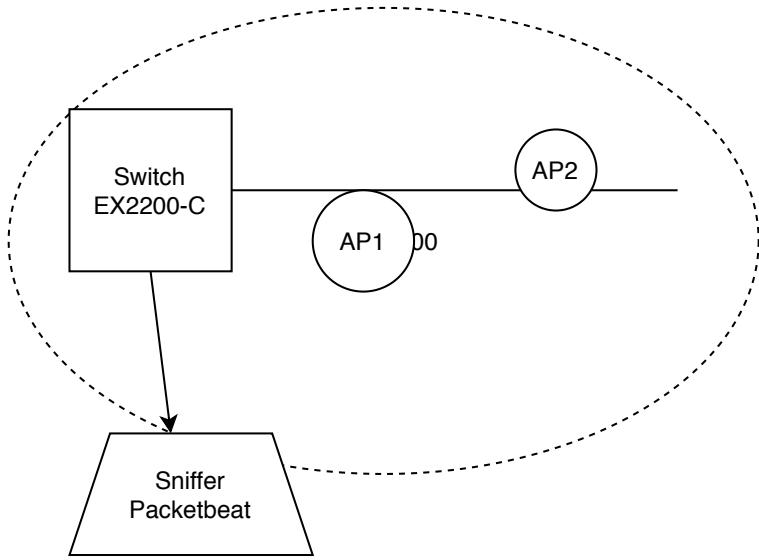
- We plan to build a basic cluster with Elasticsearch, latest stable
- Multiple ES nodes for easier upgrade, redundancy and performance
- Each have 200Gb disk and 16Gb memory allocated

# Logstash – syslog and SNMP trap receiver



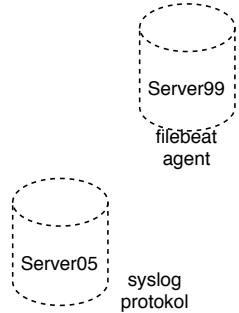
- We have network devices which can only send syslog and SNMP trap – *push events from the network*
- So enable inputs: snmptrap, tcp, udp and use UFW to redirect ports
- We have made two servers, which use VRRP to have a common address

# 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

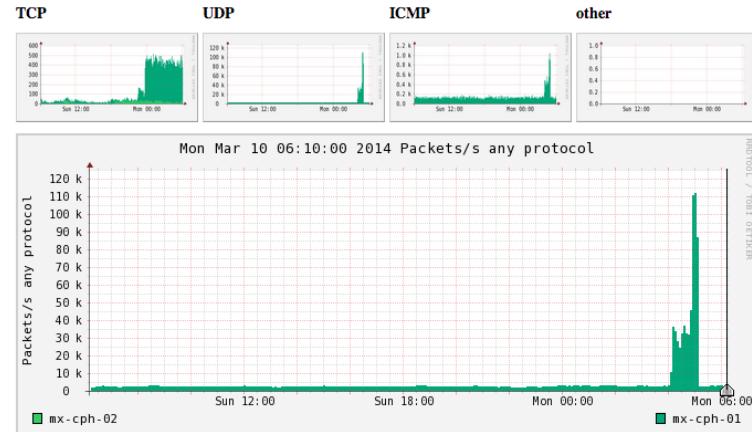


- We told the server and application people to use Filebeat and Syslog
- The Linux people decided to use syslog
- Windows servers use Filebeat <https://www.elastic.co/beats/filebeat>
- All of them send to the Logstash instances

# Baseline



Profile: DDoS



- 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>

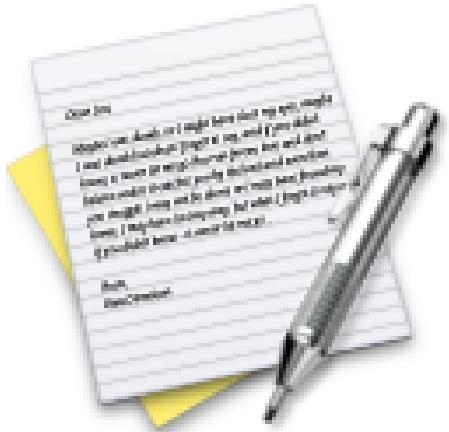
## Alerting everywhere



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 ES 7 – 30min

which is number **23** in the exercise PDF.

## 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-elasticsearch-7-10-0-correlation-cloud-visibility-detectio>

# Automatic reporting: tcpflow

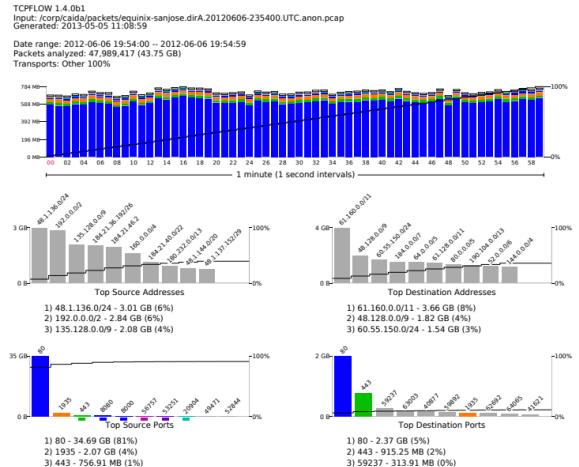


Figure 1: *tcpflow*'s one-page visualization. The color key for the timeline stacked bargraph is presented in the source and destination graphs. Each graph includes a CDF. Here we show the CAIDA equinix-sanjose.dirA.20120606-235400.UTC.anon.pcap capture

- <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>

## 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