Finding Needles in Haystacks (the size of countries)

Michael Baker
@cloudjunky
Ruxcon - October 2012

packetloop

Acknowledgements

- David Turnbull @dsturnbull
- Gerald Kaszuba @gakman
- Packetpig Committers

Two Rules

The Landscape.

Exhibit A

- CVE-2011-3192 - “Apache Killer”
- auxiliary/dos/http/apache_range_dos 2011-08-19 normal Apache Range header DoS (Apache Killer)
- Snort 1:19825
- /Range\s*\x3A\s*bytes=([\d\x2D]+\x2C){50}/Hsmi
- /Range\s*\x3A\s*bytes=([\d\x2D]+[\x2C\s]*){50}/Hsmi

Prevention Fails.

Detection is the key.

NSM - "focused on providing an intrusion analyst with the best possible information in the shortest amount of time" - NSMWiki

Network Security Monitoring

- Advocates focus on detection and that prevention will fail.
- Believes in inventoried and defensible networks.
- Build entropy from alert (attack) information.
- Provide analysts with accurate information as fast as possible.

Tools collect.

People analyze.

Network Security Monitoring

- Squil
- Argus
- Flowgrep
- Snort and Suricata
- Bro
- Network Miner
- Netwitness

It's all about Context.

Context

- Enriched information, not just IP Addresses.
- Additional intelligence on attackers.
- Allow you to perform detective work
- What if? Branch analysis and exploring data.
- Providing full fidelity and full context quickly.

Full Packet Capture

- Complete record of all network data.
- Provides the highest fidelity to analysts.
- Only way to really understand subtle, targeted attacks.
- Play, pause and rewind your network.
- No need to have a specific logging setup.

NSM + FPC

> % OPTIONS

ntop

Using n2disk for 10 Gbit line-rate packet-to-disk

on OCTOBER 14, 2012

Packet-to-disk is the ability to dump network packets to disk. This activity is important for implementing a sort of "network time machine" so that when something unexpected happens, you have the ability to access the raw packets and thus inspect the cause of the problems. Implementing efficient packet-to-disk requires high-speed packet capture, speedy disks, and efficient packet dump software.

We started to work on this field, a few years ago when creating a packet-to-disk application for 1 Gbit networks, named *n2disk*. Today we are introducing the second generation of n2disk that has been further optimised for 10 Gbit networks. Leveraging on PF_RING DNA, n2disk can dump packets on disk using the industry-standard pcap format at 10 Gbit line rate, minimal size packets. All you need to have is a fast storage system and an adequate system to run n2disk on. As you can read on the n2disk home page, we have the ability to:

- Filter packets during capture using BPF-like filters.
- Dump packets with nano-second timestamps (precise timestamping card required such as Silicom 10G timestamp adapter).
- Index packets on the fly, during packet capture, for fast packet retrieval.
- Search disk-stored packets within a specified time-boundary, using BPF-like filters leveraging on the n2disk packet search companion tools.

Unlike costly proprietary packet-to-disk solutions, n2disk can run on commodity hardware using DNA-aware network adapters. Contrary to the common belief that packet-to-disk solutions are expensive and based on proprietary (i.e. non-pcap) dump formats, n2disk demonstrates that this statement is no longer true making packet-to-disk a commodity activity.

For more information about n2disk features and configuration options, please refer to the n2disk home page and n2disk User's Guide. Those who are looking for an affordable turn-key packet-to-disk solution, can instead have a look at the nBox recorder.

bit.ly/RdrI6M

“The difficulty shifts from traffic collection to traffic analysis. If you can store hundreds of gigabytes of traffic per day, how do you make sense of it?”

- Richard Bejtlich

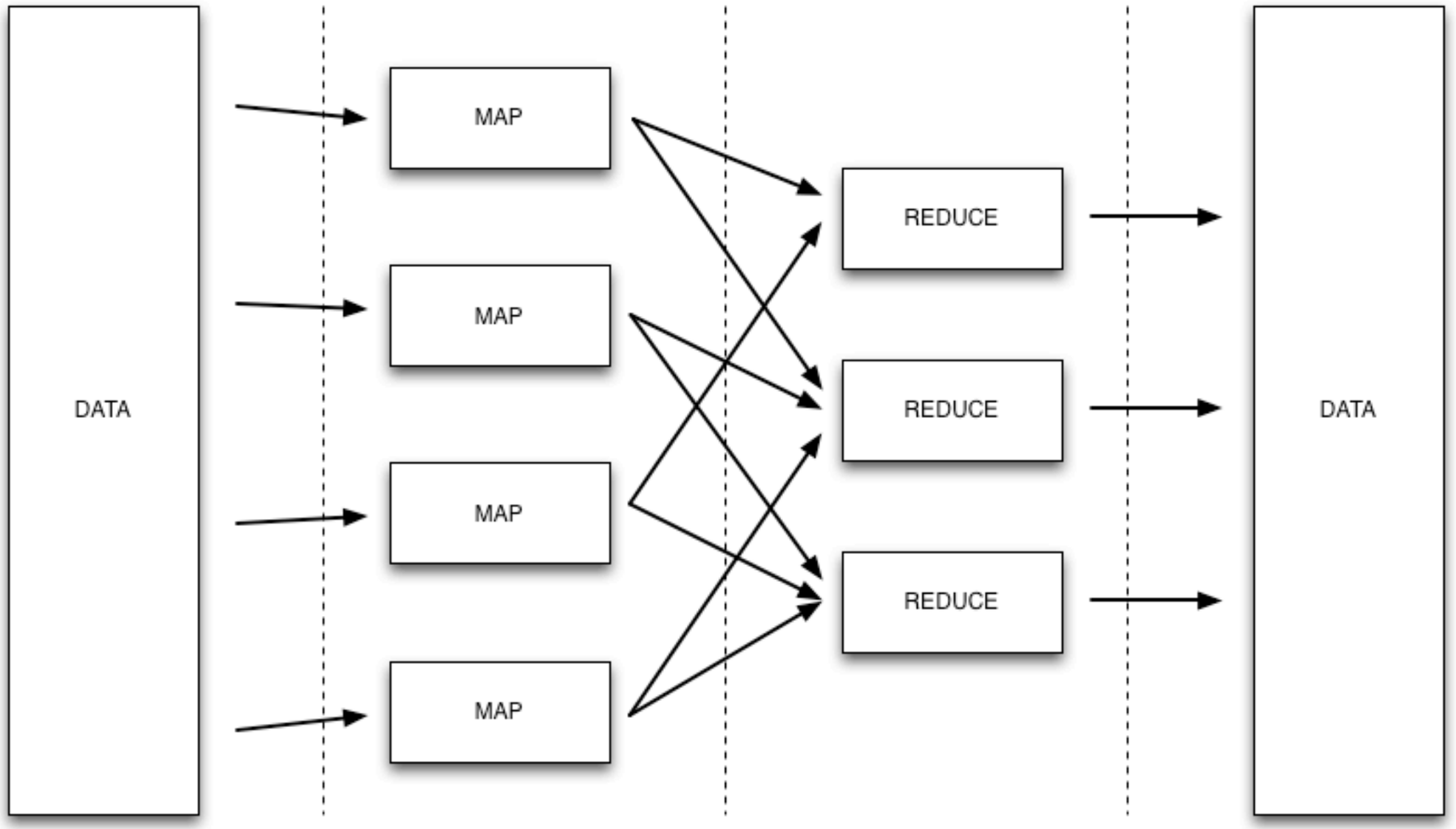
Big Data is a collection of data sets so large and complex that it becomes difficult to process using on-hand database management tools. The challenges include capture, curation, storage, search, sharing, analysis, and visualization.

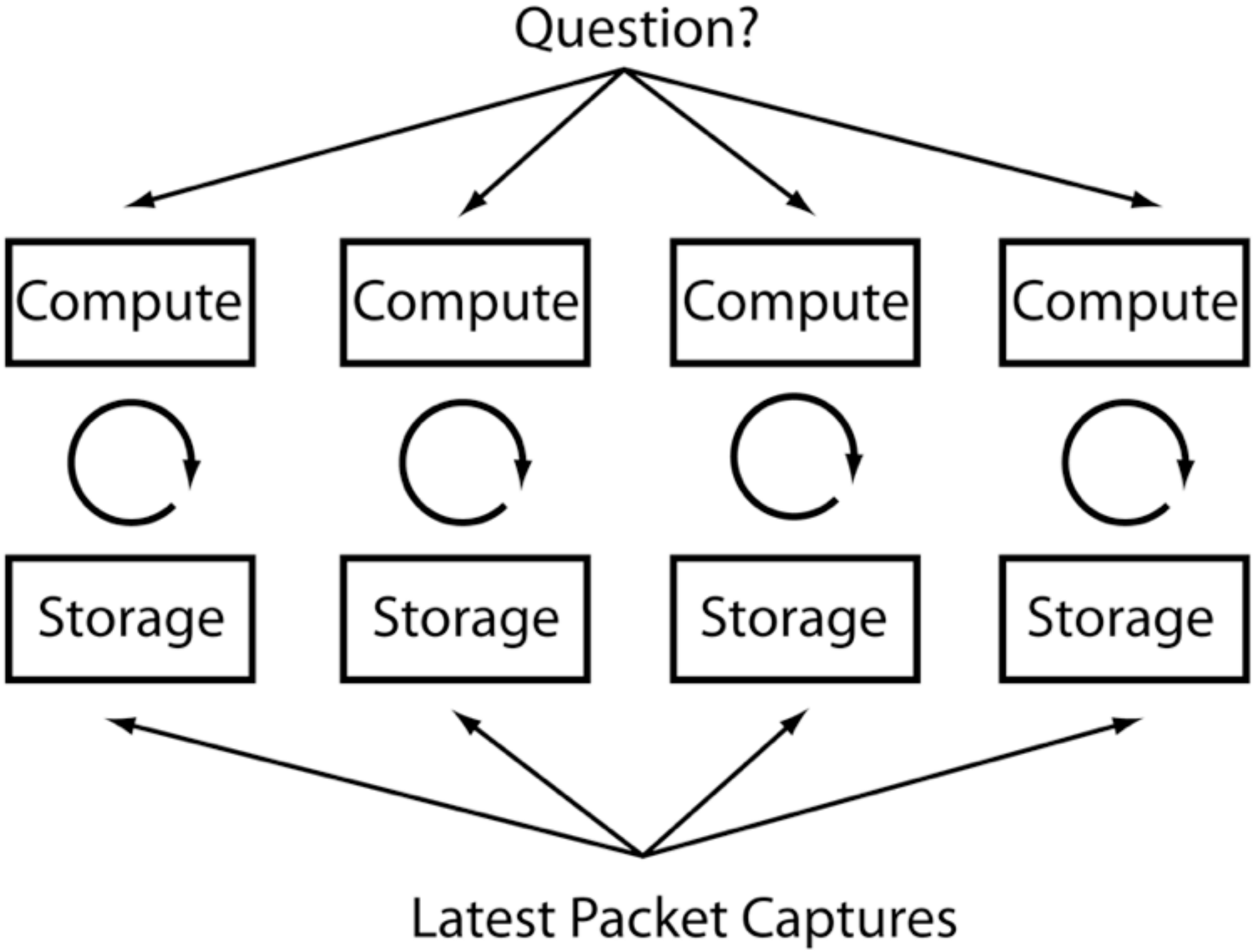
- Wikipedia

Big Data

- Cloud - Elastic compute and Cheap Storage
- Map Reduce - parallel computation
- Pig, Hive - avoid writing M/R
- NoSQL - Cassandra and Mongo

Map Reduce

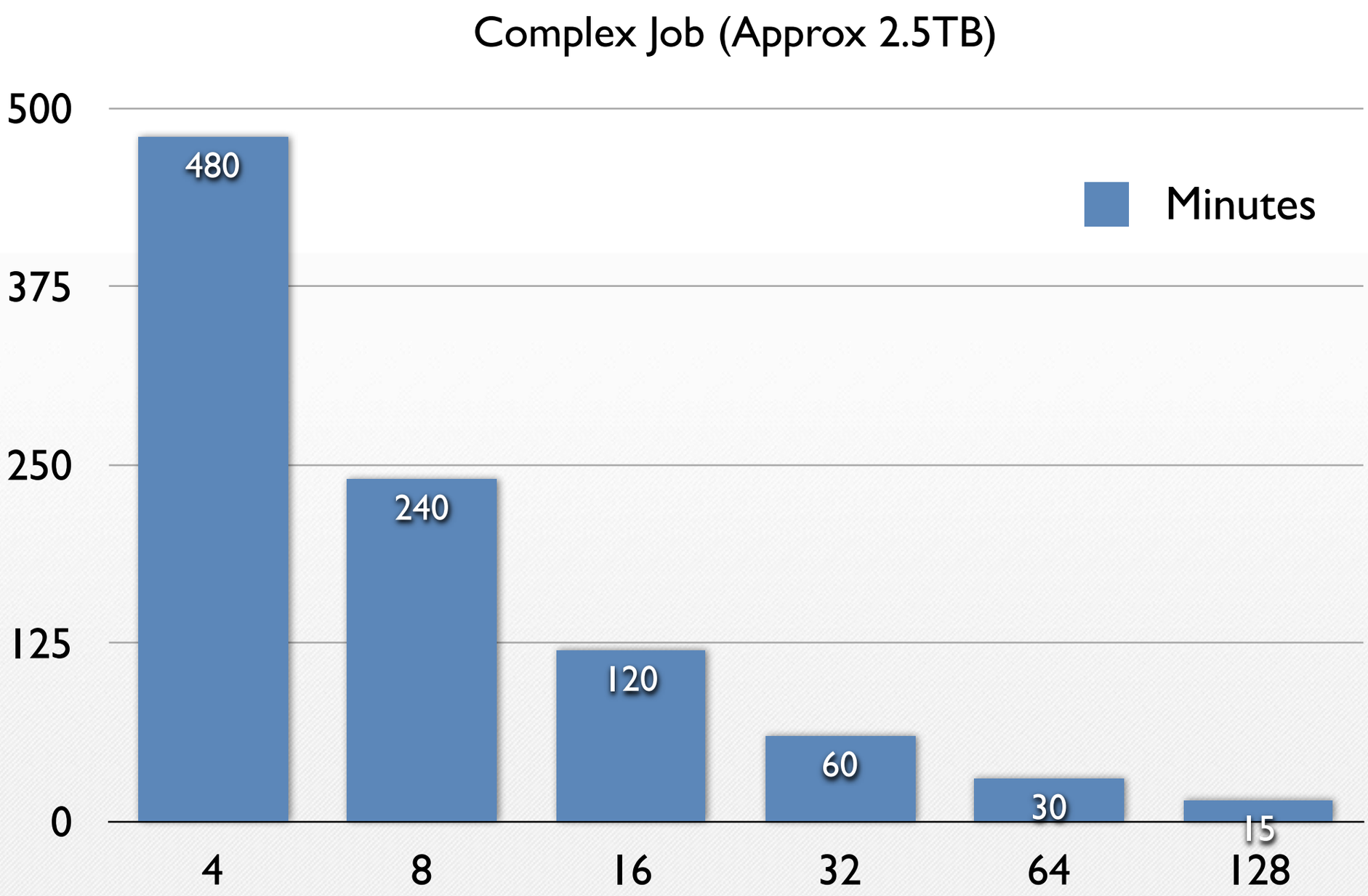



Question?
Compute
Compute
Compute
Compute
Storage
Storage
Storage
Storage
Latest Packet Captures

Big Data Scale

- I want to ask a 2.5TB question
  - Process 2.5TB, 8 hours, 4 Compute units.
  - Process 2.5TB, 4 hours , 8 Compute units.
  - Process 2.5TB, 2 hours, 16 Compute units.
  - Process 2.5TB, 1 hour, 32 Compute units.
  - Process 2.5 TB, 30 minutes, 64 Compute units.
  - Process 2.5 TB , 15 minutes, 128 Compute units.
- Scale my compute to answer my question.

Big Data Scale



History

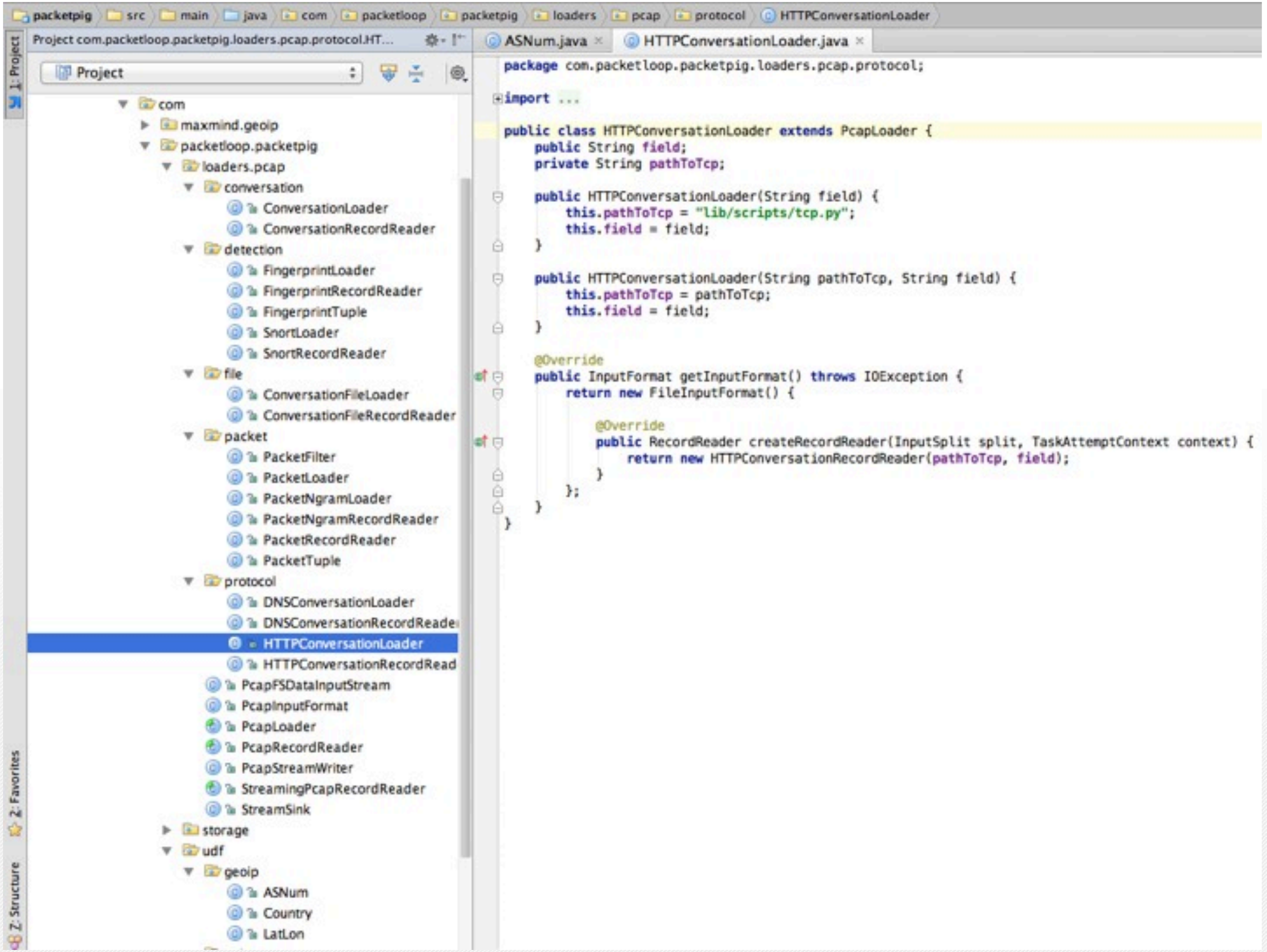
- Google Map Reduce Whitepaper (2004)
- Google File System Whitepaper (2003)
- Hadoop is an Apache Project for M/R (2007)
- Hadoop File System is a distributed file system for Hadoop nodes (2007)
- Pig is a data analysis language to ease the creation of Map / Reduce jobs that run on Hadoop Clusters (2008)

@packetpig

- @packetpig = Packets (FPC) + Pig
- Pig uses a data flow language called Pig Latin.
- Executes Map/Reduce Jobs over Hadoop Clusters.
- Works identically on-premise or in the cloud (Amazon's EMR)

Features

- Full access to IP packets at scale.
- Threat Analysis (Snort)
- Traffic Analysis.
- Flow-based deep packet inspection.
- Geo-Location
- Passive OS Detection (p0f)
- File Dissection


packetpig
src
main
java
com
packetloop
packetpig
loaders
pcap
protocol
HTTPConversationLoader
Project com.packetloop.packetpig.loaders.pcap.protocol.HT...
ASNum.java
HTTPConversationLoader.java
Project
com
maxmind.geoip
packetloop.packetpig
loaders.pcap
conversation
ConversationLoader
ConversationRecordReader
detection
FingerprintLoader
FingerprintRecordReader
FingerprintTuple
SnortLoader
SnortRecordReader
file
ConversationFileLoader
ConversationFileRecordReader
packet
PacketFilter
PacketLoader
PacketNgramLoader
PacketNgramRecordReader
PacketRecordReader
PacketTuple
protocol
DNSConversationLoader
DNSConversationRecordReade
HTTPConversationLoader
HTTPConversationRecordRead
PcapFSDataInputStream
PcapInputFormat
PcapLoader
PcapRecordReader
PcapStreamWriter
StreamingPcapRecordReader
StreamSink
storage
udf
geoip
ASNum
Country
LatLon
1: Project
2: Favorites
Z: Structure
package com.packetloop.packetpig.loaders.pcap.protocol;
import ...
public class HTTPConversationLoader extends PcapLoader {
public String field;
private String pathToTcp;
public HTTPConversationLoader(String field) {
this.pathToTcp = "lib/scripts/tcp.py";
this.field = field;
}
public HTTPConversationLoader(String pathToTcp, String field) {
this.pathToTcp = pathToTcp;
this.field = field;
}
@Override
public InputFormat getInputFormat() throws IOException {
return new FileInputFormat() {
@Override
public RecordReader createRecordReader(InputSplit split, TaskAttemptContext context) {
return new HTTPConversationRecordReader(pathToTcp, field);
}
};
}
}

Finding Zero Days

Worth a ~~coffee~~ JD?

- Motivation
- Time window
- Attacker
- Attack type
- Target
- Obfuscated
- Anonymised

Attacker Information

File Extraction

Big Data

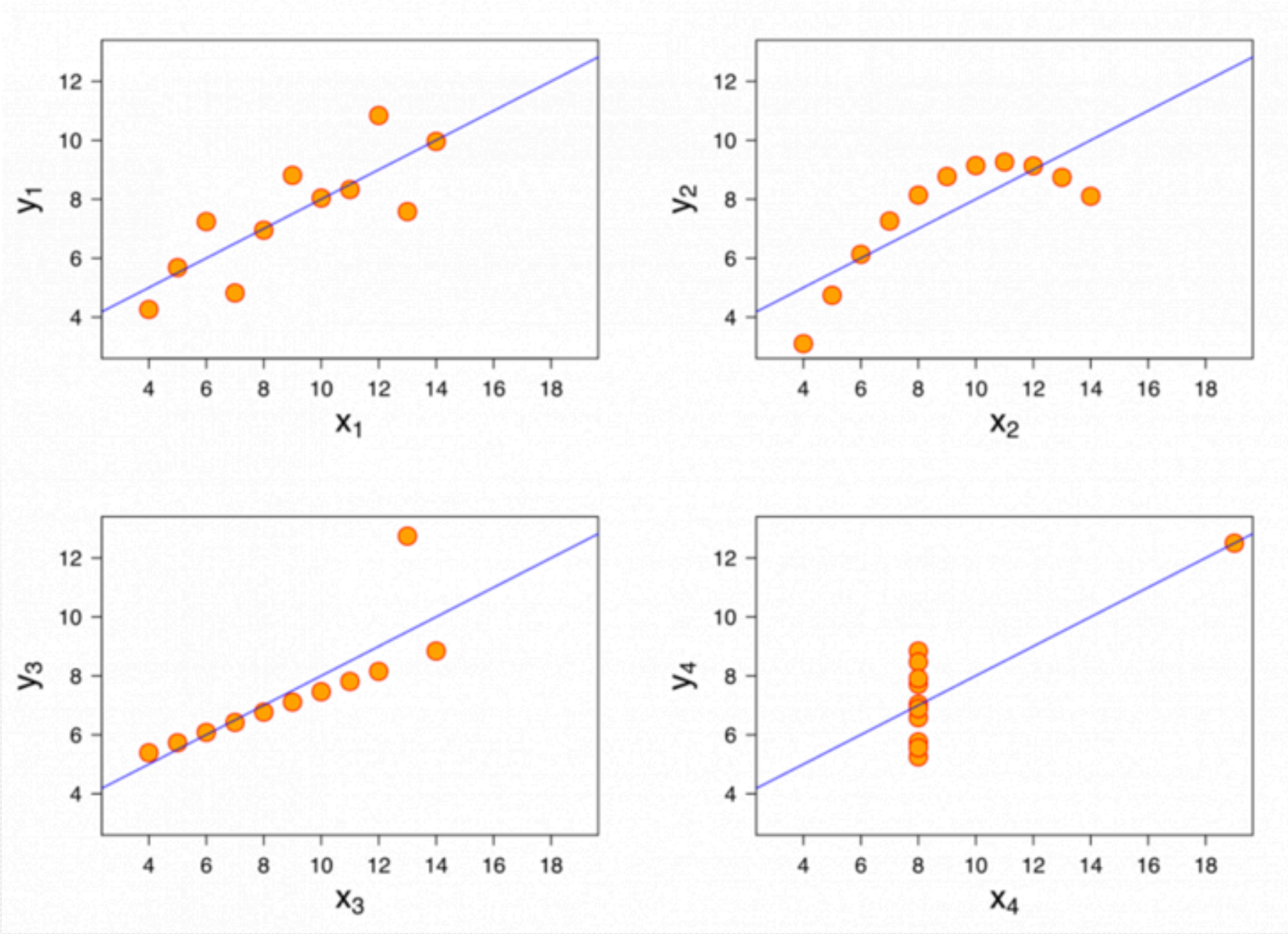
Security Analytics

Anscombe's Quartet

I		II		III		IV	
x	y	x	y	x	y	x	y
0.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

Source: http://en.wikipedia.org/wiki/Anscombe%27s_quartet

Anscombe's Quartet



Source: http://visual.ly/anscombes-quartet

Big Data Security Analytics

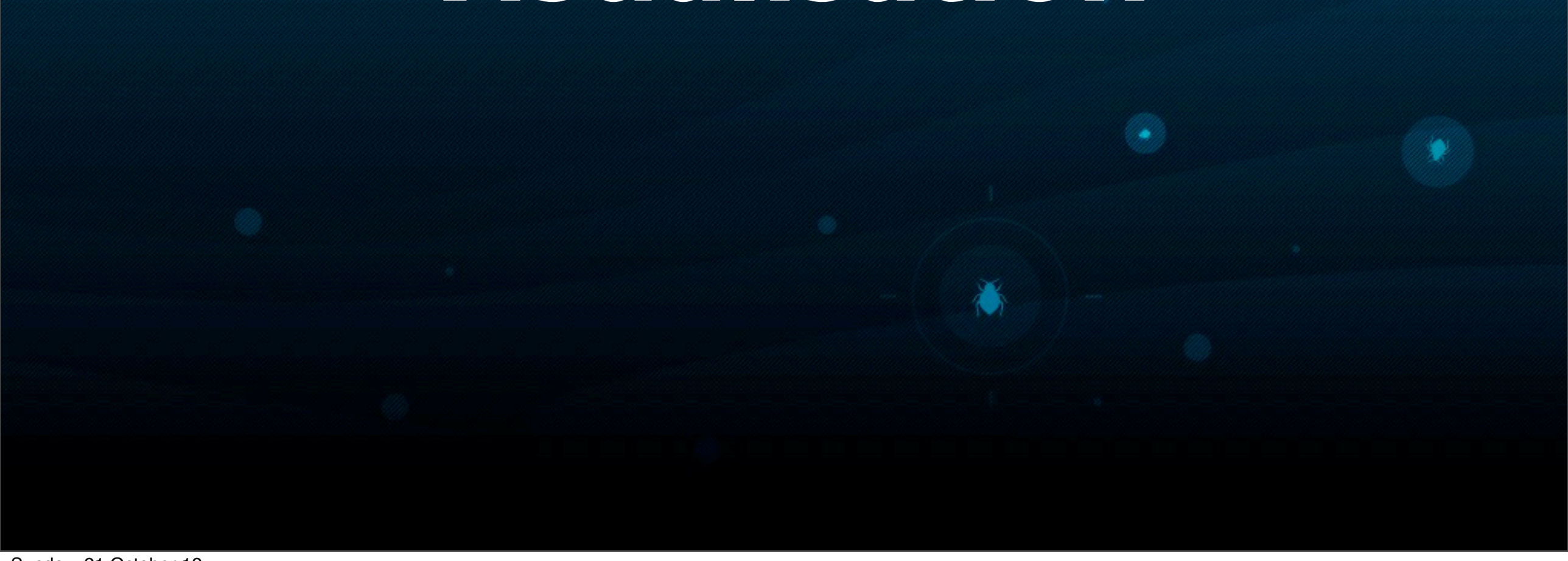
- Visualization
- Fidelity
- Interaction
- Outlier Detection
- Attacker Profiling
- Enrichment
- Transform
- Prediction and Probability
- Intelligence sharing
- Statistical Analysis
- Feature Extraction
- Machine Learning

This is not SIEM.

Not SIEM

- Full Fidelity
- Explore and explain the data (evidence).
- Play, Pause and Rewind.
- Blink and you miss it technology.
- No aggregation.
- No parsers or complex integration.
- Clear intelligence.

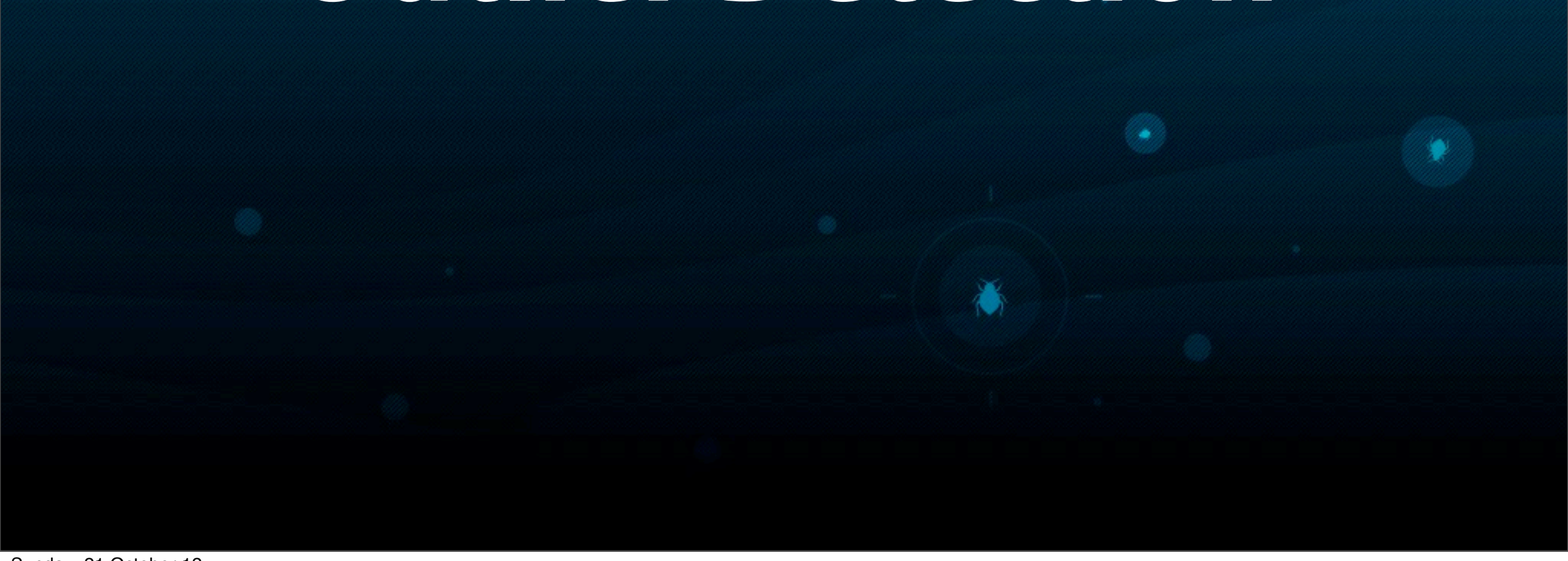
Visualisation



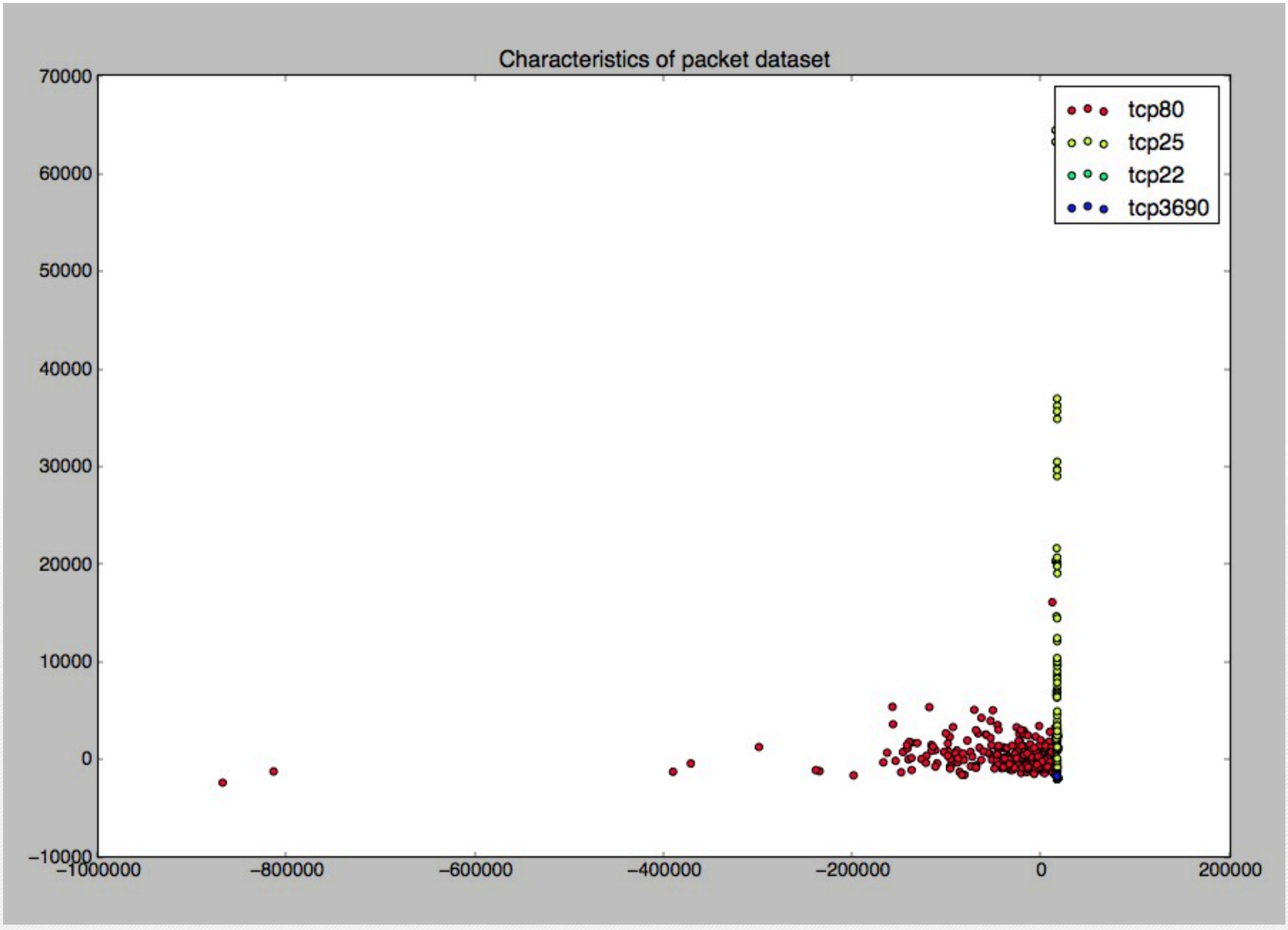
Full HD

Play, Pause, Rewind

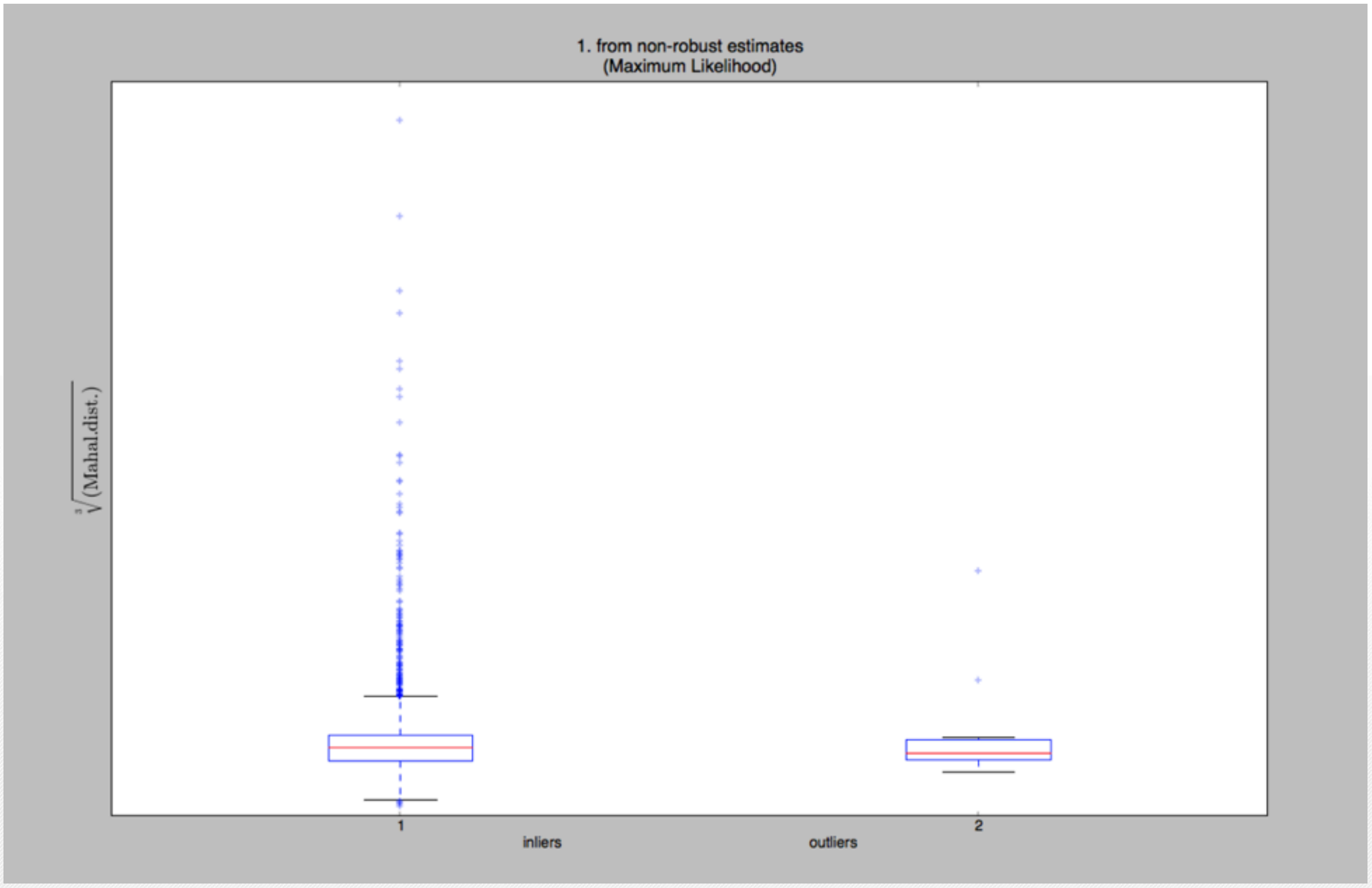
Outlier Detection



Classification


Characteristics of packet dataset
tcp80
tcp25
tcp22
tcp3690
70000
60000
50000
40000
30000
20000
10000
0
−10000
−1000000
−800000
−600000
−400000
−200000
0
200000

Novelty and Outliers


1. from non-robust estimates
(Maximum Likelihood)
$\sqrt[3]{(\text{Mahal.dist.})}$
1
2
inliers
outliers

Entropy and Covert Channels

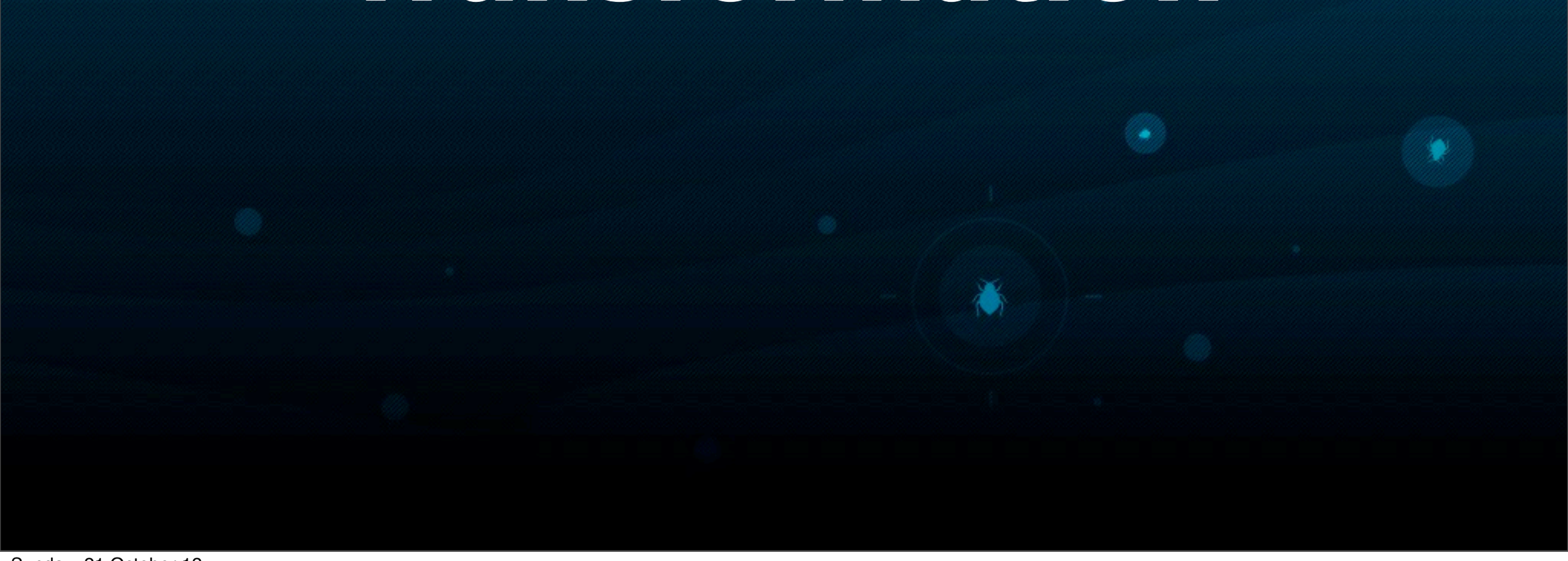
Enrichment

Geocoding

TOR

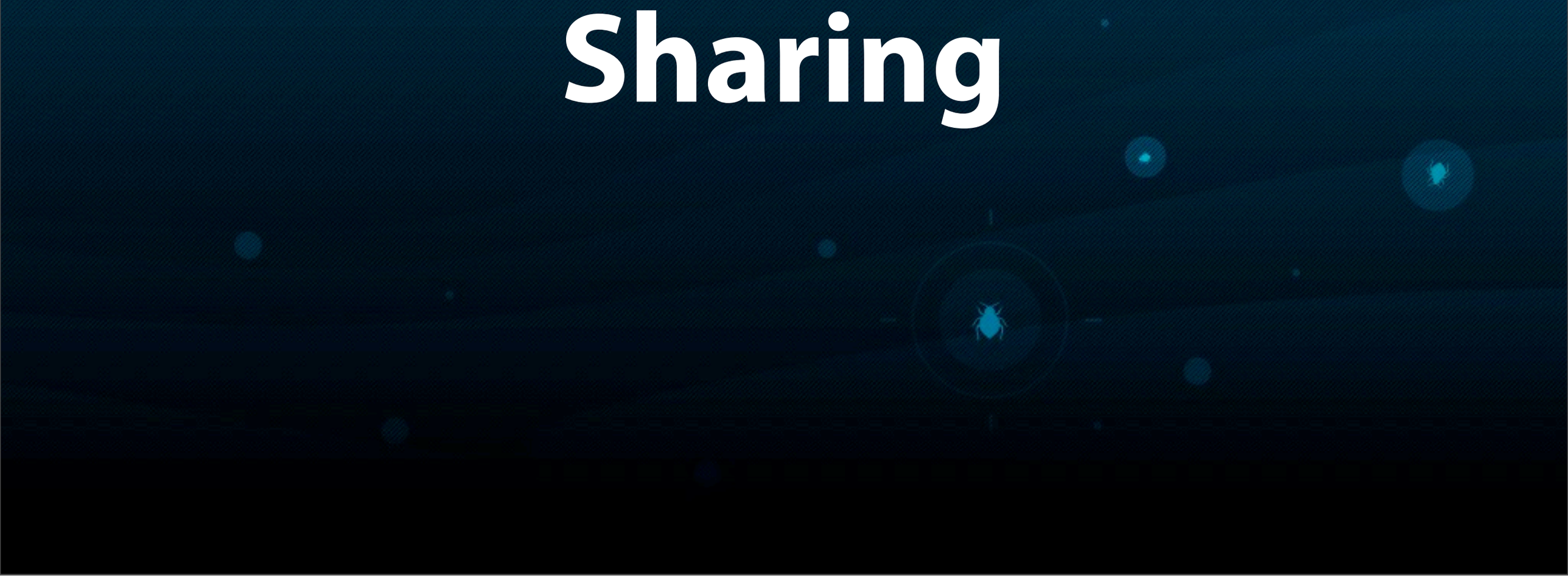
Torrent Triangulation

Transformation



Network Graphs and Relationships

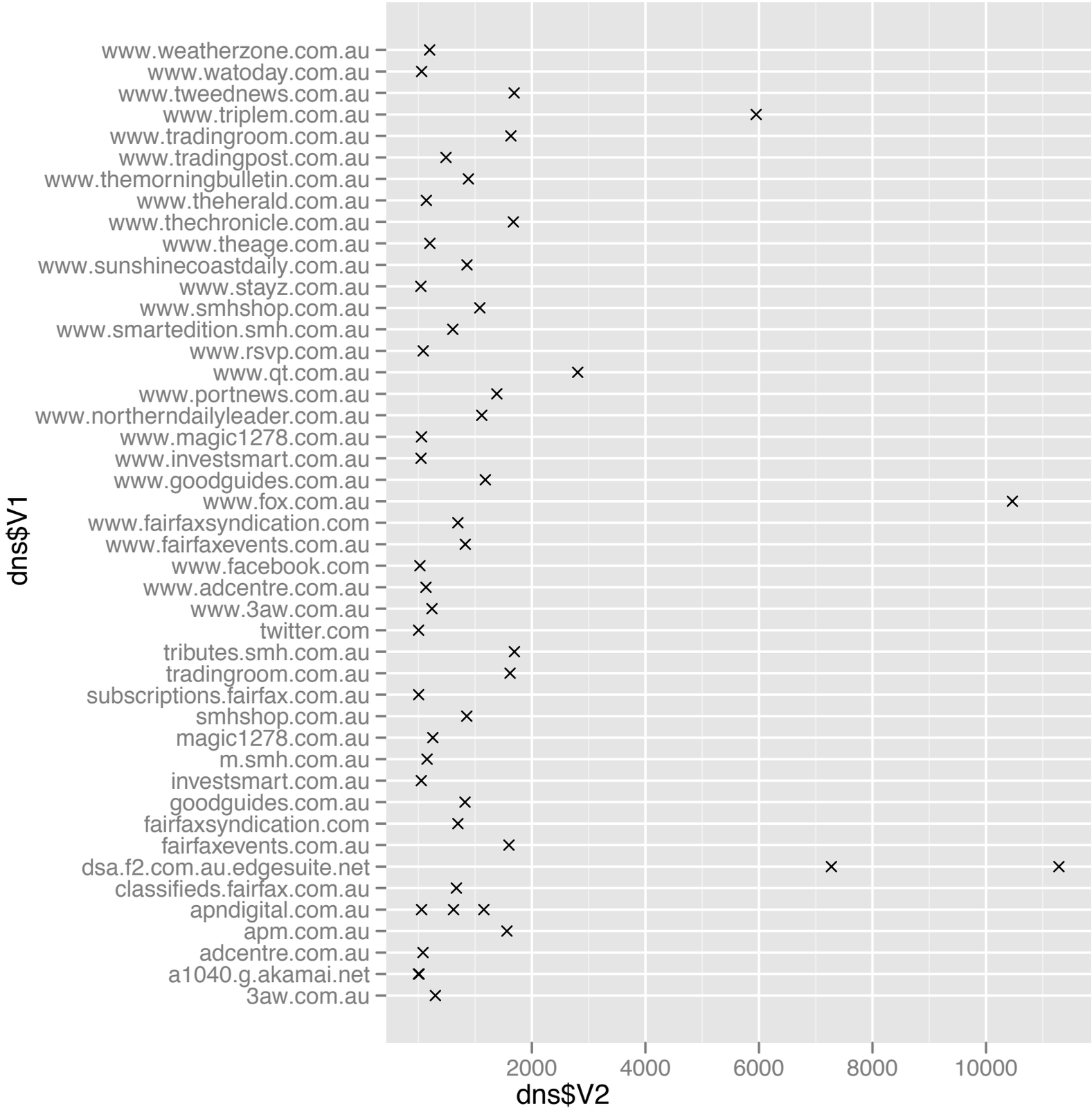
Intelligence and Metric Sharing



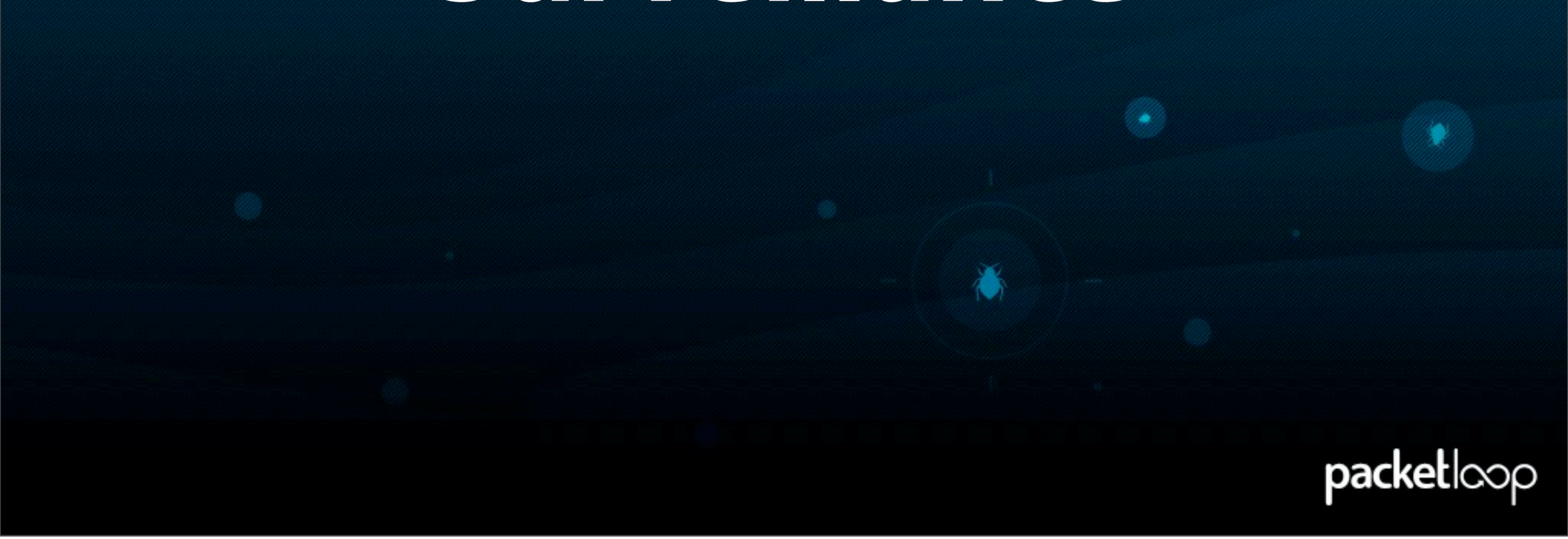
Indicators of Compromise

- OpenIOC and CyBOX
  - Open Indicators of Compromise (XML)
  - Host and Network Indicators of Compromise
- Fork a github repository
  - Execute Packetpig scripts that find bad things and visualise them

DNS and Malware


www.weatherzone.com.au
www.watoday.com.au
www.tweednews.com.au
www.triplem.com.au
www.tradingroom.com.au
www.tradingpost.com.au
www.themorningbulletin.com.au
www.theherald.com.au
www.thechronicle.com.au
www.theage.com.au
www.sunshinecoastdaily.com.au
www.stayz.com.au
www.smhshop.com.au
www.smartedition.smh.com.au
www.rsvp.com.au
www.qt.com.au
www.portnews.com.au
www.northerndailyleader.com.au
www.magic1278.com.au
www.investsmart.com.au
www.goodguides.com.au
www.fox.com.au
www.fairfaxsyndication.com
www.fairfaxevents.com.au
www.facebook.com
www.adcentre.com.au
www.3aw.com.au
twitter.com
tributes.smh.com.au
tradingroom.com.au
subscriptions.fairfax.com.au
smhshop.com.au
magic1278.com.au
m.smh.com.au
investsmart.com.au
goodguides.com.au
fairfaxsyndication.com
fairfaxevents.com.au
dsa.f2.com.au.edgesuite.net
classifieds.fairfax.com.au
apndigital.com.au
apm.com.au
adcentre.com.au
a1040.g.akamai.net
3aw.com.au
dns$V1
2000
4000
6000
8000
10000
dns$V2

Analytics or Surveillance



Big Brother on a budget: How

arstechnica.com/information-technology/2012/09/big-brother-meets-big-data-the-next-wave-in-net-surveillance-tech/

ars technica

MAIN MENU · MY STORIES: 25 · FORUMS · SUBSCRIBE NOW · SEARCH · LOG IN

TECHNOLOGY LAB / INFORMATION TECHNOLOGY

Big Brother on a budget: How Internet surveillance got so cheap

Deep packet inspection, petabyte-scale analytics create a "CCTV for networks."

by **Sean Gallagher** - Sept 28 2012, 0:50am EST

BIG DATA · NATIONAL SECURITY · PRIVACY · 61

The surveillance powers of CCTV are coming to a network near you, thanks to deep packet inspection and big data analytics.

ogglog

This feature originally ran on August 28, 2012.

bit.ly/TzcSq8

TOP FEATURE STORY

FEATURE STORY (2 PAGES)

Review: 7th-generation iPod nano does little to excite

Is this just last year's device with a bigger screen?

114

STAY IN THE KNOW WITH

LATEST NEWS

ZERO-DAY PATCHING

Microsoft already updating Office RT Preview to final version

FREE MONEY

Motorola lists the dated Android phones it will buy back for $100

HOLD ON FOR ONE MORE WEEK

Microsoft teases Metro version of Skype for Windows 8 release

No, Minnesota did not kick Coursera out of the state

FALLING LIKE A ROCK

Questions?

@packetpig @packetloop

Thank you!

http://blog.packetloop.com

packetloop