"Capture the Flag" Data Capture Experiences

George Jones <gmj@cert.org> and
Paul Conrad <jpconrad@cert.org>
Network Situational Awareness Group
CERT

This material is based upon work funded and supported by Department of Homeland Security under Contract No. FA8721-05-C-0003 with Carnegie Mellon University for the operation of the Software Engineering Institute, a federally funded research and development center sponsored by the United States Department of Defense.

Abstract

There is a need for common, accessible data sets for use in security testing, training, tuning of systems and experimentation. Live captures of data from security conferences and Capture The Flag (CTF) exercises offer one possible way of obtaining such data. In the past year CERT has captured at three such events. This talk presents our experiences, lists benefits, challenges, and thoughts on possible future captures.

The Problems

Good network security data is hard to find.

Common, accessible sets of network data for use in security testing, training, tuning of systems and experimentation are few and far between. In part this is due to some well-founded concerns organizations have:

- Exposure of sensitive data
- Legal and regulatory issues
- Negative publicity
- Providing information to bad actors

The Problems

Problems caused by a lack of data

There are a number of problems caused by the lack of good data:

- Engineers can not validate heuristics for detecting malicious network-based behaviors without common examples to test against.
  - e.g. IDS signatures
- Training is less realistic without real data.
- Research results are not reproducible.
  - Science built on testable, verifiable results.
  - Need common, shareable data.

The Problems

Problems with data that is available

There are a number of problems with data that is available:

- Synthetic data not as effective for product testing.
- Anonymized traces
  - loose critical information.
  - can not be correlated.
  - Content and attribution provide context e.g. “Paul attacking me, vs. Country X attcking Industry Y”
- Available data is out of date.
- The amount of “badness” in data sets is unknown.
- The specific instances of “badness” is unknown.

One Possible Solution

Live Data Capture Opportunities

The following are some options for data capture that might be sharable:

- Security conferences.
- Training exercises
- Capture The Flag (CTF) exercises.

One Possible Solution

Benefits of the data

- You can get it.
- It is current.
- Can contain newer/recent attack traffic (0-days)
- It is not anonymized.
- It is not synthetic.
  - Usually real people on red/blue teams.
- Fewer privacy concerns.
- Easier to share, use for training and experimentation.

One Possible Solution

Issues with the data

- "find the needle in the needle stack."
  - High signal to noise. No background traffic. No ISP "noise". No unknown hosts. Everything is "good" or "bad", no "gray".
  - Not great for scientific experiments.
- CTF Networks are not realistic
  - Small.
  - Built for hacking contests, not production.
  - Amateur design, construction, administration.
  - Not set up, run for "normal" business purposes.
  - Can still miss intra-VLAN traffic.

CERT Experience In Data Capture

Parallels to network defense underlined

- There are a number of parallels to the in-the-small captures described in the rest of this talk and the capture and analysis tasks for larger scale network defense.
- Items in underlined in the rest of this presentation have parallels in network defense.

CERT Experience In Data Capture

Possible goals of data capture at CTF exercises

- Obtain Data for development testing, training, analysis, research, etc.
- Gain experience
  - With tools
  - With analysis

CERT Experience In Data Capture

Possible goals for participating in CTF exercises

- Learn new 0-days (if Red Teams use them).
- Networking (the in-your-face kind).
- Exposure/Marketing/Recruiting.

CERT Experience In Data Capture

Capture at 3 events (first two)

- 2011 and 2012 "a local security conference"
  - Conference traffic + CTF.
  - Multiple networks, layer 2 segmenting, wireless, internet access, CTF network, live video streaming.
  - CTF hosted in cloud, missed most traffic.
  - Span/tap off central switch. All VLANs.
  - 2011: captured ~= 375 GB pcap.
  - 2012: captured ~= 169 GB pcap.

CERT Experience In Data Capture

Capture at 3 events (third)

- 2012 CCDC http://www.midatlanticccdc.org/CCDC/
  - CTF: all red/blue team traffic.
  - Blue Teams defending fake medical network.
  - Span/tap off central switch.
  - Captured ~= 16GB pcap.

CERT Experience In Data Capture

To-Do Before The Event

- Contact organizers, inquire about capture opportunities.
- Line up capture and analysis hardware.
- Get as much info as you can ahead of time.
- Prep a day or two before.
- Show up a day early for setup.
  - You learn things by talking to people before the chaos starts.

CERT Experience In Data Capture

Info You Would Like To Have Before The Event

- Capture interface types.
- Expected data rates and duration.
- Topology info.
  - Network maps.
  - Addressing plans (layer 2, layer 3).
  - Firewall, Router, Switch configs.
  - Where are red and blue teams?
  - Where is the capture?
  - How many capture points?

CERT Experience In Data Capture

More Info You Would Like To Have Before The Event

- More Topology Info.
  - What can/can't you see from there?
  - What other devices are on the net? Is there Internet connectivity ? Is there wireless? What other logs will be available?
- Asset info.
- Rules of the contest.
- Power, network drop info.
- Access to capture location (public/guarded, 24x7 ?)
- Hours of contest.

CERT Experience In Data Capture

Even More Info You Would Like To Have

- Rules for retention, sharing, use of the data after the event.
- Contacts for sharing results at end/after event.

CERT Experience In Data Capture

Hardware To Bring

- Boxes for Capture and analysis.
  - Make sure interfaces are the right kind.
  - Make sure disks are large enough.
  - Pre-load and test tools and configs.
  - Monitors, keyboards and mice if needed.
- Cables.
- Switches.
- USB Drives.
- MIFI Cards for net access.

CERT Experience In Data Capture

More Hardware To Bring

- Sacrificial laptops for net access.
  - Do you want to use your laptop on a hacker-con net?
  - Load clean before event, wipe after or run Live CD.

CERT Experience In Data Capture

Doing The Capture

- Start before the event goes live.
- Be sure you're getting pcaps.
  - Everything else can be regenerated from these.
- Do analysis during the event if possible.
- Pay attention to what is happening around you.
  - net-ops, monitoring, services, CTF red/blue teams.
- Take notes

CERT Experience In Data Capture

Data Captured and Tools

Here is a list of the data captured and the tools used

Data Capture	Tool	Data Produced
pcap capture	dumpcap	pcaps
netflow generation	YAF + rwflowpack	Netflow
IDS alert generation	Snort + Security Onion	IDS Alerts
application labels	YAF + SiLK	labeled flows: HTTP, DNS, SSH, etc.
Entity Extraction	YAF + super mediator	HTTP, DNS, email, etc.

CERT Experience In Data Capture

Analysis and Tools

Here is a list of the types of analysis and the tools used

Analysis	Tool	Data Used
Top N Lists	SiLK	Netflow
Scan Detection	SiLK	Netflow
Protocol Anomalies	SiLK, Snort, Bro	Netflow, pcaps
Behavioral Anomalies	SiLK	Netflow, pcaps
Volume Graphs	Prism	Netflow
Packet Analysis	Wireshark	pcap

CERT Experience In Data Capture

Analysis Notes

- Top N
  - Protocols
  - Ports
  - Talkers
  - AppLables
- Scans
- Protocol Anomalies (applabel)
- Behavioral Anomalies
- Volume Graphs (prism)
- Snort Hits

CERT Experience In Data Capture

Sharing

- Share results with organizers, other participants during/ after event.
- Share the data publicly if possible.
  - You can get our data at . . . we're still working on that !

CERT Experience In Data Capture

Lessons Learned

- It all comes back to goals (see above).
- Personal interaction is important.
- On-site capture is best.
  - Build trust.
  - More personal interaction.
  - Know more of topology, time-lines, services, etc.
- A working network takes priority over security and monitoring.
- It's hard to focus when events are local.
  - or when you have cell phone/email.

CERT Experience In Data Capture

What is the data good for?

- Testing YAF.
- Testing Snort/SourceFire.
- Testing/Learning other tools.
  - Replay into Security Onion
- Scientifically valid experiments (repeatable, sharable).
  - Possibly. Need more rigor in capture, labeling.
- Training.

Future Work?

Future Work?

- More captures?
  - Training events?
  - More CTF events?
  - Public release?
- More analysis? What about mobile devices?
  - Certain lab architectures miss traffic, such as mobile.
- What about cloud service?
  - Help set up CTF exercises and capture in cloud?

References

Tools

Here are links to some of the tools used referenced:

- CERT NetSA tool suite
  - http://tools.netsa.cert.org
- Drop In Network Observer
  - https://forensics.cert.org/confluence/display/dino/Home
- Security Onion
  - http://securityonion.blogspot.com

References

Public Data Sets

The following are some useful public data sets:

- OpenPacket.org
  - https://www.openpacket.org/
- A Day in the Life of the Internet
  - http://www.caida.org/projects/ditl/
- CAIDA Data Overview
  - http://www.caida.org/data/overview/
- ShmooGroup CCTF at DEFCON
  - http://cctf.shmoo.com/

References

More Public Data Sets

- LBNL/ICSI Enterprise Tracing Project
  - http://www.icir.org/enterprise-tracing/Overview.html
- UMassTraceRepository
  - http://traces.cs.umass.edu/index.php/Network/Network
- Packet Traces from the WIDE backbbone
  - http://mawi.wide.ad.jp/mawi/

References

Conferences

The following are some related conferences:

- USENIX Cyber Security Experimentation and Test (CSET) Workshop
  - http://static.usenix.org/events/cset12
- Internet Measurement Conference
  - http://www.sigcomm.org/events/imc-conference
- SecurityMetrics/MetriCon
  - http://www.securitymetrics.org/content/Wiki.jsp
- CERT LASER 2012
  - http://www.cert.org/laser-workshop/

Questions?

Questions?

- Questions?