

Multi-site Intrusion Detection System (IDS) Correlation

Paul Krystosek, PhD
CIAC (Computer Incident Advisory Capability)
U.S. Department of Energy
Lawrence Livermore National Laboratory
University of California

LLNL Disclaimer

This document was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor the University of California nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial products, process or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or the University of California, and shall not be used for advertising or product endorsement purposes.

CIAC is the U.S. Department of Energy's Computer Incident Advisory Capability. Established in 1989, shortly after the Internet Worm, CIAC provides various computer security services to employees and contractors of the DOE, such as:

- **Incident Handling consulting**
- **Computer Security Information**
- **On-site Workshops**
- **White-hat Audits**

CIAC is located at Lawrence Livermore National Laboratory and is a part of its Computer Security Technology Center. CIAC is also a founding member of FIRST, the Forum of Incident Response and Security Teams, a global organization established to foster cooperation and coordination among computer security teams worldwide.

Reference to any specific commercial product does not necessarily constitute or imply its endorsement, recommendation or favoring by CIAC, the University of California, the United States Department of Energy, or the United States Government.

What is the Department of Energy?

- The U.S. Department of Energy is a federal government agency
- It employs 152,000 people (employees and contractors)
- It consists of
 - 16 National Laboratories
 - ~90 other sites
- It is responsible for
 - Energy policy
 - Scientific research
 - Managing the national laboratory system
 - National security



What is CIAC?

- Computer Security Incident Response Team for DOE
- Started in 1989
- Founding member of FIRST
- 20 team members
- Organized as
 - Operations (Call Center, Analysis, Research)
 - Tools development (NID, SafePatch, TIPPS, IEFT)



About the Project

- We have access to IDS data from about a dozen US Dept of Energy (DOE) sites...
- Collection of Snort sensors
 - Identically configured
 - Uploading to a common point
- From there, transferred to CIAC
- Preprocessed and loaded into Oracle

The questions we attempt to answer include

- What is happening at each facility?
- What common activities are seen at multiple sites?
- Can the common activity be explained?
 - Might there be a cooperative effort involved?
- What do we know about the source IPs whose activity is logged by the IDS sensors?
- How does the activity change over time?
 - Can we spot anomalies in time to do something about them?

Are there existing ways to get the information we want?

- Many companies make systems to correlate data from their own products (ISS RealSecure, and Acid for example)
- At least two companies offer products that will correlate data from various vendors' IDS sensors
 - ArcSight
 - Guarded Net's NeuSecure

Are there existing ways to get the information we want?

- For special cases, yes
 - Analyst's Notebook, Link Notebook
 - On-Line Analytic Processing – OLAP
- They work great within their own limits
 - Link Notebook: ~2,000 records
 - OLAP: several million records
- Remind me to come back to these

What is happening at each site?

- Characterize the activity
 - By source
 - By destination
 - By type
 - By duration
- Spot changes over time

What is happening at each site?

- Start with the easy one
- Basically, Snort alerts are triggered by
 - Scanning or other recon activity
 - Malicious activity such as exploit attempts
 - False alarms
 - Misconfigurations
 - Questionable activity
 - Who knows what else

An exercise in counting

- Count everything in sight
- Count the counts where applicable
- Percentages are useful in addition to absolute counts
 - Can compare different stuff in same report
 - Can compare same stuff in different reports

Basic Counts

- How many “Xs” did we see this week?
 - Countries
 - Source and destination IP addresses
 - Alert Ids and destination ports
- The always popular Top Ten (by volume)
 - Countries
 - Source and destination IP addresses
 - Alert Ids

More complex counts

- Pairs of things
 - Source IP – Destination (IP, Ports, Alerts)
 - In Perl, append one to another and use as a key
 - `$count{addr.",", ".port}++`
 - Sort, and you've got your top ten

More complex counts

- DestinationIP per SourceIP
 - Broad vs deep scan
- SourceIP per DestinationIP
 - Popular facility or targeted

More complex counts

- Duration of activity (number of hours actually)
 - `$count{$sourceip.",",$.dayhour}++`
 - Then count hours per SourceIP
 - The range of values is interesting
- Site counts

What common activities are seen at multiple sites?

- Are the “Top Tens” similar?
 - Between sites?
 - Is the Cross Site report different?
- Are the changes from week to week similar?
- Any trends?

Can the common activity be explained

- What, but not always why
- A simple one would be that a particular Source IP scanned the Internet, hence every site would see the same activity
- What might look like a coordinated scan can have a different explanation

For example

- One day Alert Id 483 rang over 600,000 times
- Not unusual in itself, but it gets better
- It happened at several sites
- At exactly the same time
- At an average rate of 8,000 per second

Example concluded

- It turns out that it was just a ping sweep
- But the Class B that was swept is allocated among several sites
- It was suspicious because a week later Nachi rang the same way
- Of course, we didn't know about Nachi at the time

What do we know about the source Ips?

- They generally do the same thing
 - Over and over and over
- Recurring ones seem to be
 - broad or deep
 - fast or slow
 - continuous or bursty

What do we know about the source Ips?

- One time only source IPs probably fall into several distinct categories, but we haven't studied them that much yet
- We know where they live
- Well, at least where they want us to think they came from

A correlation example

- Question: is anyone targeting DOE sites?
- If they are, should we worry?
- Many Internet related things follow what is called a Power Law
 - Last year's conference gave me the idea

A correlation example

- How many source IPs “visited” more than one DOE site?
- A lot!

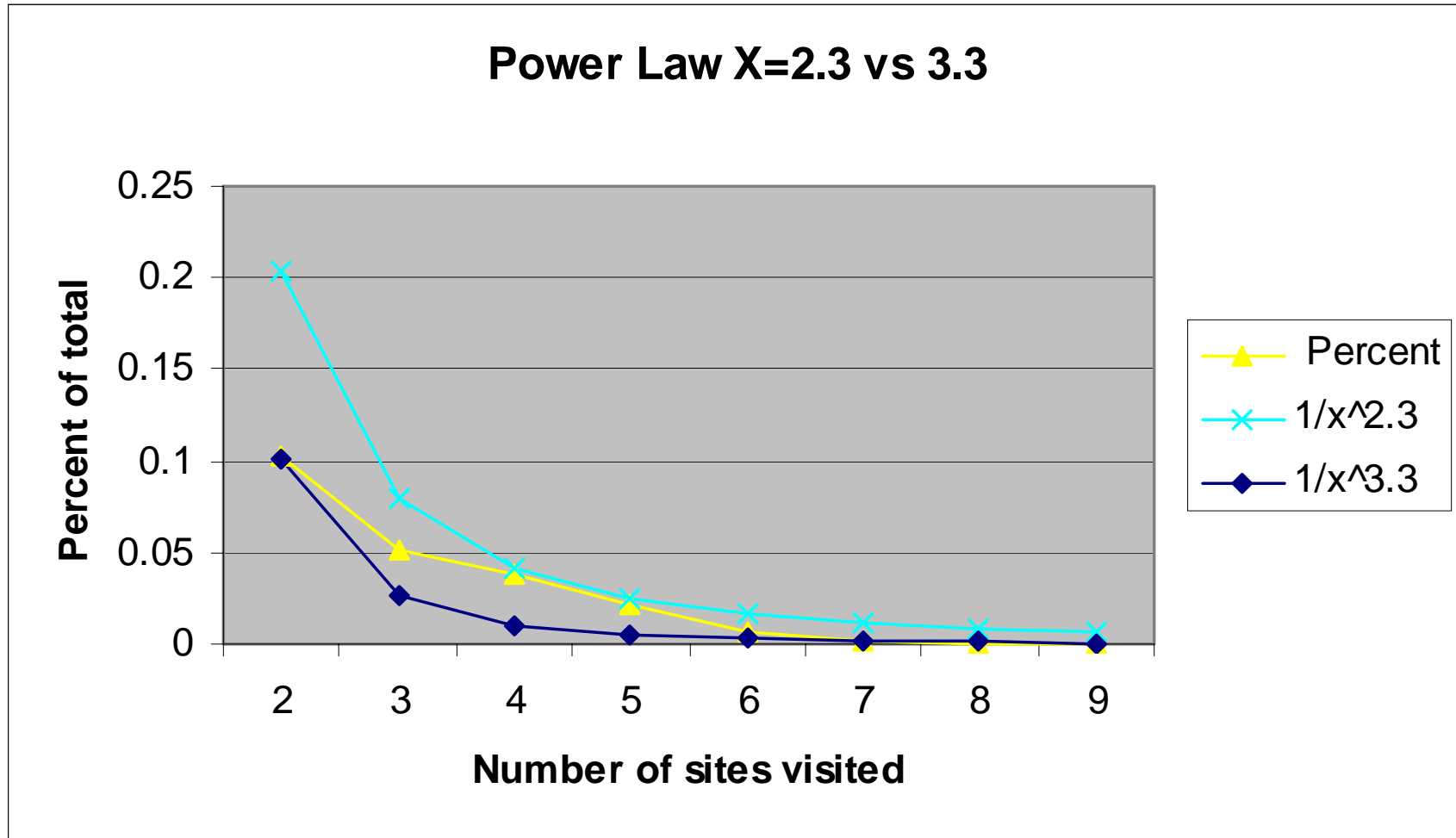
A correlation example

- Count how many sites each source IP “visited”
- Sort and look at the list
- Process that list to determine how many source IPs visited 1, 2, ... all sites
- At the time this example was run there were 9 sites

Correlation example

Distribution of source IP site counts			
Site Count	Source IP Count	Percent	
1	63954	0.777443	
2	8477	0.103049	
3	4195	0.050996	
4	3155	0.038353	
5	1751	0.021286	
6	508	0.006175	
7	169	0.002054	
8	47	0.000571	
9	6	0.000061	

Power Law Example



Power Law Example

- According to my interpretation of Power Law, we probably see fewer multi-site visits than expected

How does the activity change over time?

- Incidents.org uses a very specific algorithm
 - It looks at 30 days compared to 2 days
 - Scales and determines trends

Can we spot anomalies in time to do something about them?

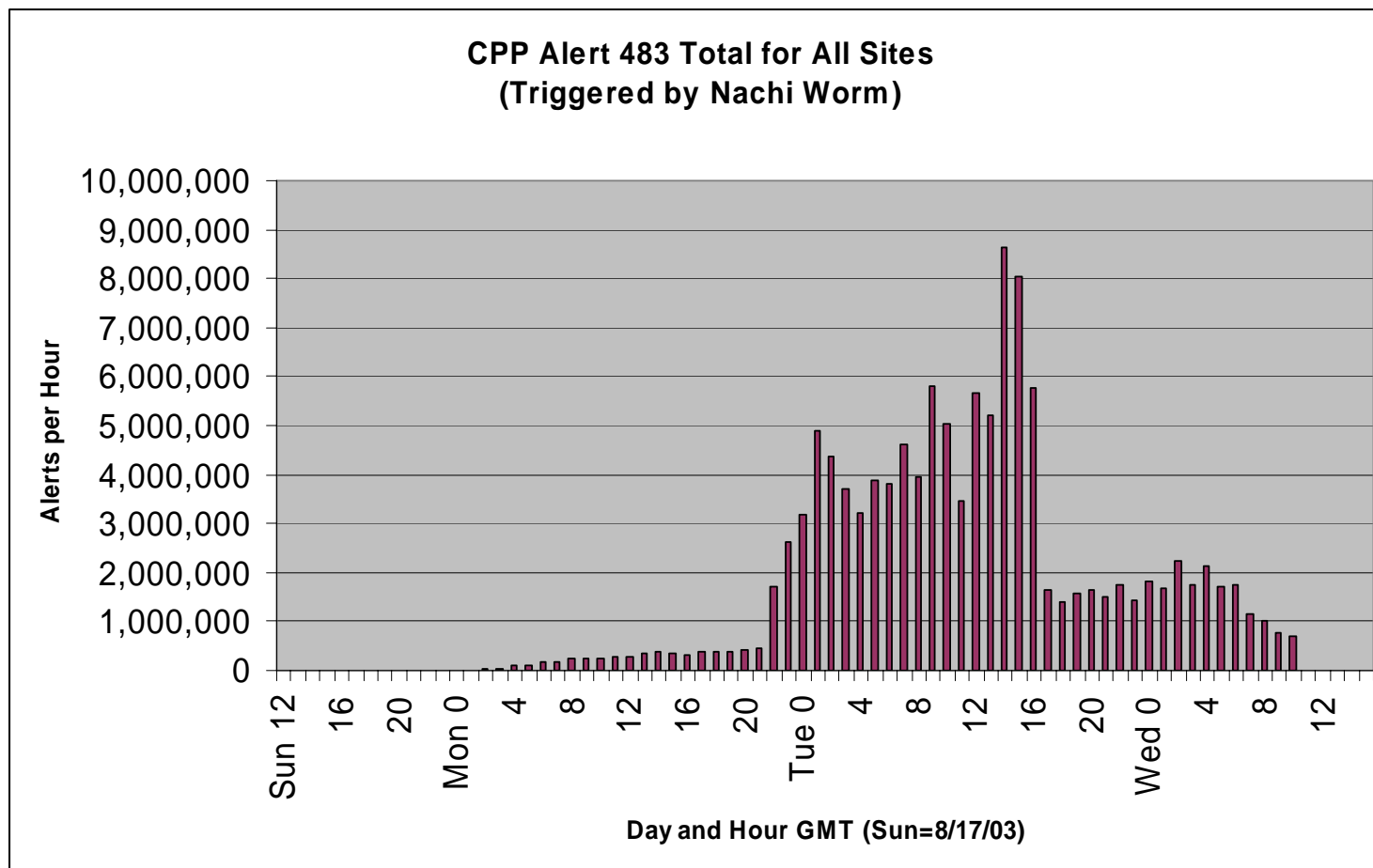
- The incidents.org model needs two days to spot a trend
 - Some activity becomes critical in hours
 - Can it be adapted to shorter time scale?

What are the problems with analysis of data over time?

- Time synchronization
 - Solved
- Data transfer cycle time
 - Programmable
- Missing data
 - The big problem
- Processing power vs analysis interval
 - Don't get greedy unless you have cash

Example 1: Missing Data

- What does this graph tell you?



Example 1: Missing Data

- At 2200 hours Monday a sensor came back on line
- At 1600 hours Tuesday that sensor went off line
- Even without the spike from the errant sensor, it is a steep ramp up (until they reconfigured Snort without #483)

Example 2: Weird stuff

- What does it mean when the
 - number of ports \approx number of destination ips?
- It caught my eye ...
- This was Snort data, so some signature was triggered, but which one?
 - #524 BAD-TRAFFIC tcp port 0 traffic

OK, so I'm interested

- Look at all Alert 524 records
- 11 Source IPs, several sites, 1,000s of hits
- All different destination ports
- Unless the destination IP is the same
 - Same or different source IP
 - Same destination IP
 - Same destination port

Could it get stranger?

- Of course ... analyst next door gets call
 - See any traffic with window size=55808?
 - Yes ...
- At the time it was called the “stealthy trojan” (not too stealthy)

Things we'd like to add

- Use what you know about source IPs
 - Maintain a *Watch List* and flag occurrences in new data
 - Maintain a *Benign List* and remove IPs on it from further consideration
- Prioritize the Snort Alerts and look at the bad ones

More stuff we'd like to add

- How do we find new bad stuff?
 - Compute every count and aggregation you can think of (within reason) and look at them for patterns
 - Graph anything that might provide a visual pattern not easily computed
 - Look for trends

What can we borrow from other fields?

- Digital Signal Processing
- Data mining
- SETI
- ATC
- Cryptanalysis

- And always ... Think Big!!

References

- www.arcsight.com
- www.guarded.net/
- www.i2.co.uk
- <http://www.olapreport.com/>
- <http://www.olapcouncil.org/>
- <http://www.andrew.cmu.edu/~rdanyliw/snort/snortacid.html>

More References

- <http://isc.incidents.org/>