Argonne NATIONAL LABORATORY

DLP Detection with Netflow

Christopher Poetzel
Network Security Engineer
Argonne National Laboratory

FloCon 2011
Jan 11, 2012

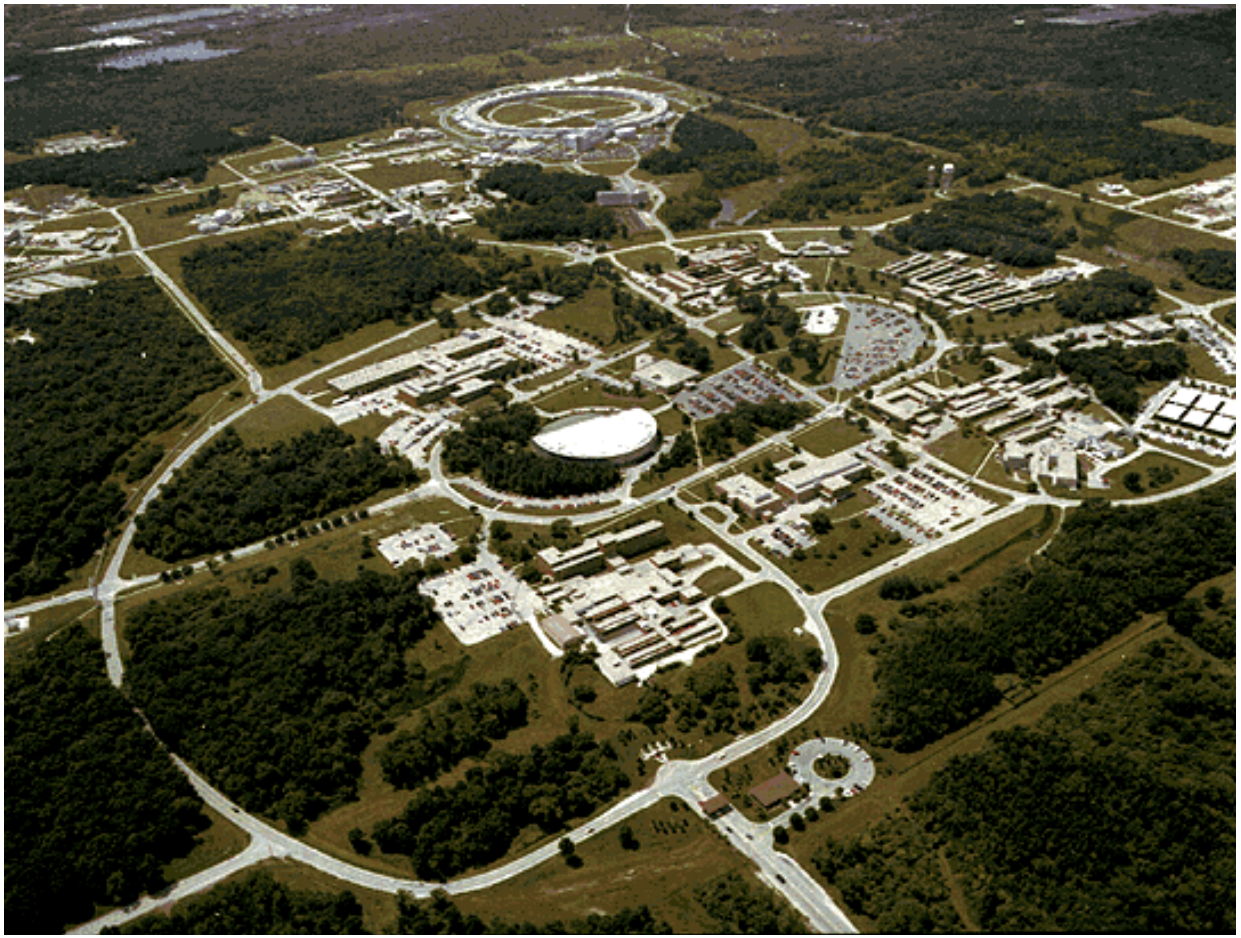
U.S. DEPARTMENT OF ENERGY

Who Am I?

- Christopher Joseph Poetzel
- University of Wisconsin-Madison
  - BS Computer Science
- Argonne National Laboratory
  - summer student through college
  - 10 years full time
- Network/Security Engineer
  - Firewall/VPN/Network Administrator
  - IDS/Netflow Scripting
  - Proxy/URL Filtering

Brextyn Ayers Poetzel
Nov 5th, 2010

Argonne National Laboratory



IT Environment Challenges

- Diverse population:
  - 2500 employees
  - 10,000+ visitors annually
  - Off-site computer users
  - Foreign national employees, users, and collaborators
- Diverse funding:
  - Not every computer is a DOE computer.
  - IT is funded in many ways.
- Every program is working in an increasingly distributed computing model.
- Our goal: a consistent and comprehensively secure environment that supports the diversity of IT and requirements.
- Balance Science, Security, and Architecture.

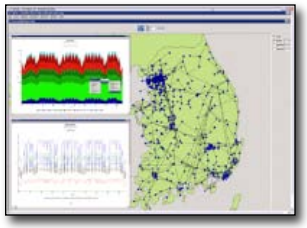
Argonne is managed by the UChicago Argonne LLC for the Department of Energy.

Emphasis on the Synergies of Multi-Program Science, Engineering & Applications

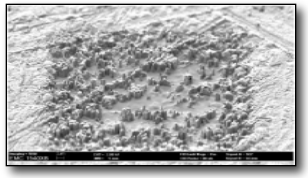
Computational Science

Accelerator Research

Fundamental Physics



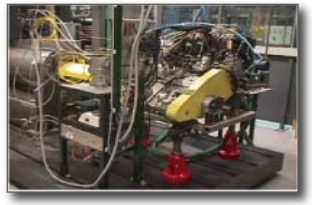
Infrastructure Analysis



Materials Characterization

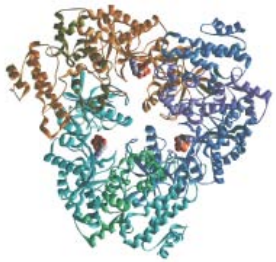
Catalysis Science

Argonne NATIONAL LABORATORY



Transportation Science

User Facilities



Structural Biology

Nuclear Fuel Cycle

.. and much more.

High Level Split of Argonne Divisions

Scientific

- Advanced Photon Source
- Biology
- High Entergy Physics
- Environmental Sciences
- Super Computers

- Mission is to do Science
- More open and collaborative with world
  - Less controlled by Central IT
- **Full outbound restrictions**

Operations

- HR, Finances
- Plant and Facility Management
- Medical
- IT Computer Support, Core Networking
- Cyber Security

- Mission is Support Science
- Less open and little collaboration
  - More Controlled by Central IT
- Access to Sensitive Information
  - PII Records, Payroll, Medical
  - Benefits, Travel System
- **Limited Http, HTTPS (some ftp)**

Data Loss Prevention (DLP)

- **Data Loss Prevention** (**DLP**) is a computer security term referring to systems that identify, monitor, and protect data in use, data in motion, and data at rest through deep content inspection, contextual security analysis of transactions, and with a centralized management framework.
    - Protect Data in use: endpoint actions
    - Protect Data in motion: network actions
    - Protect Data at rest: data storage

- The systems are designed to detect and prevent the unauthorized use and transmission of confidential information.
- The Data to protect is dependant on organization
  - PII (Social Security Numbers, Birth Dates, Addresses)
  - Credit Card Numbers
  - Source Code
  - Internal Only Documents

- Many Many Vendors in this Game
  - McAfee, BlueCoat, RSA, Symantec, Trend ……. BECAUSE

DLP Happens .. All the time .. Even to Me

- WikiLeaks: Nov 2010
  - Government Documents leaked for all to see
  - Arrests Made, USA Government “Embarrassed”, National Security “Threatened”

- Gawker Media Hacked: Dec 12, 2010
  - 1.3 million user names and passwords exposed after user database compromised
  - 500MB Torrent file of all accounts/passwords
  - Gawker Advises users to change passwords or delete account

- Heartland Payment Systems (Credit Card Processing): May 15th, 2008
  - 130,000,000 Credit Card Numbers Stolen
  - Settlement with VISA: $60,000,000.00 Jan 2010
  - Settlement with AMEX: $3,538,380.00 Dec 17, 2009

- University of Wisconsin-Madison: Nov 26, 2010
  - 60,000 names and identification card numbers including Social Security numbers stolen from server (1 was me)
- http://datalossdb.org

DLP happens, so now what

- Early 2009, Argonne Cyber Security Program Office says DLP as a capability we would like to have.

- How can this be done given the following:
  - No money for vendor solution
  - No complete desktop network control of all hosts
  - Small amount of time to commit to project
  - Automated System
    - minimal human interaction
    - We do not have 24X7 analysts or operations center
    - We do not want be chasing down alerts all the time

  - We are not web traffic cops. We are not trying to stop people from getting to Facebook/Yahoo/etc
    - Want to be alerted on large unauthorized offsite uploads that might be DLP
    - Want to catch those “abuse” cases of people web surfing all day/night long

- What is the our best bang for out buck?

Our Solution

- A Netflow based solution to look for anomalous amounts of offsite data within the last hour.

- Focus on areas of greatest risk
- Alert us to things “out of normal”
- Configurable
  - Ability to exclude ips
  - Ability for different thresholds for different networks
- Automated Email Alerting

Focus on areas of greatest risk

- Operations Divisions provide the greatest area of risk
  - Contains the meat of sensitive data
- Jobs are not about collaboration, about support
- Offsite traffic is limited to Http, Https and thus easier to model and understand

Operations

•HR, Finances
• Plant and Facility Management
• Medical
• IT Computer Support, Core Networking
• Cyber Security

- Mission is Support Science
- Less open and little collaboration
  - More Controlled by Central IT
- Access to Sensitive Information
  - PII Records, Payroll, Medical
  - Benefits, Travel System
- **Limited Http, HTTPS (some ftp)**

Alert us to things “out of normal”

- Using netflow we base lined the normal hourly amount of offsite web traffic for 1 month.
  - Fairly simple netflow script
- On Average, Per subnet, offsite Web traffic threshold
- Weekdays
  - 6am-6pm, 25 MB
  - 6pm – 6am, 5 MB
- Weekends, 5MB

Configurable

- Exclude known offsite uploaders by IP Address
  - Stored in a mysql database table
- MB Thresholds are on a per subnet basis
  - Also in a mysql database table

Automated Email Alerting

- ALERT for Excessive OFFSITE WEB Traffic

- FWInterface: sample_yellow network
- FWNetwork: 146.137.XXX.0
- FWIntDescr: Sample Yellow network

- Dest: Offsite NON-ANL on TCP 80,443

- TimeStart: Monday, 2010-12-13 11AM
- TimeEnd: Monday, 2010-12-13 12PM

- Offsite MB
- For Subnet: 38.096

- Threshold for 1 Host During Period: 25 MB/hour for single host
- Further Information for Alarm Period
- # --- ---- ---- Report Information --- --- --- #
- # Fields: Total
- # Symbols: Disabled
- # Sorting: Descending Field 2
- # Name: Source IP
- # Args: flow-stat -f9 -S2

- # IPaddr flows octets packets
- #
- 146.137.58.24 704 27035481 28856
- User:Doe, Jane DNS:csi3388XX

- Top 25 Dest Hosts
- # recn: ip-destination-address*,flows,octets,packets,duration
- post.craigslist.org,89,25080978,21459,197888 **← Key Line in Alert Email**
- a184-84-255-8.deploy.akamaitechnologies.com,44,416136,745,2060800
- 159.53.64.105,85,383093,1324,137472
- ** others removed **
- # stop, hit record limit.

- 146.137.58.25 1596 5510900 49389
- 146.137.58.30 82 1380209 25425
- 146.137.58.42 492 1196430 5126

- Apparently this user was uploading something large to craigslist during work hours.
  - Work related??

Script Logic / Flow-Tools Guts

- Create ACL to watch for traffic from network Y (include exemptions)

- Determine Offsite Traffic in last hour for network Y (146.137.X.Y)
  - Run Netflow on Border Router to get Offsite Mb amount for subnet for past hour
  - flow-cat $flowargs | flow-filter -f /tmp/$Tempfile -S check1 -P 80,443 | flow-stat -f9 -S2

- Check amount against thresholds
  - Thresholds run against database limits
- Send Alert Email if threshold tripped

- 356 line perl script, backend database table for thresholds, exclusions, and subnets to watch
- Fairly Efficient / Quick
  - Watching 49 networks for DLP detection
  - Average runtime is 5minutes
  - Took less than a week to come together

What the solutions does

- First insight into DLP for those networks where it matters
  - HR, Financial People, Lab Directors, etc
- Identifies people uploading large amounts of data to offsite services
  - Facebook
  - Online Email attachments
  - Snapfish/Walgreens/ETC
  - YouTube Videos
  - Or something large heading offsite that shouldn't be
- Identifies afterhours personal doing lots of web surfing in the wee hours of the morning
- Exemptions and different thresholds do not bury us with false positives
- Helps us know our network better

What this solution is not

- Does not actually stop DLP, just helps detect it
  - Focused only on the network detection side of DLP
- Gives no information on data offloaded
  - Not available within netflow
  - Can obtain with use of local PCAP device
- No Polices like a vendor solution
  - No inspection of traffic leaving (social security numbers, credit card, resumes, etc)
- Will not catch DLP when
  - Network MB volume is low
  - Local Argonne network is not being monitored

Future

- Solution has done its job for past 2 years as an early detection system
  - It is far from perfect but has helped to
    - Find some legitimate offsite uploads that needed to be more “controlled”
    - Find those egregious web surfers
- If we were to progress this script/solution to the next level
  - Watch offsite levels by IP address, not by network
  - Include some automatic data gathering from our PCAP software to give insight into data pushed offsite
  - Automatic trending of thresholds
- We are investigating commercial DLP Solutions
  - Any recommendations please let me know

Takeaways

- DLP is a problem and it does happen
- Our quick and simple DLP solution is a great example of how netflow statistics can be used to in various productive ways
- At Argonne, our staffing situation limits us from any real-time operator style netflow interface
  - Only real-time netflow interactions is once an alarm/alert has been triggered
  - If a commercial or home-brew tool can not send out automated alarms in some manner, we will not use it
- We have been using netflow for cyber security and network related endeavors for 9+ years.
  - It is an invaluable tool for out cyber security and network personal.

All done

- Thanks for the ear
- Questions
- Cpoetzel at anl.gov