Network Profiling with SiLK

George Jones, Austin Whisnant
CERT Network Situational Awareness Group

Notices

Profile Outline -1

Motivation

Methodology

- Gathering information
- Selecting a data set
- Identifying active assets
- Cataloging common services
- Cataloging other services
- Cataloging leftover assets

Related Topics

Wrap-up

Network Situational Awareness

First, a word from our grand motivating principal:

"The systematic gathering, analysis, and interpretation of data from local and remote networks [**network**] regarding structure, applications, traffic, and resources [**situational**] to produce actionable information for decision making in network operations and defense [**awareness**]"

– Tim Shimeall

Or, simply…

"Tell [the general] what he needs to know"

Network Profile

A network profile is an inventory of all of the assets on a network, their characteristics and their associated purpose.

A network profile can also include:

- Traffic volumes
- Network maps

Other Stuff to Profile

IP addresses

External Networks

- Countries, ASes could be considered external networks

Activities

- Tunneling
- Beaconing
- Scanning
- Specific types of malware – given known ports and protocols as well as possibly behavior
- DOS and DDOS
- P2P

Motivation

Network situational awareness

Network administration

Security administration

Guidance for purchasing and staffing

Trending and monitoring

Profile Outline -2

Motivation

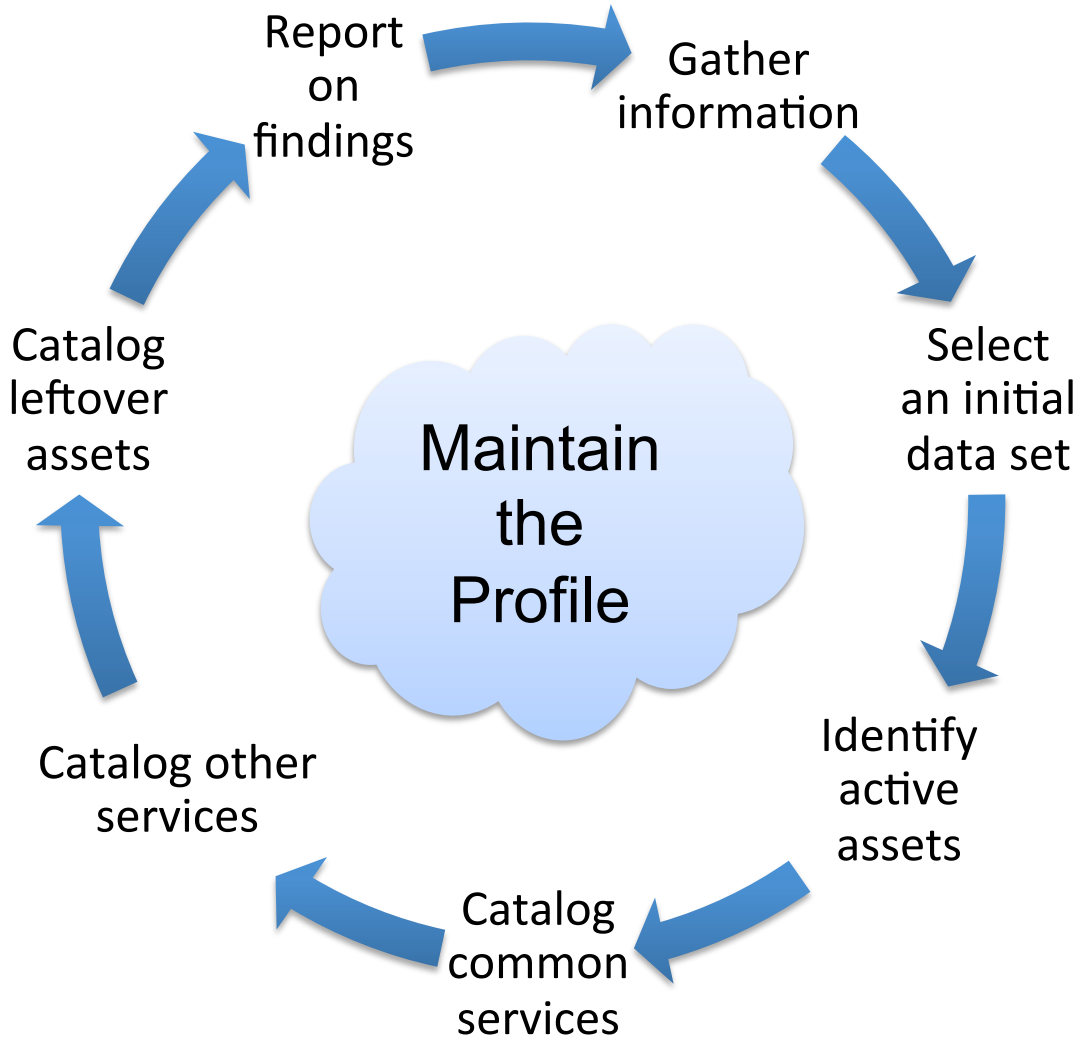
Methodology

- Gathering information
- Selecting a data set
- Identifying active assets
- Cataloging common services
- Cataloging other services
- Cataloging leftover assets

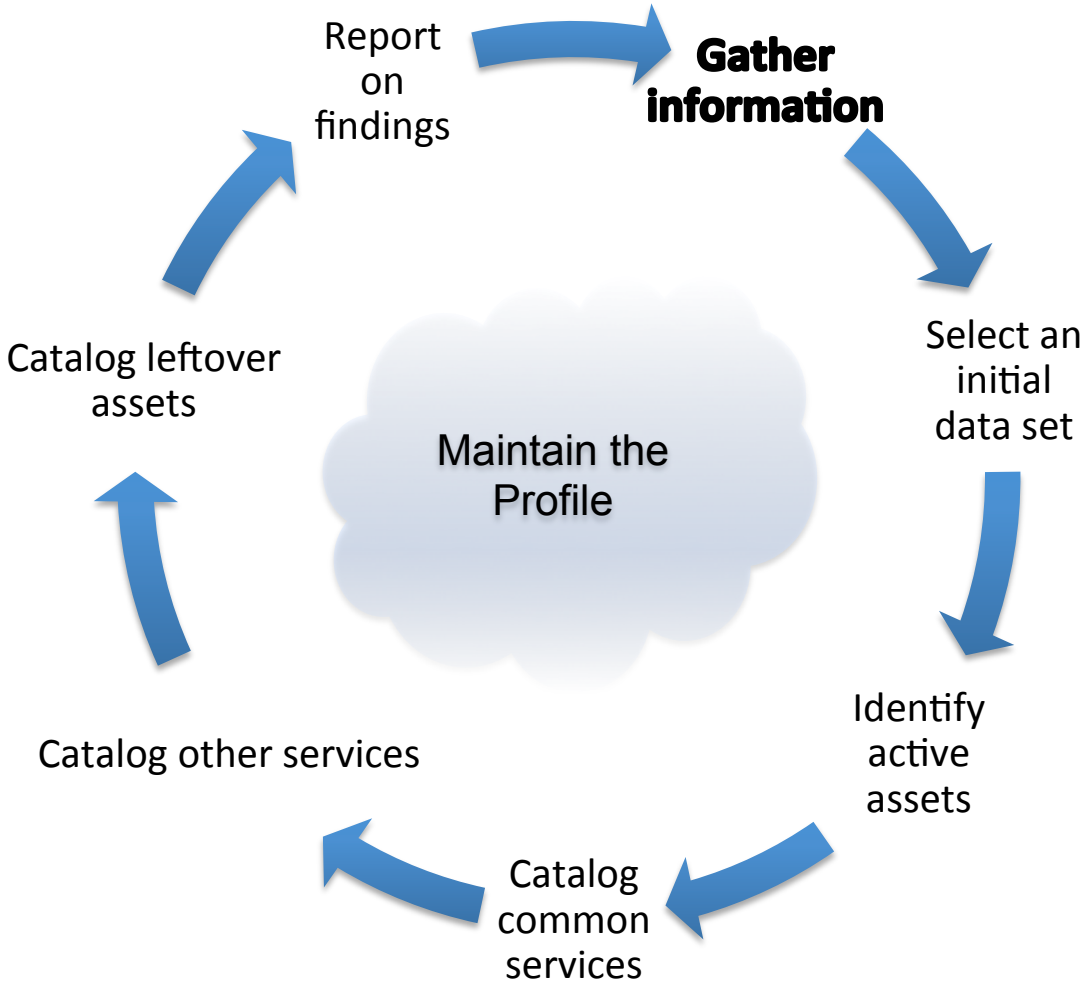
Related Topics

Wrap-up

Network Profiling Process Cycle



Process Cycle: Gather



Choosing a Network to Profile

Live vs. Captured Data?

Production vs. non-Production?

“small” vs. “large” volume?

What’s available: flow, full packet, logs, other?

Are DNS and WHOIS lookups meaningful?

Relatively recent data?

Is the network documented?

Are servers, services known?

Is there Internet traffic?

Is the ratio of attack traffic to normal traffic skewed?

Is the number of servers, clients, protocols, etc. “normal”?

Was the network built for exercise or production?

Gather Available Information

Network Diagrams

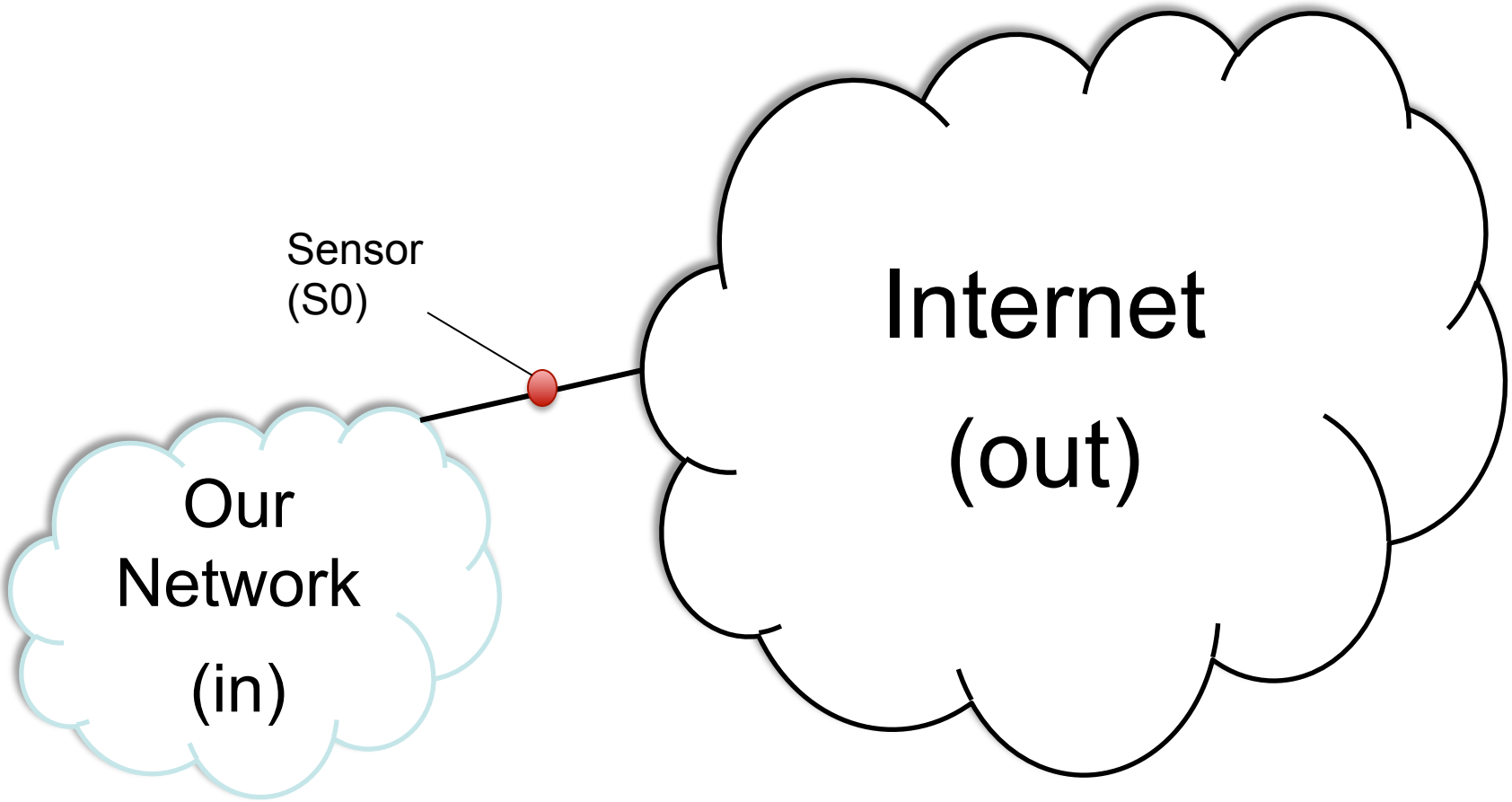
For live networks consider `dig`, www.robtex.com

For operational networks, talk to netops, IT, management, etc.

Consider active scanning (nmap, nessus) if allowed.

Policies

“In” and “Out” in SiLK



A Word About Sensor Placement

You only see things where you have instrumentation.

You will have a hard time seeing “insider threat” activity with netflow at the border.

- You might want authentication, mail and web server logs, registry dumps, disk images, full packet capture, etc.

Encryption, tunnels, VPNs, NATs, proxies, anonymization services (tor), etc. can cause problems.

The “inside/outside” model is becoming dated in the face of mobile devices and ubiquitous network access.

Sensor Exercise

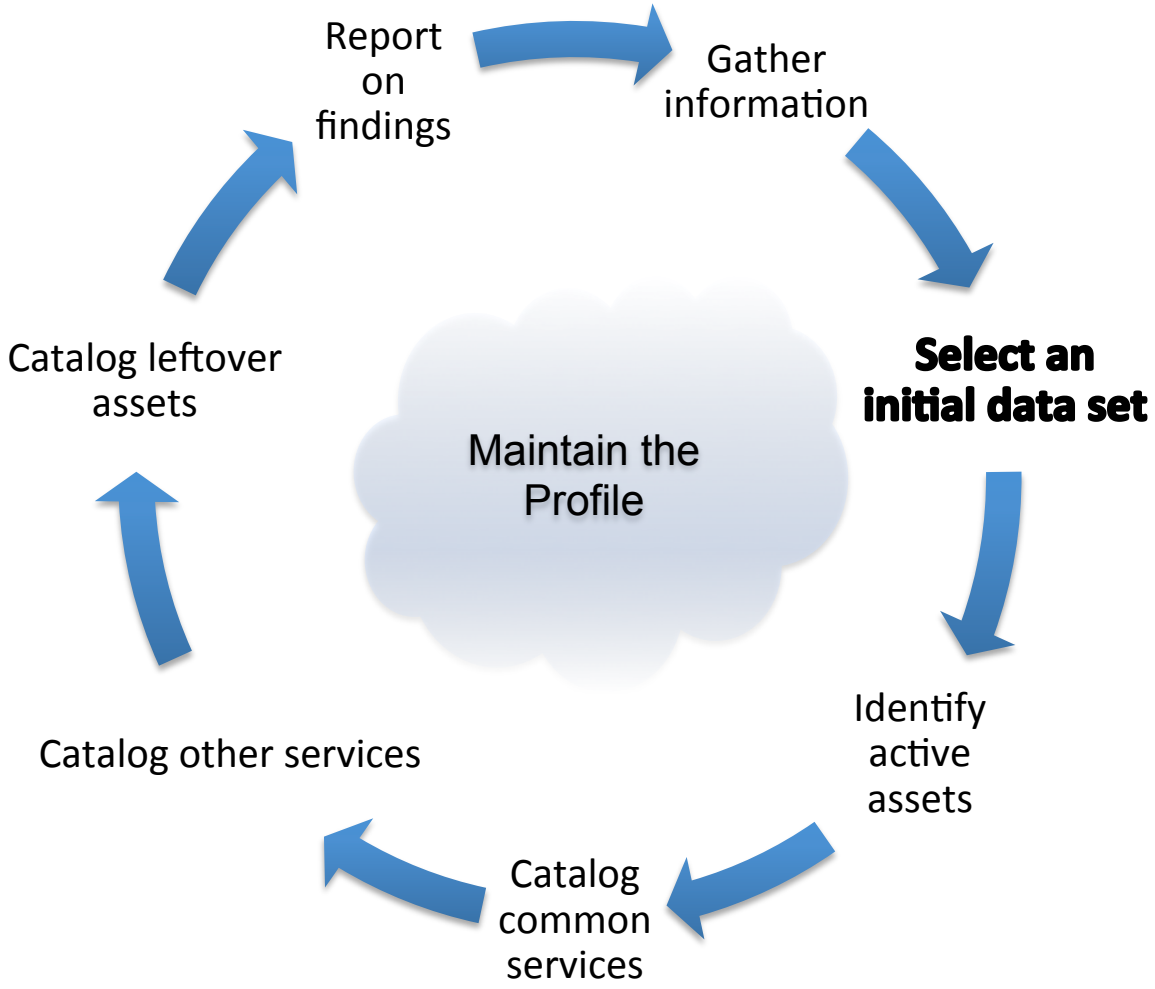
How many sensors are in the data set?

What are the internal network blocks?

Example Profiling Spreadsheet

Internal IP	Protocol	Internal Port	Internal Name	External IP	External Port	External Name	Comments

Process Cycle: Initial Data Set



Selecting an Initial Data Set

Should be:

- Representative
- Small enough to have reasonable query time

Avoid sampled data.

Start with the busiest time of day.

Choose outbound traffic if dividing by direction.

Sample Set Validation

Should match what you “expect” (i.e., what is “typical” for your network)

- Typical ports
- Typical protocols
- Typical traffic patterns

Sample Data Exercise

Create a sample data set named "sample.raw" (outbound, one day).

- How many records does it have?

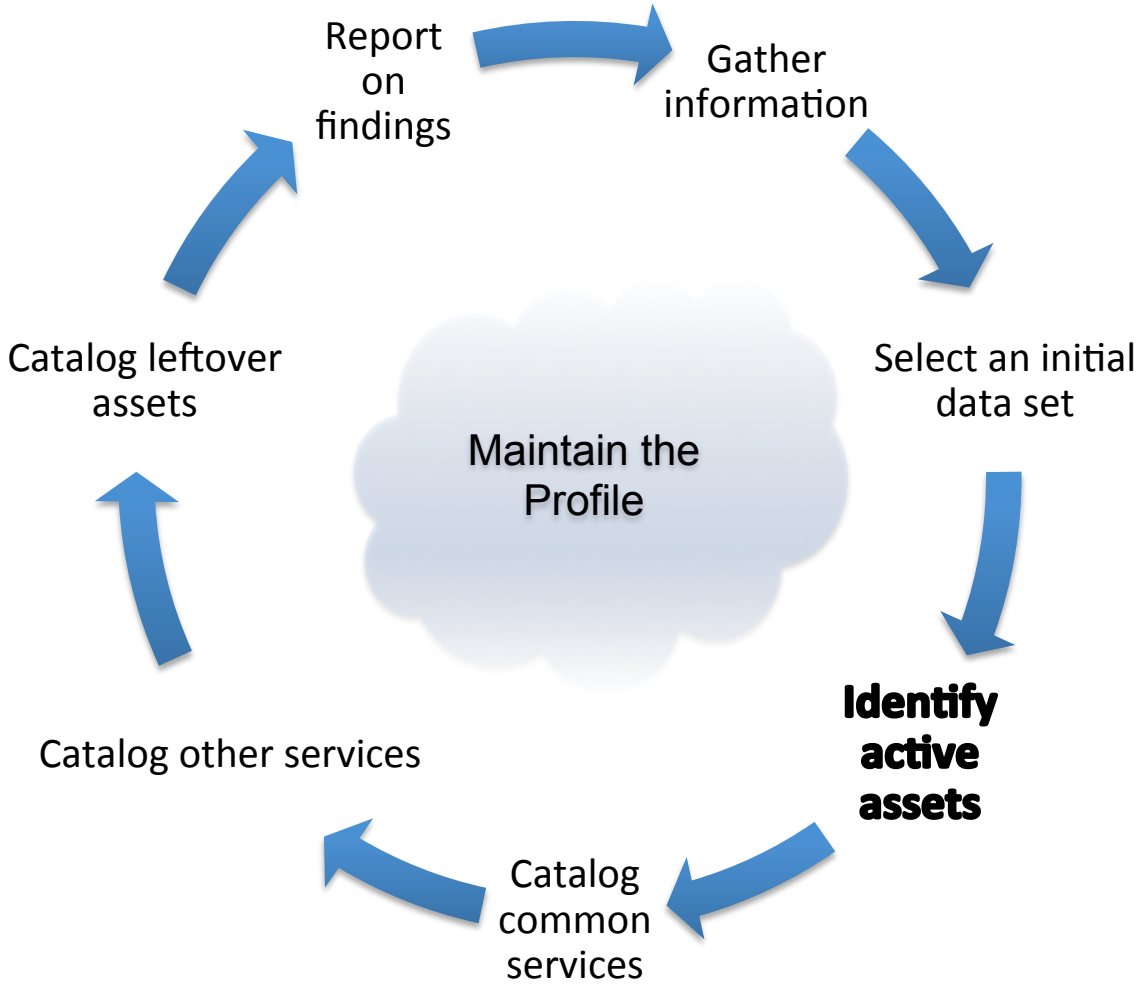
Find the top protocols.

Find the top services provided.

Find the services requested most by internal addresses.

Look at a breakdown of the traffic volume over the sample duration.

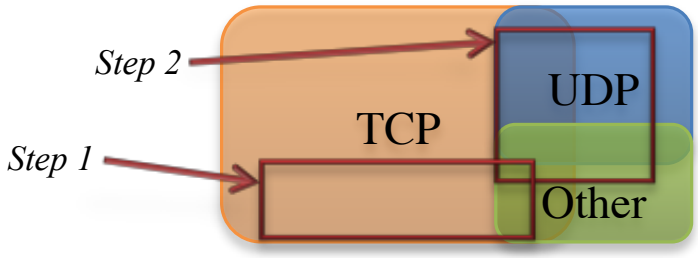
Process Cycle: Active Assets



Active Hosts

Identify active hosts that have TCP connections.

Identify active hosts that have other connections.



Merge the two sets of addresses together.

Keep in mind, failover circuits may not show up in this traffic.

Data Anomalies

Transit traffic

Asymmetric routing

Assets Exercise

Find TCP “talkers” (tcp_talkers.set)

- How many addresses are there?

Find other talkers (other_talkers.set)

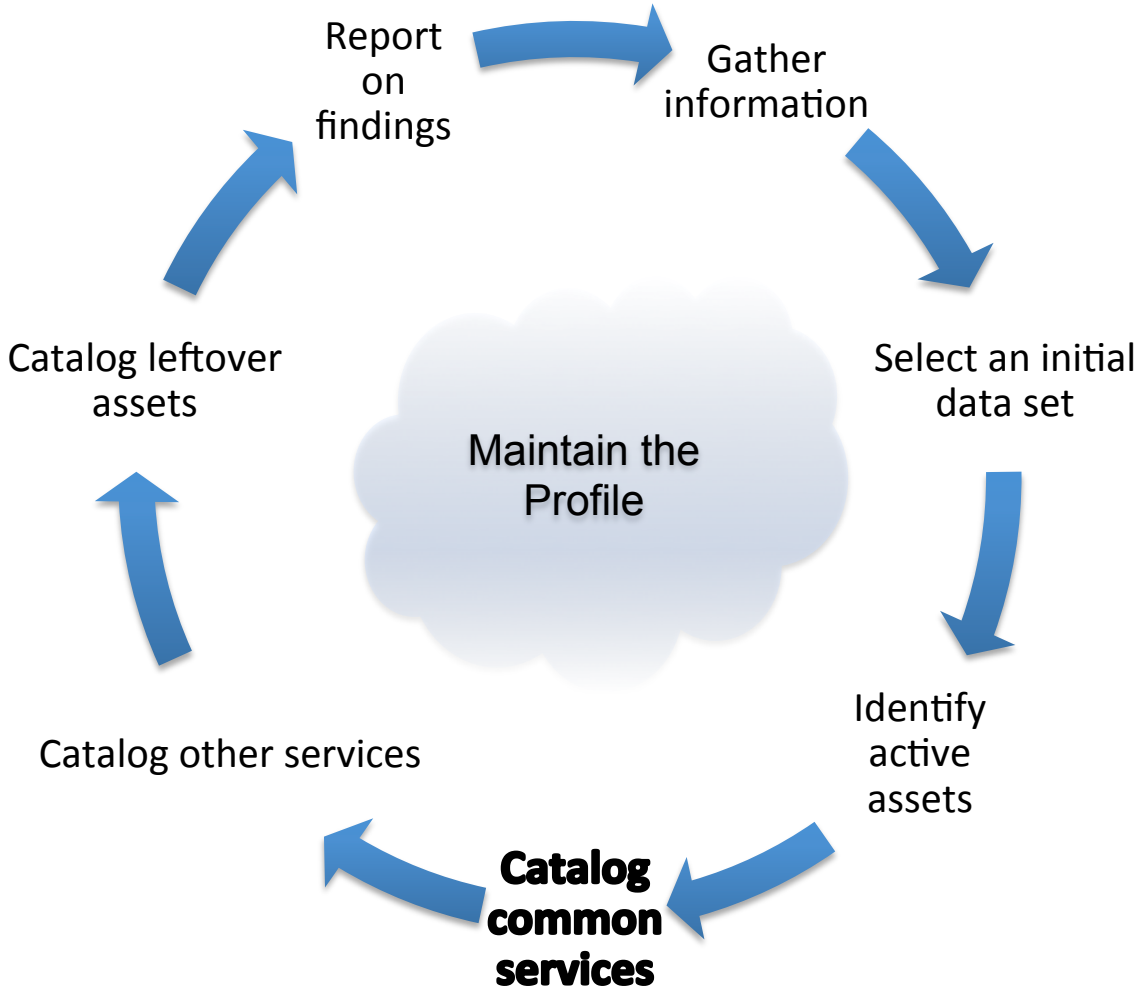
- How many addresses are there?

Combine the two (talkers.set)

- How many addresses are there?

What does the network structure look like?

Process Cycle: Catalog Services



Profiling a Service

Identify common ports and protocols.

Note architecture and any special behavior (flags, multiple protocols, etc.).

Pull client and server traffic separately.

Find hosts that account for at least 1% of service traffic.

Save addresses and ports.

Perform supplemental analysis.

Validation

Examining data you already have (flow, pcap, logs, etc.)

- Packet sizes, timing, protocol info, actual data, external hosts

Domain name resolution

Accessing the service

Telnet (raw connect) to the service

Active scanning (nmap, nessus)

3rd party services: robtex, google, whois

Service Anomalies

Traffic does not always exactly follow the expected pattern.

- Historical behavior
- Multiple services on a single box
- Unconventional devices
- General protocol anomalies

Profiling Web Traffic

How does web traffic work?

How to find servers with SiLK?

How to find clients with SiLK?

Validation

Anomalies

Web Service Exercise

Identify web servers and the ports on which they operate.

- Remember, 1% of the web server traffic

Identify web clients and the ports on which they operate.

Profiling Email Traffic

How does email traffic work?

How to find servers with SiLK?

How to find clients with SiLK?

Validation

Anomalies

Email Service Exercise

Identify email servers and the ports on which they operate.

Identify email clients and the ports on which they operate.

Remember, 1% of the traffic

Profiling DNS Traffic

How does DNS traffic work?

How to find servers with SiLK?

How to find clients with SiLK?

Iterative vs. Recursive servers

Validation

Anomalies

DNS Service Exercise

Identify DNS servers.

- Which are recursive and which are iterative?

Identify DNS clients.

Profiling VPN Traffic

How does VPN traffic work?

How to find concentrators with SiLK?

How to find site-to-site with SiLK?

Validation

Anomalies

VPN Service Exercise

Identify VPN “servers.”

- Which are concentrators and which are site-to-site?

Profiling FTP Traffic

How does FTP traffic work?

How to find servers with SiLK?

How to find clients with SiLK?

Validation

Anomalies

FTP Service Exercise

Identify FTP servers.

Identify FTP clients.

Which clients/servers are active and which are passive?

Using Advanced SiLK

Tuple files

- Port to protocol mappings
- IP to port mappings

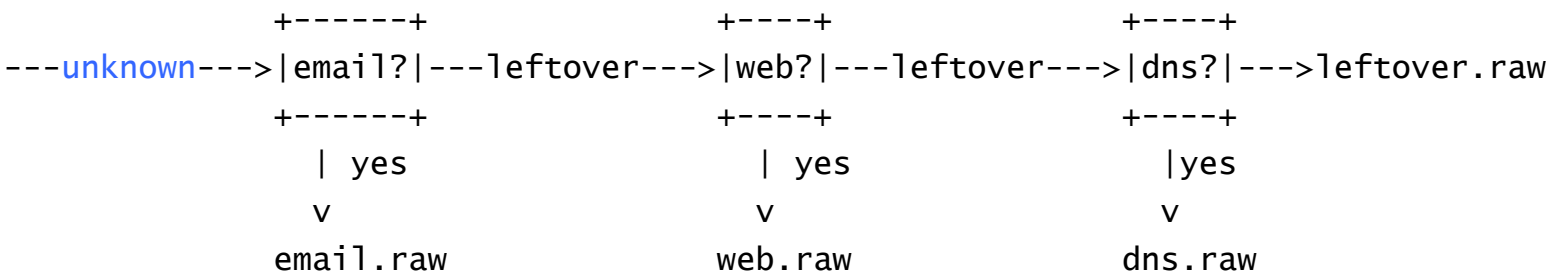
Pmaps

- Labeling ports as specific services
- Labeling IP addresses

Pipes

- Combining multiple queries into a longer, more efficient query

Everything Is Leftovers



Start with unknown flow.

Test for “service” (Web, DNS, etc.).

- Save flows that pass test to file.
- Pass flows that fail (leftovers) to next test.

At the end, you have the unanalyzed data.

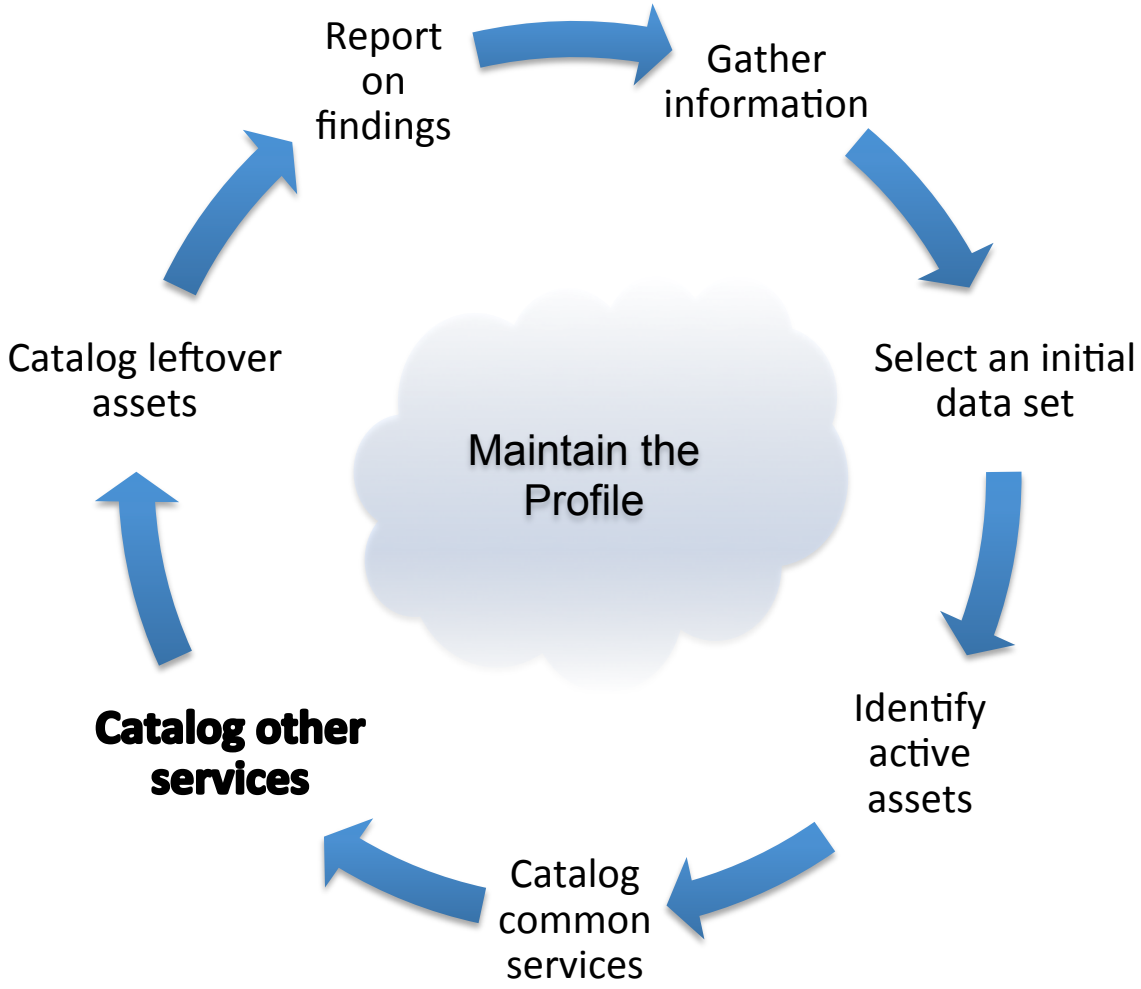
- Hopefully small. Start profiling Ips.

Leftovers Exercise

Advanced exercises:

- Try incorporating the “cataloging common services” queries into one single (piped) command
- Try incorporating tuple files into cataloging common services
  - Hint: `man rwfilter` (--tuple-file & --tuple-fields)
- Try incorporating pmaps into cataloging common servies
  - Hint: `man rwpmapbuild`

Process Cycle: Other Services

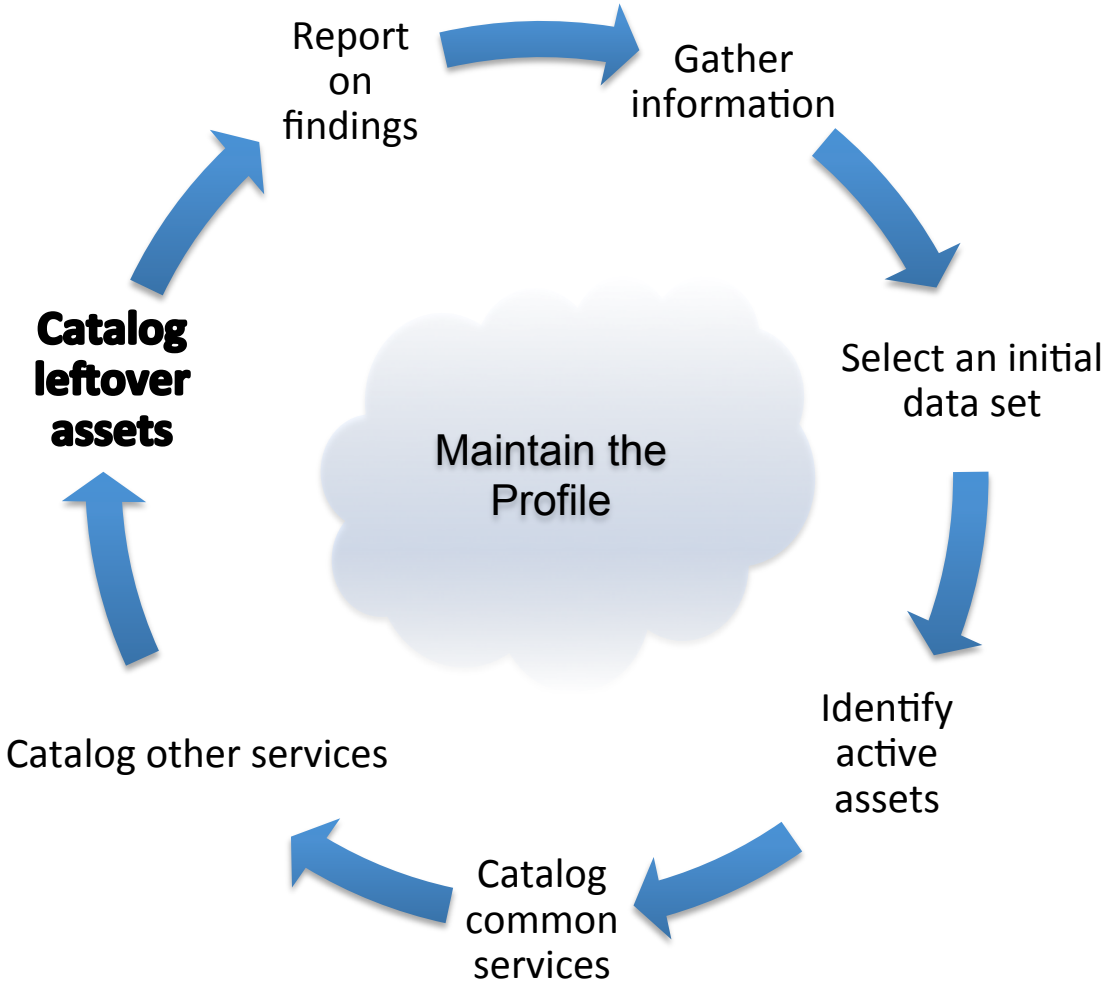


Other Services Exercise

Profile the other top services being used (found in previous exercise).

- List internal IP address, ports, and protocol for each service.

Process Cycle: Leftover Assets



Leftover Assets

Expand the time frame for the sample set and see if there are any hosts that have not yet been profiled.

The leftovers can be profiled individually.

Determine the most used port and protocol for each service.

- A more detailed profile of the address may be done later.

Common Findings

Services already profiled

Common services over encrypted or legacy ports

Other well-known services

Routers

Leftover Assets Exercise

Determine which active addresses have not yet been profiled.

List the top service each is running/requesting (ports and protocols).

Process Cycle: Report

Profile Outline -3

Motivation

Methodology

Related Topics

- IP address profiling
- Activity profiling
- External network profiling
- Trending

Wrap-up

IP Address Profiling

Can tell a lot about a host based on its network activity

Useful for network profiling if the network is small or there are active addresses that do not have common traffic, or if more detail is needed about a specific address

What types of traffic and at what volumes are produced

Which IP addresses does it communicate with

Its OS and running applications based on port usage and timing

IP Address Exercise

Choose a host and enumerate its services.

Can you tell what OS it is running?

What software?

What are its traffic volumes?

Who does it communicate with?

Activity Profiling

Often based on characteristics other than ports and protocols

- Timing, flags, traffic volume, packet/flow sizes

Examples

- Beaconing will likely have short flows with a small number of packets going out of the network at regular intervals.
- A DDOS attack might have a high number of flows with large packets and a specific flag set.

Scanner Exercise

Catalog scanners

- Decide how scans look in network flow (both vertical and horizontal)
- Create a formal definition of both types of scanner traffic in `rwfilter` notation
- Create a spreadsheet with a column of scanner IP addresses and the time of each scan
- For each address and time pair, document the type of scan and scan target IP address range
- Profile scanning hosts by cataloging their ports and services

Application Exercise

Application mismatches

- Create a list of all internal IP addresses that have conflicting port information and application fields.

Who's doing it?

Who are they contacting?

When?

Tunneling Exercise

Tunneling

- Long duration flows over common or uncommon ports

Who's doing it?

Who are they contacting?

When?

Malicious Activity Exercise

Other possible malicious activity

- DOS, DDOS, P2P, Beaconing

Who's doing it?

Who are they contacting?

When?

Domain Access Exercise

Connections to known “bad” domains

- Who’s doing it?
- Who are they contacting?
- When?

Pick a random set of 20 domains from the external hosts to use as bad domains.

External Network Profiling

Similar to internal network profiling except we do not have the same kinds or the same amount of flow data

What types of traffic and at what volumes are there?

What external addresses are doing the talking?

- Often NATed

What internal addresses are doing the talking?

Is there any malicious traffic?

- If so, what types and how much?

Network Activity Exercise

Identify active address blocks.

Identify active IP addresses.

Determine services of each address.

Map the external network.

Is there any difference between the results from different sensors?

Trending

Helps maintain the profile

- Only the assets that change need to be re-verified.

Gives admins an overall picture of the network

Daily/hourly snapshots of traffic volumes and top services

What changes?

Trending Exercise

Make bar graphs of traffic for each hour.

- Choose whatever category you want (protocol, port, application, scanning volume, etc.).
  - Are there changes from day to day?

For more information

"Network Profiling Using Flow" Tech Report
http://www.sei.cmu.edu/library/abstracts/reports/12tr006.cfm

SiLK – documentation and software
http://tools.netsa.cert.org/silk/

SiLK live CD

http://tools.netsa.cert.org/livecd.html

SiLK ToolTips
https://tools.netsa.cert.org/confluence/display/tt/Tooltips

FloCon Proceedings
http://www.cert.org/flocon/proceedings.html

Questions?