Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

CERT

Network Flow Analysis Using SiLK (v1.1.3)

Training provided by the
CERT Network Situational Awareness Group

January 2009

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

Course Modules

- Introduction
- Basic SiLK Tools
- Advanced SiLK Tools
- Unix Scripting
- Visualization
- Conclusion

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

CERT

Module Outline

- About the Training
- Introduction to Flow
- Introduction to Unix
- Beginning Analysis

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

CERT

About The Training

This Course Is...

- A "How-To" session on analyzing network flow data
- A tutorial with "What is this" and "Try It" exercises
- Part of a suite of courses being offered

This Course Assumes...

- Fundamental TCP/IP knowledge
- Interest in Network Operations or Security

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

Logistics

- Schedule
- Facilities
- Network Connection
- Handouts
- Training Material and Analysts' Handbook
- Quick Reference Guides

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

Why Flow Analysis?

What type of analysis has been done with flow?

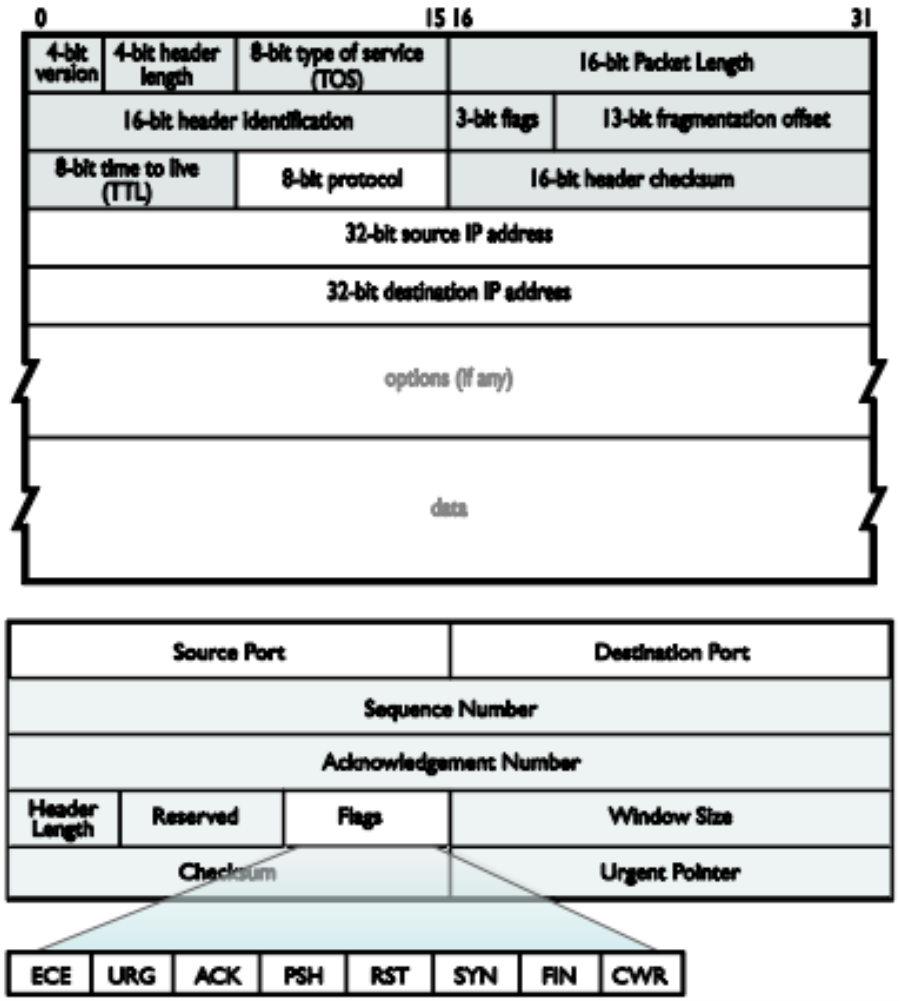
- Forensics support
- Inventory for large, dynamic networks
- Usage profiling (bandwidth studies)
- Waste (how much traffic is recreational)?
- Identify worm precursors
- Spam Detection

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

In The Beginning...

First there were packets

- Source & Destination IP
- Size Options
- Routing Information (TTL, fragmentation, QoS, etc) &



Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

CERT

Then Protocols added Ports

Accounting for Packets

The ISP asks "Who Is Using My Bandwidth?"

- Only the routers know
- Very high volume
- Routers know packets, we need them to summarize

NetFlow was developed by Cisco Systems in 1996

- Proprietary
- Evolved into the primary network accounting method
- Supported by most major routers
- IETF Standard on IPFIX (based on Cisco Netflow v9), RFC5101

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

What Is a Flow?

A flow is an aggregated record of packets
SiLK flows are defined by five unique attributes

- Source and Destination IP
- IP Protocol (TCP, UDP, ICMP, IPSec, etc)
- Source and Destination Port

These five keys form a “tuple”

- Similar to a “primary key” in a database record

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

What does a flow know?

Each unique flow (tuple) has associated attributes

- Timing (start, stop)
- Volume (packets, bytes)
- TCP flags
- Collection Location (sensor, traffic type)
- [Next Hop IP]

Flows get flushed when they close

- Timeouts, TCP FIN/RST, Router resources low

Our flows are *unidirectional*

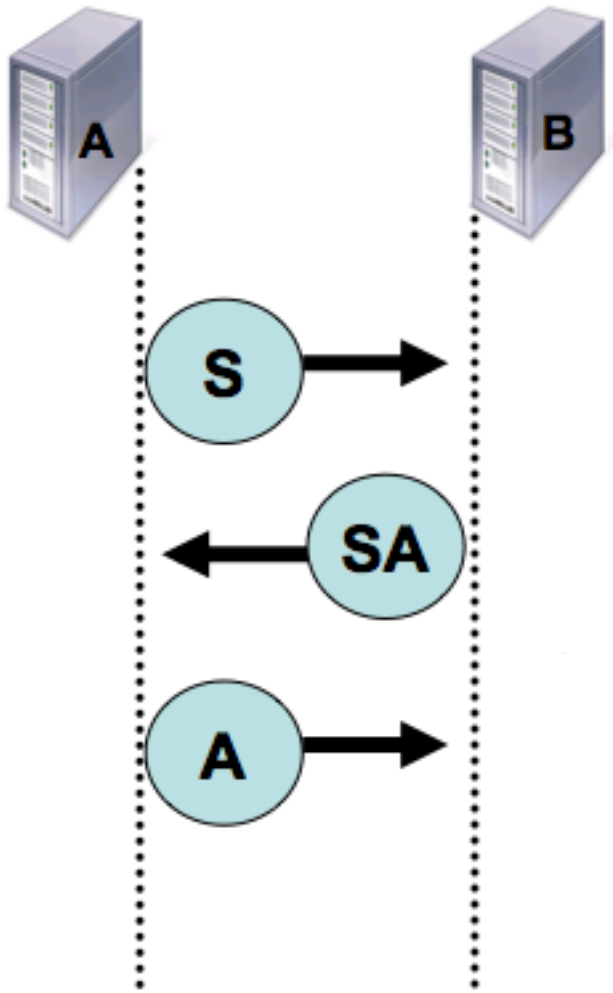
- The unique key is [sIP, sPort, dIP, dPort, protocol]

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

Flows are half duplex

For the simple example of a TCP 3-way handshake, consider how flows are counted.

- Flow 1 is created when the sensor observes the first packet between hosts A and B.
- Flow 2 is created with the second packet. Swapped IPs means a new flow.
- With the third packet, and ACK, Flow 1 is updated since the source and destination addresses and ports match. updated.



Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

CERT

What Is This #1

```
            sIP|            dIP|sPort|dPort|pkt| bytes|flags|
63.236.206.174|    72.24.144.5|44800|   25| 21| 19606|FS PA|
   72.24.144.5| 63.236.206.174|   25|44800| 17|  1066|FS PA|
63.236.206.174|    72.24.144.5|44800|   25|  1|    40|  R  |
63.236.206.174|    72.24.144.5|44800|   25|  1|    40|  R  |
63.236.206.174|    72.24.144.5|44800|   25|  1|    40|  R  |
63.236.206.174|   72.24.146.90|44800|   25|  1|    40|  R  |
  72.24.146.90| 63.236.206.174|   25|44800|  1|    49|F  PA|
```


Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

What Is This #2

```
            sIP|            dIP|pro|pkts|bytes|          sTime|
 66.142.134.179|72.24.150.186|  1|   2|  122|00:00:00.582|
 66.142.134.179|72.24.148.123|  1|   2|  122|00:00:00.911|
 66.142.134.179| 72.24.146.95|  1|   2|  122|00:00:01.783|
 66.142.134.179|72.24.159.123|  1|   2|  122|00:00:01.895|
 66.142.134.179|72.24.145.227|  1|   2|  122|00:00:02.220|
 66.142.134.179| 72.24.154.87|  1|   2|  122|00:00:02.329|
 66.142.134.179|72.24.149.212|  1|   2|  122|00:00:02.550|
 66.142.134.179| 72.24.158.18|  1|   2|  122|00:00:02.766|
 66.142.134.179| 72.24.150.34|  1|   2|  122|00:00:02.875|
 66.142.134.179|72.24.153.102|  1|   2|  122|00:00:02.879|
 66.142.134.179| 72.24.144.61|  1|   2|  122|00:00:03.421|
 66.142.134.179|  72.24.129.2|  1|   2|  122|00:00:03.530|
 66.142.134.179|72.24.129.224|  1|   2|  122|00:00:03.642|
 66.142.134.179|72.24.151.196|  1|   2|  122|00:00:04.184|
```


Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

What Is This #3

```
          sIP|          dIP|sPort|dPort|pkt|flags|      sTime|
72.24.144.12|   68.8.27.65|63126|   80|  7|FS PA|00:01:31.232|
  68.8.27.65| 72.24.144.12|   80|63126|  5|FS PA|00:01:31.232|
72.24.144.12|   68.8.27.65|63277|   80|  8|FS PA|00:01:42.642|
  68.8.27.65| 72.24.144.12|   80|63277|  8|FS PA|00:01:42.642|
72.24.144.12|   68.8.27.65|63330|   80|  7|FS PA|00:01:51.052|
  68.8.27.65| 72.24.144.12|   80|63330|  5|FS PA|00:01:51.052|
                         [pause]
72.24.144.12|   68.8.27.65|63707|   80|  8|FS PA|00:02:47.722|
  68.8.27.65| 72.24.144.12|   80|63707|  8|FS PA|00:02:47.831|
                         [pause]
72.24.144.12|   68.8.27.65|63957|   80|  8|FS PA|00:03:20.036|
  68.8.27.65| 72.24.144.12|   80|63957|  8|FS PA|00:03:20.036|
                         [pause]
72.24.144.12|   68.8.27.65|64504|   80|  8|FS PA|00:04:12.501|
  68.8.27.65| 72.24.144.12|   80|64504|  8|FS PA|00:04:12.501|
```


Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

What Is This #4

```
          sIP|          dIP|sPort|dPort|pkts|flags|       sTime|
72.24.129.20|82.80.30.150|   80| 1220|  152| S PA|00:00:23.602|
82.80.30.150|72.24.129.20| 1220|   80|   90| SRPA|00:00:23.602|
72.24.129.20|82.80.30.150|   80| 1221|1126| S PA|00:00:23.710|
82.80.30.150|72.24.129.20| 1221|   80|  413| SRPA|00:00:23.710|
72.24.129.20|82.80.30.150|   80| 1223|   63| S PA|00:00:26.341|
82.80.30.150|72.24.129.20| 1223|   80|   39| S PA|00:00:26.341|
72.24.129.20|82.80.30.150|   80| 1224|    8| S PA|00:00:26.883|
82.80.30.150|72.24.129.20| 1224|   80|    7| SRPA|00:00:26.883|
82.80.30.150|72.24.129.20| 1223|   80|    1|  R A|00:01:33.068|
```


Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

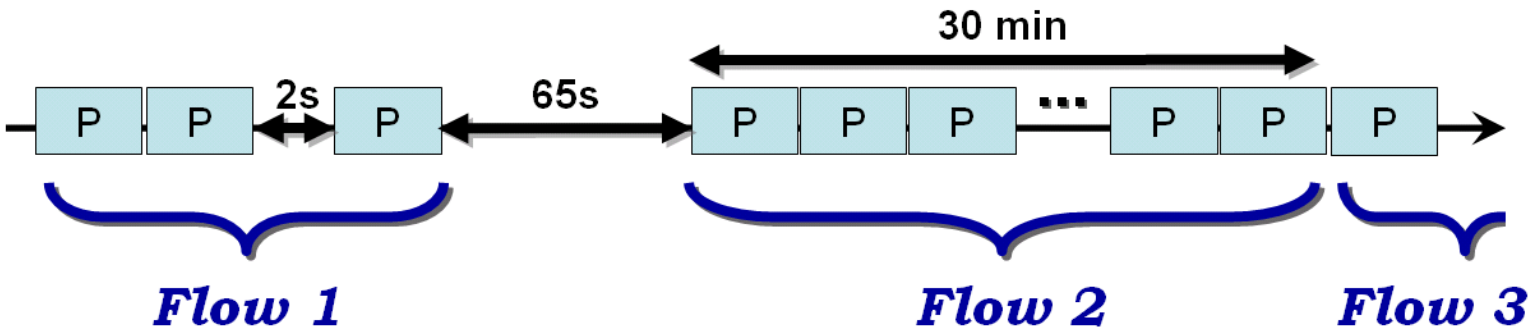
It's all a matter of timing

The flow buffer has to be kept manageable
Inactivity timeout:

- If there's no activity within [30] seconds, flush the flow

Active timeout:

- Flush all flows open for [30] minutes



Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

What Is This #5

sIP	dIP	sPort	dPort	pro	Pkt	byte	sTime	dur
8.97.138.194	72.24.145.68	500	500	17	1	112	00:02:31	0.000
72.24.145.68	8.97.138.194	500	500	17	1	112	00:02:31	0.000
8.97.138.194	72.24.145.68	500	500	17	2	224	00:13:53	40.498
72.24.145.68	8.97.138.194	500	500	17	2	224	00:13:53	40.498
8.97.138.194	72.24.145.68	500	500	17	2	224	00:25:10	45.582
72.24.145.68	8.97.138.194	500	500	17	2	224	00:25:10	45.582
8.97.138.194	72.24.145.68	500	500	17	1	112	00:36:03	0.000
72.24.145.68	8.97.138.194	500	500	17	1	112	00:36:03	0.000
8.97.138.194	72.24.145.68	500	500	17	1	112	00:43:19	0.000
72.24.145.68	8.97.138.194	500	500	17	1	112	00:43:19	0.000
8.97.138.194	72.24.145.68	500	500	17	3	336	00:47:30	46.088
72.24.145.68	8.97.138.194	500	500	17	3	336	00:47:30	46.088
72.24.145.68	8.97.138.194	500	500	17	1	112	00:53:32	0.000
8.97.138.194	72.24.145.68	500	500	17	1	112	00:53:32	0.000
72.24.145.68	8.97.138.194	500	500	17	2	208	00:58:42	0.000
8.97.138.194	72.24.145.68	500	500	17	20	2232	00:58:49	90.095

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

What Is This #6

```
          sIP|          dIP|sPort|dPort| pkts|flg|       sTime|     dur|
 72.24.147.6|58.210.70.72|35282|   22|29640|PA |00:00:11.361|1800.63|
58.210.70.72| 72.24.147.6|   22|35282|29633|PA |00:00:11.911|1800.08|
 72.24.147.6|58.210.70.72|35282|   22|30824|PA |00:26:23.092|1800.82|
58.210.70.72| 72.24.147.6|   22|35282|30825|PA |00:26:23.092|1800.82|
 72.24.147.6|58.210.70.72|35282|   22|29346|PA |00:56:24.020|1800.90|
58.210.70.72| 72.24.147.6|   22|35282|29347|PA |00:56:24.020|1800.90|
 72.24.147.6|58.210.70.72|35282|   22|31107|PA |01:00:10.783|1800.20|
58.210.70.72| 72.24.147.6|   22|35282|31113|PA |01:00:11.301|1800.68|
 72.24.147.6|58.210.70.72|35282|   22|29227|PA |01:26:25.036|1800.95|
58.210.70.72| 72.24.147.6|   22|35282|29228|PA |01:26:25.036|1800.95|
 72.24.147.6|58.210.70.72|35282|   22|30880|PA |01:56:26.096|1800.82|
58.210.70.72| 72.24.147.6|   22|35282|30878|PA |01:56:26.096|1800.82|
 72.24.147.6|58.210.70.72|35282|   22|30302|PA |02:00:11.301|1800.65|
58.210.70.72| 72.24.147.6|   22|35282|30287|PA |02:00:11.843|1800.10|
 72.24.147.6|58.210.70.72|35282|   22|31998|PA |02:26:27.028|1800.90|
58.210.70.72| 72.24.147.6|   22|35282|31999|PA |02:26:27.028|1800.90|
 72.24.147.6|58.210.70.72|35282|   22|32764|PA |02:56:28.040|1800.88|
```


Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

What Is This #7

sIP	dIP	sPort	dPort	pkt	flags	sTime	dur
72.24.144.17	10.25.235.38	40395	80	45	S PA	1:59:34.81	1759.18
10.25.235.38	72.24.144.17	80	40395	44	S PA	1:59:34.81	1759.07
10.25.235.38	72.24.144.17	80	40395	40	PA	2:29:39.82	1797.62
72.24.144.17	10.25.235.38	40395	80	40	A	2:29:39.93	1797.51
10.25.235.38	72.24.144.17	80	40395	40	PA	3:00:23.46	1800.17
72.24.144.17	10.25.235.38	40395	80	40	A	3:00:23.57	1800.17
10.25.235.38	72.24.144.17	80	40395	40	PA	3:31:09.83	1797.52
72.24.144.17	10.25.235.38	40395	80	40	A	3:31:09.93	1797.52
10.25.235.38	72.24.144.17	80	40395	40	PA	4:01:53.42	1797.72
72.24.144.17	10.25.235.38	40395	80	40	A	4:01:53.51	1797.64
10.25.235.38	72.24.144.17	80	40395	35	RPA	4:32:37.18	1560.50
72.24.144.17	10.25.235.38	40395	80	34	A	4:32:37.29	1520.89
72.24.144.17	37.52.53.241	40395	80	13	FS PA	5:18:41.57	0.48
37.52.53.241	72.24.144.17	80	40395	18	FS PA	5:18:41.63	0.43
72.24.144.17	42.15.190.19	40395	80	9	FS PA	8:21:01.15	4.14
42.15.190.19	72.24.144.17	80	40395	6	FS PA	8:21:01.15	4.14
42.15.190.19	72.24.144.17	80	40395	1	A	8:21:05.29	0.00
72.24.144.17	10.46.227.72	40395	80	7	FS PA	9:21:24.36	0.22
10.46.227.72	72.24.144.17	80	40395	6	FS PA	9:21:24.47	0.22
72.24.144.17	18.113.57.14	40395	80	6	FS PA	9:39:43.67	0.11
18.113.57.14	72.24.144.17	80	40395	4	FS PA	9:39:43.67	0.21

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

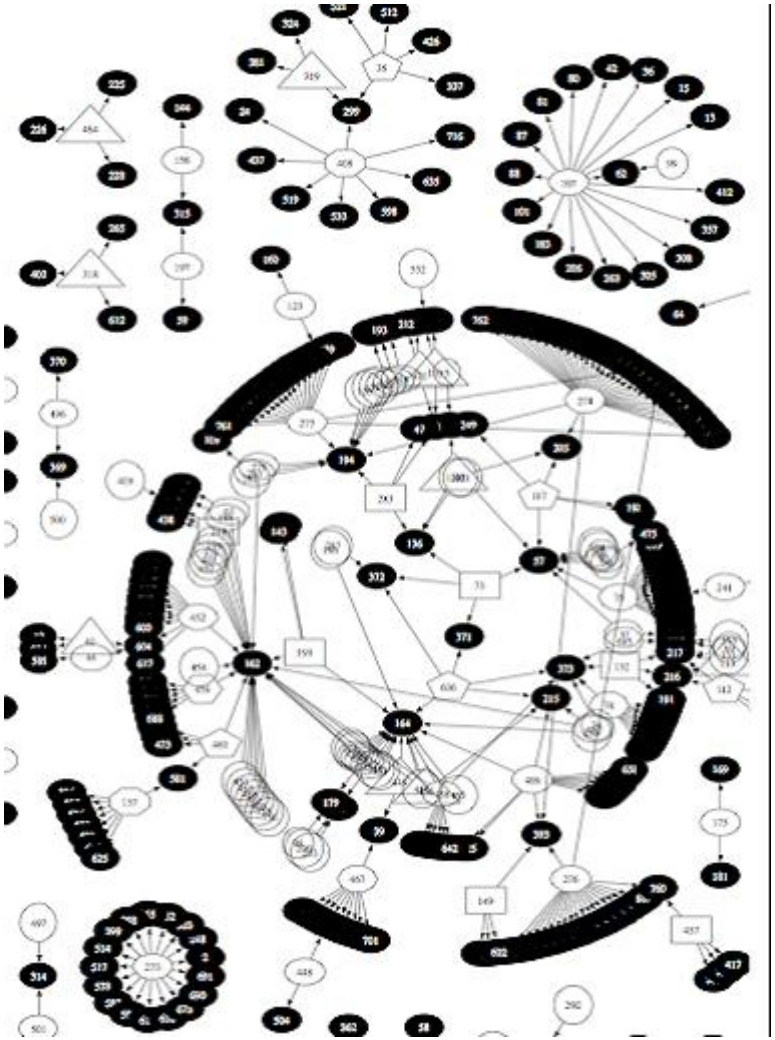
CERT

Where do I collect flows?

Flow is often collected at the border

- Watch internal and external communications
- Identify services on your network
- Identify resources your machines use regularly

Most routers can generate flows



Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

CERT

Flow vs. IDS

IDS

+ Content inspection
  - Presents an interpretation of raw data
  - Tuning means discarding false positive data

Flow

- No content available

+ Gives direct observations
+ No tuning, keep everything

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

Flow vs. Firewall

Firewalls

+ Block unwanted traffic
  - Not intended as a historial record; logging is secondary

Flow

- Completely passive
+ Logging is primary
+ Audits the firewall

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

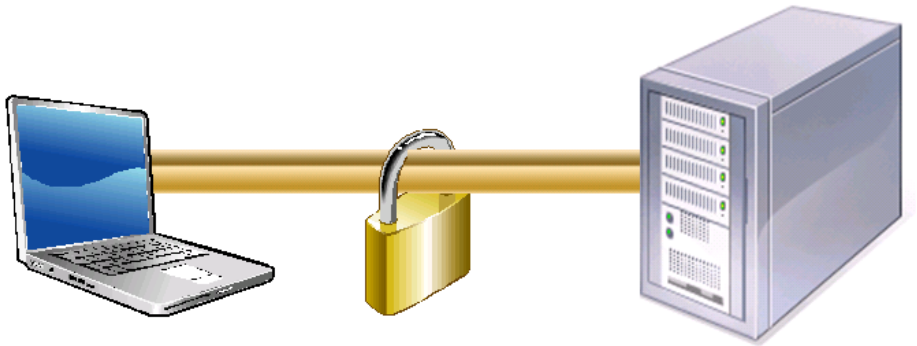
Got a question? Flow can help.

- What's on my network?
- What happened before the event?
- Where are policy violations occurring?
- What are the most popular web sites?
- How much volume would be reduced with a blacklist?
- Do my users browse to known infected web servers?
- Do I have a spammer on my network?
- When did my web server stop responding to queries?
- Who uses my public DNS server?

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

About ssh

- ssh creates a secured connection between your computer and the ssh server
- ssh is your primary tool for moving things between you and your analysis server



Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

Try It #1!

Login
Check for access to data (`ls /data`)
Type "`which rwfilter`"
Type "`rwfilter --help | more`"
Logout (optional)

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

Try It #2!

Move a file from the server to your workstation:

```
scp server:/remote/path/to/file.ext
/local/directory/
```

Move a file from your workstation to the server:

```
scp /local/file.ext server:/remote/directory/
```


Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

Text Editors

Write only: `echo "blah" > file`
Simple: `vi`
Flexible but not always available: `emacs`
Other simple text file tools

- `cat`: print it out
- `more`, `less`: print it out one page at a time
- `head`, `tail`: print out just the beginning (or end)
- `wc -l`: count the number of lines

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

Getting around

Some other commands you may need:

- `cd`: change directory
- `ls`: list the current directory contents
- `mkdir`: make a directory
- `rm`: remove a file
- `cp`: copy a file
- `logout` or `exit`: log out

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

Try It #3!

1. Create a sample file on the server.
2. Move the file from the server to your local machine, and open it in the local text editor. Change the file and move it back to the server.
3. Use head and tail to display the second line of a file which contains 5 lines.

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

Where's the GUI command prompt?

- It's not quite the same as Windows:
- You'll always be working from a command prompt
- We'll be doing lots of text manipulation
- There's occasional CR-LF messiness
- Data can get big, but that's usually OK

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

Get used to using pipes

- Pass output from one command as input to another
- Stop things with ctrl+c
- Also watch out for ctrl+s (suspend), restart with ctrl+q
- Also watch out for ctrl+z (put in background), continue with `fg`

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

About SiLK

- The System for internet Level Knowledge
- `http://tools.netsa.cert.org`
- Packing System
  - Accepts Netflow
  - Stores data in a very space-efficient binary flat file

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

Analysis Suite

- Used to query binary flat files from the packing system
- Some mirror Unix text tools for operate on binary flow files e.g., cut, uniq, sort, split
- Some work with large IP data collections sets, bags and prefix maps
- All support ad-hoc analysis needs

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

What SiLK Does

- Optimized for extremely large data collections
- Very compact record format
- Large amount of history can stay on line
- Command line interface
- Keep data in the native binary format as long as possible
- Retrospective analysis
- Most useful for analyzing past network events
- May feed an automated report generator
- Good for forensics (what happened before the incident?)

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

Flavoring your Flows

Without content data, flows often seem very bland
SiLK flavors flow data with add-ons:

- Address sets e.g., blacklists
- Address Bags give a value to an address
- Prefix Maps give an arbitrary label to a group of addresses e.g., Country Code Mapping
- Hooks for custom libraries

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

About Classes and Types

SiLK assigns each flow record a CLASS and a TYPE

Class

- Duplicates the purpose of the router
- Sample classes might be Border, Internal, Customer
- We will simply use "All"

Type

- Separate inbound from outbound
- Queries often run against a single type to improve performance
- Other types are common also
- in, inweb, out, outweb, null

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

The Flow Repository

The Repository: a directory structure holding binary flow files

Directory structure based on:

- Sensor
- Type
- Year
- Month
- Day

File name based on:

- Type
- Sensor
- YYYYMMDD.HH

All times are GMT

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

Try It #4!

We're using anonymized flow in the repository at /data. SSH in to the server and determine:

1. Which dates is data available for?
2. What classes and types of data are available?

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

The Training Repository

Based on LBNL Anonymized data set
`http://www.icir.org/enterprise-tracing/Overview.html`
Sensor name and date/time locates data within the repository

- S0 – anonymized general flows
- S1 – anonymized scanning flows, different anonymization
- Selected dates and times in 2004 and 2005
- Avaliable data types:
  - out, outweb: source internal, destination not internal
  - in, inweb: source not internal, destination internal

Timeouts

- 1800s (30 min) active timeout
- 60s inactive timeout

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

Dates in Sample Data

- 2004/10/04:20-22
- 2004/12/15:08-23
- 2004/12/16:01-06,16-23
- 2004/12/17:00-03
- 2005/01/06:19-23
- 2005/01/06:00-06,10-23
- 2005/01/08:00-05

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

Try It #5!

We've glossed over the nuance of how SiLK handles ICMP flows. Type in the following command and look at the output:

```
rwfilter --type=in --start-date=2004/10/04 \
--protocol=1 --max-pass-records=10 \
--pass-destination=stdout \
| rwcut --fields=sip,sport,dip,dport,icmptypecode
```

1. How does SiLK store ICMP type and code information?
2. What did this command actually do?

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

What have we done so far?

- An Introduction to Flow
- A Brief Discussion of Unix
- A Flow Analysis Teaser

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

Section Outline

Basic SiLK Tools

- rwfilter
- Printing and Sorting Tools
- Counting Tools
- Other Tools

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

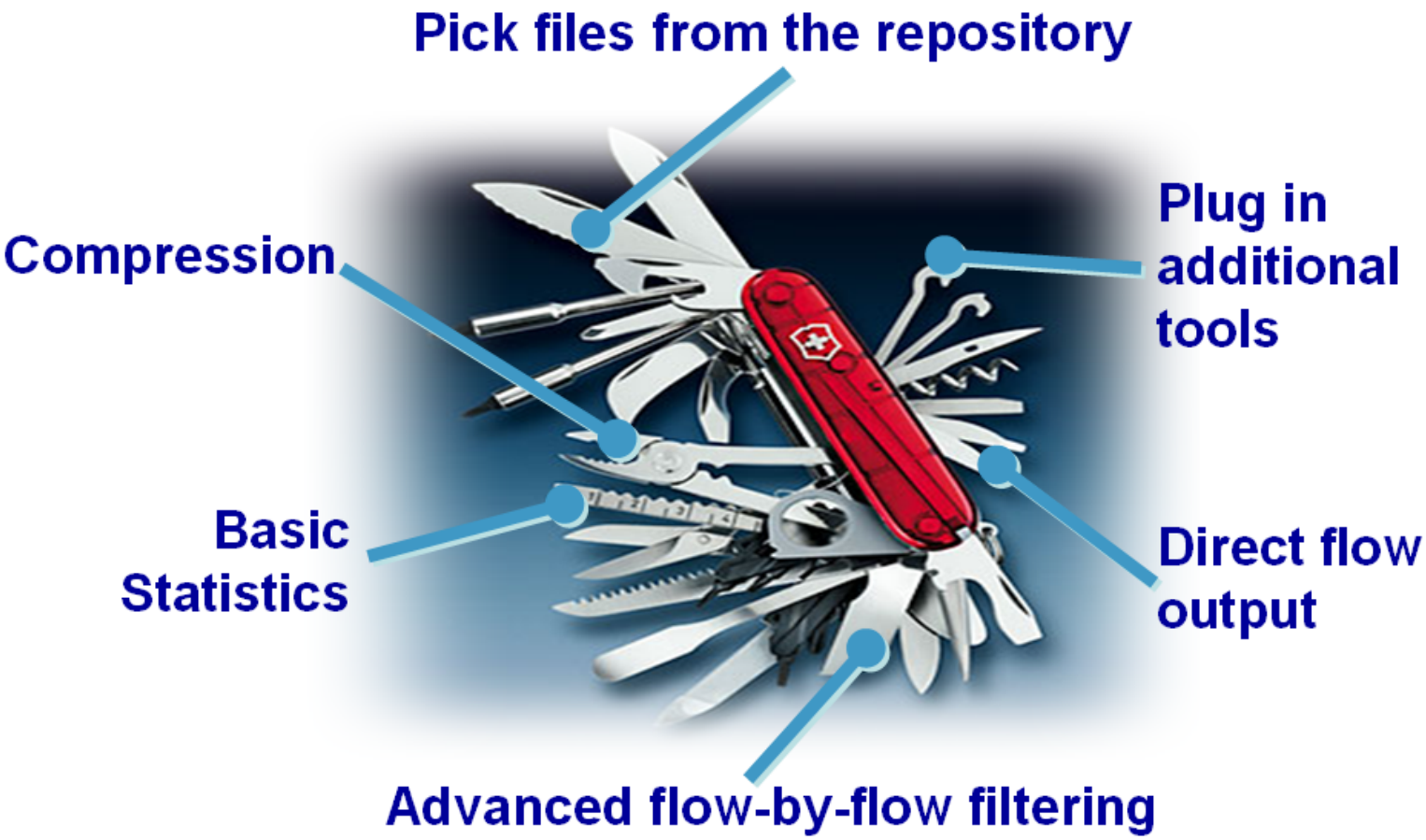
So much to do, so little time...

We can't discuss all parameters for every tool Resources

- Analyst's handbook
- SiLK Reference Guide (hardcopy man pages)
- `rw[something] --help`
- `man rw[something]`
- `http://tools.netsa.cert.org`

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

rwfilter



Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

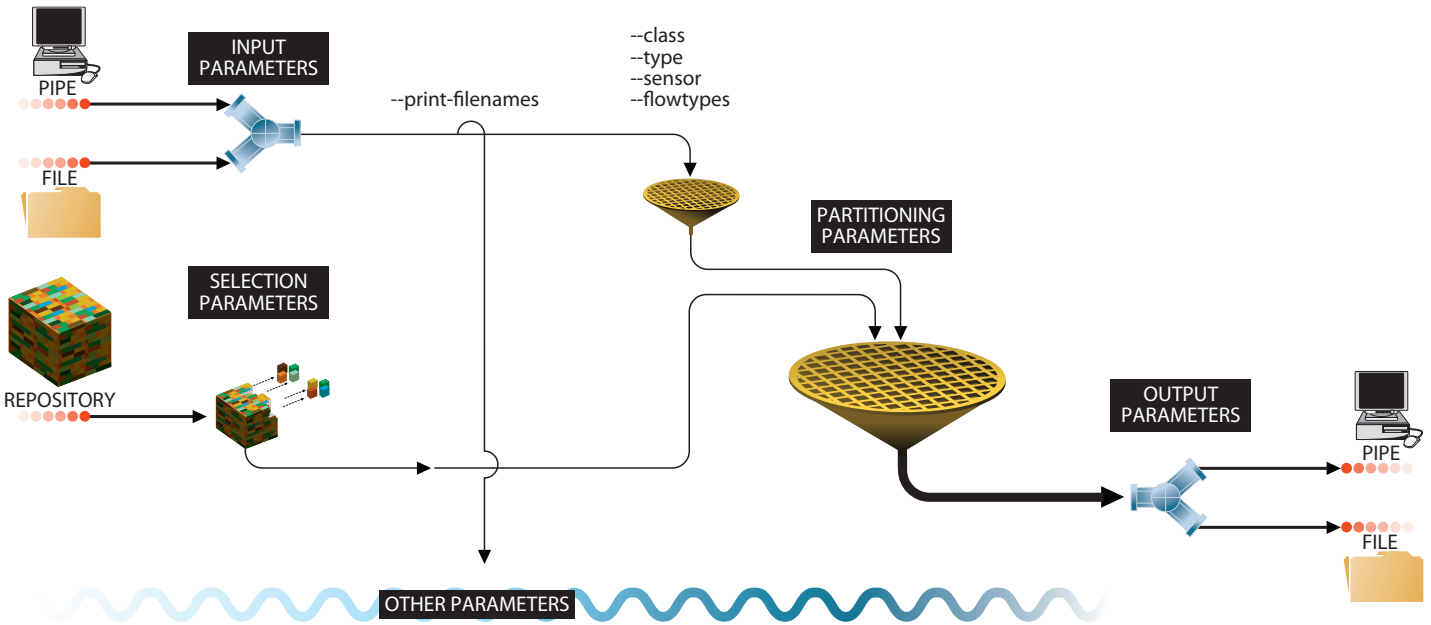
CERT

rwfilter Command Structure

- Most of the time: any order of parameters
- Parameters may be abbreviated to unique prefix
- Five different groups of parameters:
  - Input – file, repository, pipe
  - Selection – which part of repositiory
  - Partitioning – which flows among the selected
  - Output – going where (pipe, file)
  - Other – IP version, filter statistics, etc.

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

rwfilter Command Flow



Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

CERT

rwfilter Requirements

Each rwfilter call must have:

- Somewhere to get records from:
  - File name
  - `--input-pipe=stdin` or other pipe
  - Repository (default or `--data-rootdir=./myarchive`) with selection parameters (type, sensor, start-date, end-date, class)
- Some description of what records are wanted (partitioning parameters)
- Some description of where records should go:
  - `--pass=myfile.rw`
  - `--fail=stdout`
  - `--print-statistics`

```
rwfilter --start-date=2008/12/05:00 \
--end-date=2008/12/05:03 --type=all \
--protocol=6 --packets=1-3 --pass=dec05.rw
```

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

Selection Parameters

These options control access to repository files

- `--start-date=2007/10/03:00`
- `--end-date=2007/10/03T03`
- `--sensor=S0`
- `--class=all`
- `--type=in`

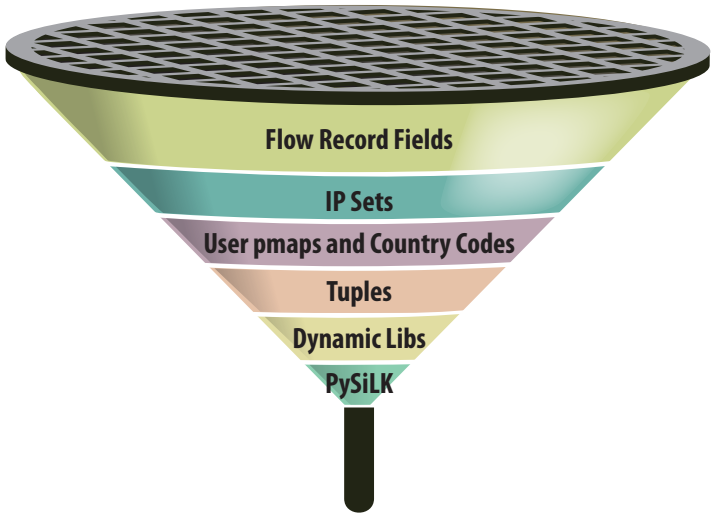
Alternatively, use a pipe or a file

- `--input-pipe=stdin` – Useful for chaining filters through stdin/stdout
- `myfile.rw` – Useful for filtering previous results

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

Partitioning Parameters

Partioning is the most complex



- Partitioning parameters form an “and” expression
- Too few parameters means too much output
- Can refine partitioning with another rwfilter call
- Some of these are beyond the scope of this course

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

CERT

Flow partitioning parameters: Record Fields

Pass records based on flow record fields; one is required

- `--[not-]saddress, --[not-]daddress`: Wildcard like 12.5,7,9.2-250.x
- `--protocol`: IP protocol
- `--sport, --dport, --aport` TCP, UDP ports (caveat: ICMP)
- `--tcp-flags=SF; --flags-all=S/SA; --fin-flag;...`
- `--icmp-type; --icmp-code`
- `--bytes, --packets, --bytes-per-packet`

At least one partitioning parameter is required

- Use `--proto=0-` to pass all

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

Flow partitioning parameters: Flow Record Time

start-date, end-date choose repository files, but do not look at the actual flow records

- `--stime`, `--etime`: choose flows that start (or end) within a time range
- `--active-time`: flows active in a time range
- Time format: YYYY/MM/DD:HH:MM:SS
- Time range format: [Time]-[Time]

Duration

- `--duration=1-10`: number of seconds the flow was active

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

Flow partitioning Parameters: Flags

- `--tcp-flags=[FSRPAUEC]`
- `--fin-flag`, `--syn-flag`, etc.
- `--flags-all=[FSRPAUEC]/[FSRPAUEC]`
- `--flags-initial=[FSRPAUEC]/[FSRPAUEC]`
- `--flags-session=[FSRPAUEC]/[FSRPAUEC]`

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

Flow partitioning Parameters: Advanced

Some of these will be discussed later:

- `--max-pass`: limit the number of records passed
- `--sipset`, `--dipset`, etc: limit to specfici IP addresses
- `--ipport`: IP/port pairs
- `--pmap`; prefix map
- `--dynamic-library`: dynamically loaded library
- `--scc`, `--dcc`: country codes
- compression

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

Output Parameters

rwfilter leaves the flows in binary(compact) form

- `--pass`, `--fail`: direct the flows to a file or pipe
- `--all`: destination for everything pulled from the repository
- One output is required but more than one can be used

Other useful output:

- `--print-statistics`
- `--print-volume-statistics`

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

Other Parameters

- `--dry-run`: test the command (useful for scripting)
- `--ipversion=6`: process IPv6 data
- `--print-filenames`: print files from which flow records came
- `--help`: print condensed help text
- `--man`: print manual page
- `--version`: print configuration info
- `--threads`: parallelize rwfilter run

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

Try It #6!

The time to run an initial query against the repository often depends on the number of files which will be accessed. How many files in the repository will be opened with this command?

```
rwfilter --sensor=s0 \
  --start-date=2004/12/15:19
```

(note: you have to add extra parameters to this command to make it work)

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

Try It #7!

Often you will want to track an individual address or address block. Develop a filter command to retrieve:

- Flows to the 131.243.10.0/24 CIDR block,
- Leaving our network,
- On 12/16/2004 at 17:00 GMT,
- And save the flows in the file `netblock.rw`.
  How many packets, bytes and flows were retrieved?

How many packets, bytes, and flows were retrieved?

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

Try It #8!

Let's look for short, bursty outbound ssh traffic. Develop a filter command that does the following:

- Pulls out all outbound ssh (TCP port 22) flows,
- On 12/17/2004,
- Between 00:00 and 04:00 GMT,
- That lasted less than 60 seconds,
- With an average of more than 60 bytes per packet,
- And store the result in a file named short-ssh.raw

How many records did you retrieve? How many files in the repository were opened?

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

Try It #9!

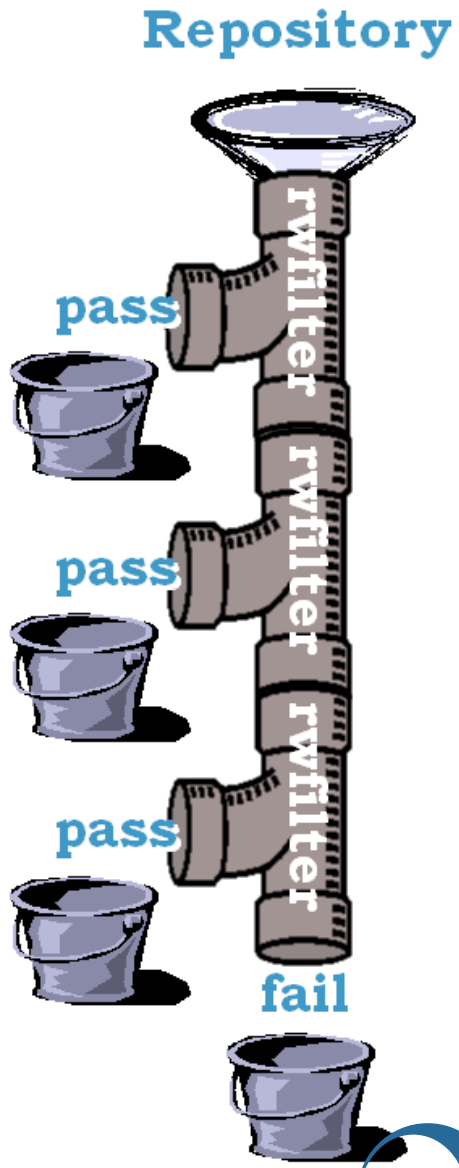
Examine traffic trends. What is the change in mail traffic volume between 19:00 and 20:00 hours on 12/15/2004 for the mail server at 128.3.26.249?

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

Chaining filters

It is often very efficient to chain rwfilter commands together

- Use `--pass` and `--fail` to segregate bins
- Use `--all` so you only pull from the repository once



Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

CERT

What Is This #8

```
rwfilter \
  --start-date=2007/09/30 \
  --type=outweb \
  --bytes=100000- \
  --pass=stdout \
| rwfilter \
  --input-pipe=stdin \
  --duration=60- \
  --pass=long-http.rw \
  --fail=short-http.rw
```


Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

Try It #10!

Let's revisit the last example for some more analysis. For the mail server at 128.3.26.249, and looking only at outbound traffic for the 19:00 hour on 12/15/2004, use a single command to find out both:

- The total number of SMTP flows (TCP port 25), and
- The number of flows which were for outbound messages

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

Common rwfilter Typos

- `--port` or `--destport`: not an option name
- `--saddress=file`: pointing to a filename; should be an IP
- `--sip=10.1.2.3`: sip specifies an IPSet; use saddr for addresses
- `--start=2005/11/04:06:00:00` start-date and end date use only down to the hour
- `---start-date`: should be only two dashes
- `-- start=2007/05/22`: no space between `--` and the option

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

But I can't read binary...

rwcut provides a way to display binary records as human-readable ASCII

- Useful for printing flows to the screen
- Useful for input to text processing tools
- You'll usually only need the `--fields` argument

sip	packets	sval	flags	application
dip	bytes	dval	initialflags	icmptypecode
sport	Sensor	in, out	sessionflags	attributes
dport	scc	dur	dur+msec	type
protocol	dcc	stime	stime+msec	stype
class	nhip	etime	etime+msec	dtype

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

Pretty Printing SiLK Output

Default output is fixed-width pipe delimited data

```
                sIP|            dIP|pro|pkts|bytes|
   207.240.215.71| 128.3.48.203|  1|    1|   60|
   207.240.215.71|  128.3.48.68|  1|    1|   60|
   207.240.215.71|  128.3.48.71|  1|    1|   60|
```

Tools with text output have these formatting options

- `--no-titles`: suppress the first row
- `--no-columns`: suppress the spaces
- `--delimited`; `--column-separator`
- `--legacy-timestamps`: better for import to Excel

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

Try It #11!

Create a file ssh.rw that contains all outbound SSH flows from 12/16/2004:17. Experiment with rwcut and Unix text tools to try and sort out records:

1. Can you tell which flows are from internal SSH servers, and which are from external SSH servers?
2. Which flows look like SSH keep-alives?
3. Which flows had the most data transfer?

Try to write rwfilter commands against ssh.rw to query these records, and display them with rwcut

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

rwsort

Why sort flow records?

- Records are recorded as received, not in time order (look at records from the last exercise)
- Analysis often requires finding outliers

rwsort options

- fields (same as rwcut) is required
- in, out (stdin / stdout are defaults)
- For improved sorts, specify a buffer size
- For large sorts, specify a temporary directory

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

I only believe what I see

You'll be tempted to work with text-based records

- It's easy to see the results and postprocess with other tools (e.g., perl)
- It takes a lot of space, and it's much, much slower

Guiding Principle: Keep flows in binary format as long as possible

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

Try It #12!

Often HTTP beaconing consists of very small HTTP requests. Let's get a feel for what HTTP data looks like, even before we start to find these beacons.
What do the smallest outbound HTTP web client flows look like on 12/15/2004?

- First, find them using rwsort
- Second, find them using sort
- Which was faster?

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

Counting Tools

The suite contains several counting tools:

- `rwcount` - count across time
- `rwaddrcount` - count across addresses
- `rwuniq` - count on arbitrary field combinations
- `rwstats` - descriptive statistics and counts

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

rwcount

Basic counting:

- `rwcount myfile.rw > count_file`
- Produces byte, packet and flow totals by time

Common Options:

- `--bin-size`: changes the size of each bin (in seconds)
- `--skip-zeroes`: should empty bins be printed?

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

rwaddrcount

Basic counting:

- `rwaddrcount --[print-option] myfile.rw`
- `--use-dest` to work with dIP; default is sIP

Print Options:

- `--print-stat`: Lists total number of addresses found
- `--print-ips`: Just print out the IP address, nothing else
- `--print-recs`: Lists bytes, packets, records, times for each address

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

rwstats

Great for generating top-N, bottom-N lists
Group by (choose one or two):

- Addresses
- Ports
- Protocols

Output Limit

- Count
- Top, Bottom
- Threshold (specific value range)
- Percentage

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

rwuniq

The more general case for rwstats
Mirrors the unix “uniq -c” command

- Creates a giant hash table where you define the key
- Memory is expensive, so we can’t uniq everything

Common Options:

- `--fields`: same as cutting and sorting
- `--all-counts`: collect bytes, packets and flows
- `--bin-time`: size the bins when uniq’ing on time

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

Try It #13!

Find scans traffic in the sample data (while the anonymizer removed some of the simple scans, they didn't find them all). When you find one, answer the following questions:

- What type of scan was it?
- When did it start/end?
- How fast was it?
- What did the scanner discover?

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

Try It #14!

This example showcases the very useful `--dip-distinct` feature of rwuniq:
For 2004/12/15, how many clients connected to the highest volume web servers?

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

Oops...I forgot where this came from...

rwfileinfo

- Each SiLK file (flows, sets, bags, prefix maps, etc.) has a header which logs data
- rwfileinfo prints out that data
- For flow files, it also (usually) keeps a history of the commands used to generate the file

Try It!

- rwfileinfo *.rw

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

When the files are LARGE

rwcat

- Send a binary rw file to stdout

rwappend

- Join multiple files together

rwsplit

- Carve large files into pieces

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

How long will this really take?

rwfglob

- Find out which files will be pulled from the repository
- Find out whats available and whats missing
- Use the output in other file-processing scripts

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

Maintain anonymity

rwnetmask

- Mask off low order bits of source and/or destination addresses

rwrandomizeip

- Randomly replace source and destination addresses

rwtuc

- Change text flow data into binary (opposite of rwcut)

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

Who was that?

rwresolve

- Perform a DNS lookup on text output
- Caveat: it uses your analysis host's DNS resolver
- Caveat: DNS is subject to change

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

What have we done so far?

Basic SiLK Tools

- rwfilter
- Printing and Sorting Tools
- Counting Tools
- Other Tools

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

Section Outline

Advanced SiLK Tools

- Sets
- Bags
- Prefix Maps

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

Blacklists, Whitelists, Books of Lists...

Too many addresses for the command line?

- Spam block list
- Malicious web sites
- Arbitrary list of any type of addresses

Create an IP set!

- Individual IP address in dotted decimal or integer
- CIDR blocks, 192.168.0.0/16
- Wildcards, 10.4,6.x.2-254

Use it directly within your filter commands

- `--sip`, `--dip`, `--anyset`

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

Set Tools

`rwset`: create sets from binary flows
`rwsetbuild`: create sets from text
`rwsetcat`: print out an IP set into text (**very useful**)
`rwsetmember`: test if IP is in given IP sets
`rwsettool`: perform set algebra (set, union, intersection) on multiple IP sets

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

Try It #15!

Flow is also very useful for creating network inventories. What /24 net blocks are populated within my network? Which block has the densest population?

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

Other uses of IP Sets

Perform set arithmetic on IP data

- What addresses on my spam blacklist are also bot infected?

Randomly select items for sampling

- `rwsettool --sample --size=100`

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

Bags: sets with attitude

Bags are generally IPSets with an associated integer

- Usually a count or sum
- Could also be ports or protocols

Bags can make sets

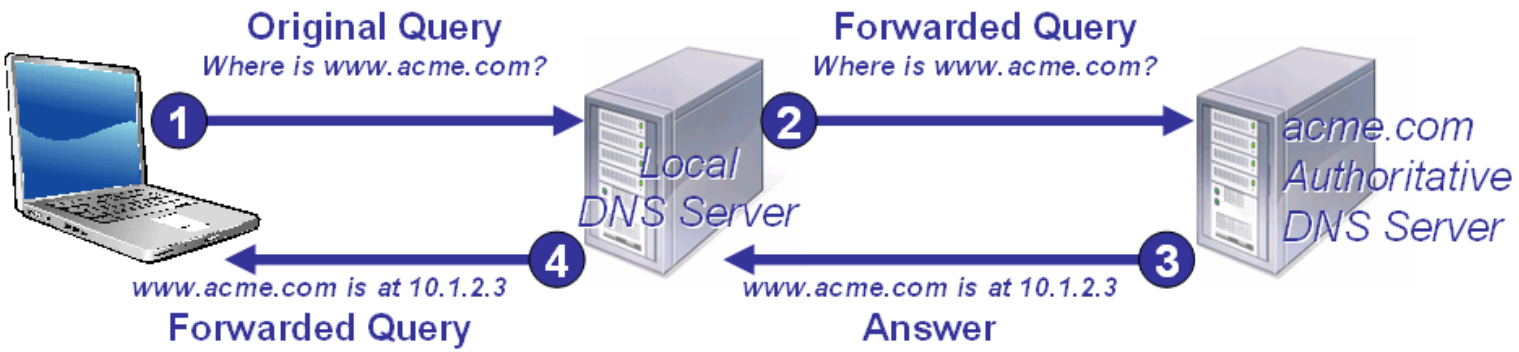
Math operations can be performed on bags

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

Try It #16!

Let's look for DNS clients that are using an external DNS resolver.

- First, let's take a moment to review DNS:
- When a client wants an address, it asks its local DNS server
- The local DNS server does all the work

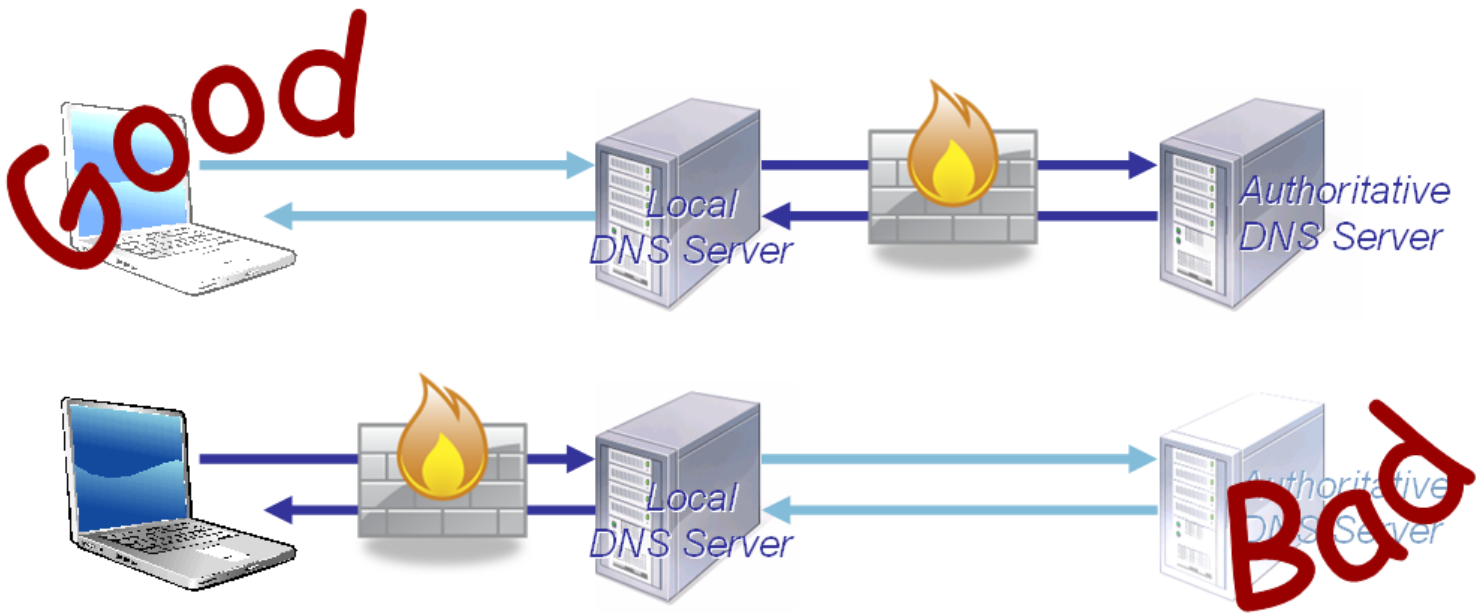


Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

Try It #16!(2)

The Local DNS Server should be local

- Can be assigned manually or by DHCP
- Up to three can be assigned, but often only one is used



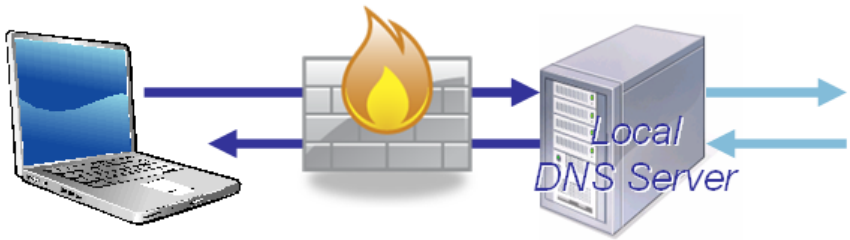
Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

CERT

Try It #16!(3)

Once again, let's look for DNS clients that are using an external DNS resolver:

- Use bags to count the number of outbound DNS connections per address,
- Create a candidate set from that bag of addresses with more than 100 outbound flows, and
- Find the number of unique destination addresses for the candidates.



Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

Prefix Maps(pmaps): sets with bling

Assign an arbitrary label to address prefixes

- Start with a text file of IP ranges and labels
- Order from least to most specific
- Compile the text file with rwpmapbuild
- Print out the pmap with rwpmapcat

The input file:

```
10.0.0.0/8                Private Unassigned
192.168.0.0/16            Private Unassigned
172.16.0.0/12             Private Unassigned
10.0.1.100 10.0.1.200 Workstation DHCP
10.0.1.1   10.0.1.50  Servers
10.0.2.1   10.0.2.50  Servers
10.0.3.1   10.0.3.50  DMZ Servers
```

No other pmap tools (?!?)

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

Using pmaps

pmaps don't have their own tools, they fit in with existing tools

- rwfilter
- rwsort
- rwcut
- rwuniq

This allows you to add your own fields to flow

- Query all your servers:
  `rwfilter --sval="Servers","DMZ Servers"`

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

Port-based pmaps

It's also possible to create prefix maps based on ports.

- Useful for well-known service ports; e.g., IRC, HTTP
- Also useful for ICMP

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

Try It #17!

Create an ICMP prefix map from the ICMP types (or types and codes).

- Look at unassigned ICMP type/code values that are in use. Which ICMP types receive the most traffic?
- Note: ICMP type/code values are assigned by RFC792; a summary table is available from IANA at `http://www.iana.org/assignments/icmp-parameters`

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

Are we there yet?

Advanced SiLK Tools

- Splitting and merging
- Sets
- Bags
- Prefix Maps

More reliance on examples to demonstrate these concepts

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

Scripting

Why script?

- Repeatable analyses
- Encapsulating syntax
- Composing complex commands

How to script?

- Shell scripting (we'll use bash)
  - Good reference: `http://tldp.org/LDP/abs/html/`
- Python (beyond this class, but widely used)
  - Good reference: `http://docs.python.org/tut/`

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

Getting started

Put your typed commands into a file

- From our second example, `ls -lR /data |grep "/data"`

Run the file

- `bash script.sh` or
- `sh script.sh` or
- `chmod +x script.sh; ./script.sh`

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

Using variables

`name="value"` (quoting is optional)

- Use the variable by adding a "$" prefix
- Everything is case sensitive

Display values with "echo"

- e.g., `echo "Variable var is $var"`

`$1, $2, ...` `$9` are input parameters

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

What Is This #9

```
cmd="rwfilter"
#  selection
cmd="$cmd --type=in,out"
cmd="$cmd --start=$1"
#  partition
cmd="$cmd --protocol=1"
#  output
cmd="$cmd --print-volume"
#
# execution
date>>logfile
echo "$cmd">>logfile
$cmd
```


Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

Loops

For loops

```
for variable in list
do
  stuff
done
```

`list` is a space-delimited set of values

While loops

```
while [ condition ]
do
    stuff
done
```

Note: The square brackets and spaces are required

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

Conditionals

```
if [[ condition ]]
then
  do something
elif [[ condition ]]
then
  do another thing
else
  do final thing
fi
```

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

Try It #18!

Let's do some investigation of spam activity.
Re-run the query from the spam exercise in the previous section, and save the results to a file. From the single text file, find:

- What is the worst spammer by byte count and by record count?
- Which unique blacklist entry numbers were seen?
- Did any spammers send to multiple internal mail servers?
- As you run the initial command, send it to the background, kill it, then run it again

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

Try It #19!

Trend the buildup of source addresses over the course of a day. Write a script that will:

- Accept an input date
- Create a cumulative set file without any addresses in it
- Then, for each hour in the input date:
  - Create a set of all the outbound source addresses
  - Add the outbound source addresses to the cumulative set
  - Count the addresses in the hourly and cumulative set files

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

What have we done so far?

Unix Scripting

- More Unix Commands and Techniques
  - Working with processes
  - Working with text
- Automating Daily Tasks
  - Variables
  - Loops
  - Conditions

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

Section Outline

Visualization

- Basic graphs
- The Excel Cookbook
- The Gnuplot Cookbook
- Advanced graphs

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

A picture is worth a thousand data points

Exploratory

- Throw something against the wall
- Is anything going on?

Explanatory

- Find the point inside the big picture
- Place the point in context
- Avoid technical obscurities

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

You've Got Style

Colors, patterns and shapes are important

- Useful for separation and contrast
- Draw attention to important features

Use multiple curves

- Compare and contrast data sets

Avoid unnecessary clutter

- Limit point count or point size
- Limit text size; avoid redundant text

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

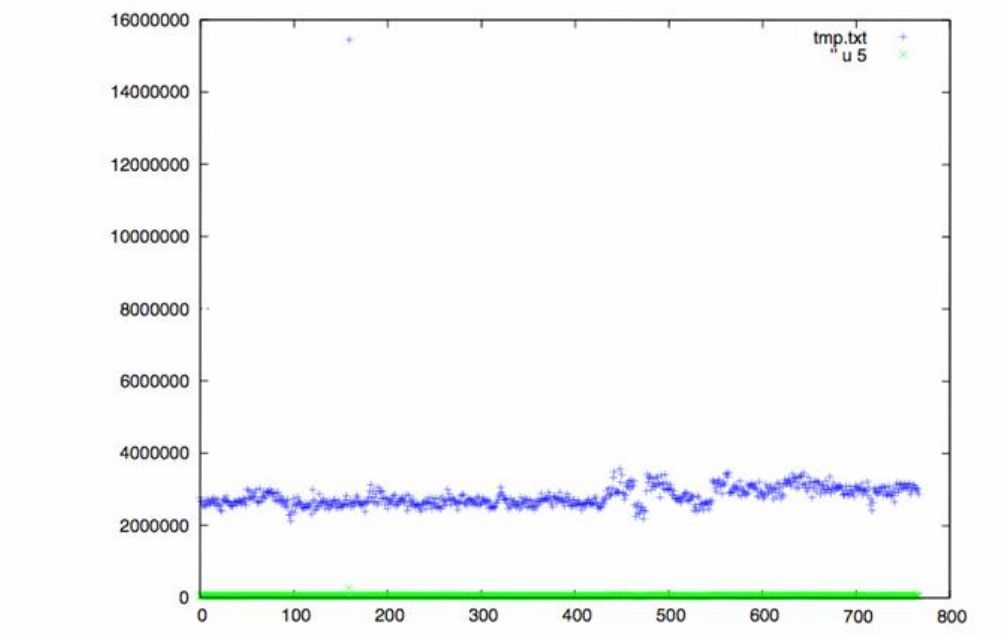
What I meant, not what I said

Avoid bad visualizations which

- Look overly complex
- Promote the insignificant
- Lose the context of the issue
- Contain no information or too much information
  - There's a difference between data and information

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

The Worst Plot Ever

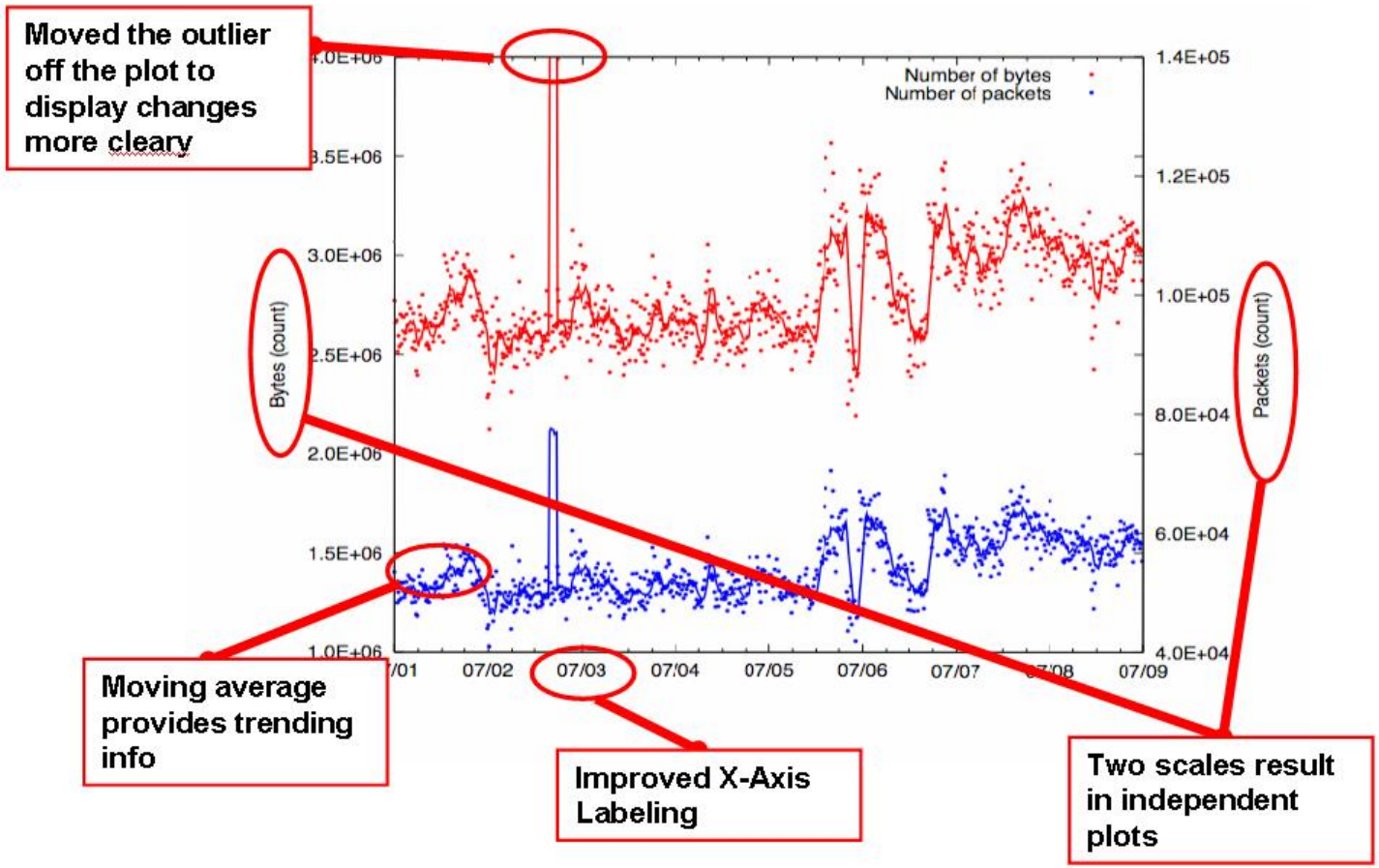


Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

Why is this the worst plot ever?

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

Same data, better plot



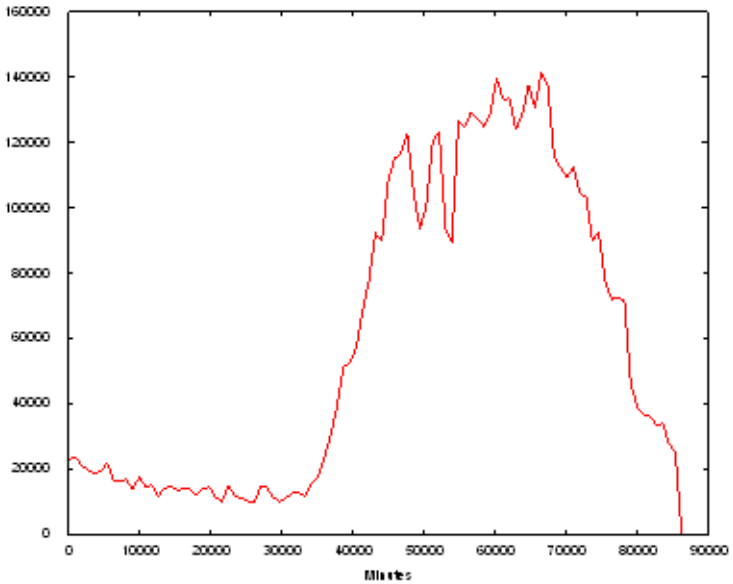
Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

CERT

Getting started: two primary types

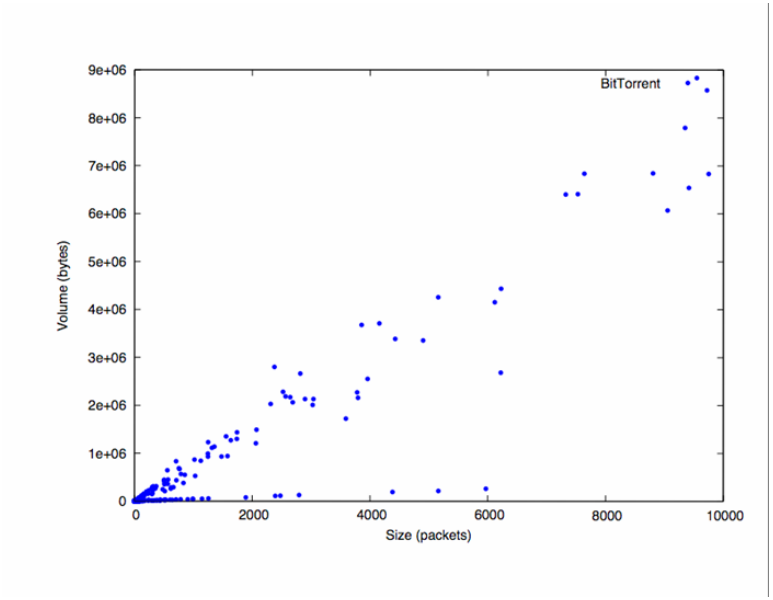
Time-series graphs

- Display how a value changed over time



Scatterplots

- Compare two distinct values



Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

CERT

Scale data consistently!

Byte volumes

- Always use “Megabits per Second” or “Mbits per second”
- Note that Mbps = Mbits per second; MBps = MBytes per second
- Mbps = (Bytes $*8$)/(seconds in the time bin)

Packet counts

- Always count packets per second

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

Try It #20!

Our task for the rest of this section will be to get a better understanding of high-port to high-port TCP traffic. Before we begin graphing, we need the data.

- Extract outbound TCP flows,
- for 12/15/2004,
- where both ports are greater than 1024,
- with both SYN and ACK flags set,
- with 5 or more packets,
- and save the output as "`highport.rw`"

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

Getting started with Excel

Excel works on the client

- Data must be loaded into a client-side spreadsheet
- Transfer bulk data with scp
- Transfer small data sets with cut and paste

Once the data's on the client, we can graph

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

Formatting data for Excel

Column Delimiters

- Excel likes comma separators, csv files
- Excel can turn pipe-delimited data into columns
  - Menu — Data — Text to Columns

Date/Time Data

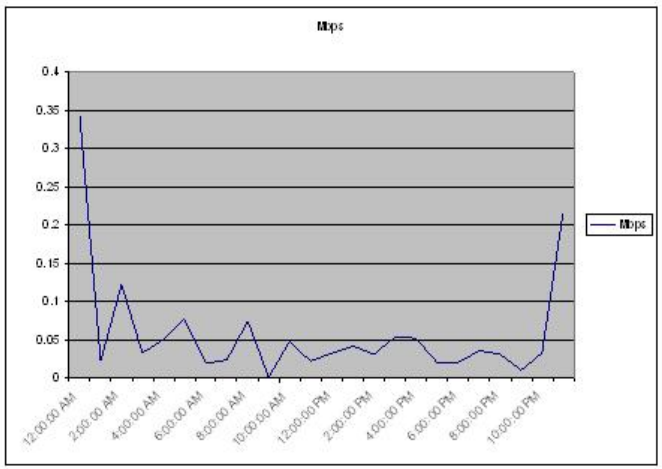
- Excel likes “legacy-timestamps”

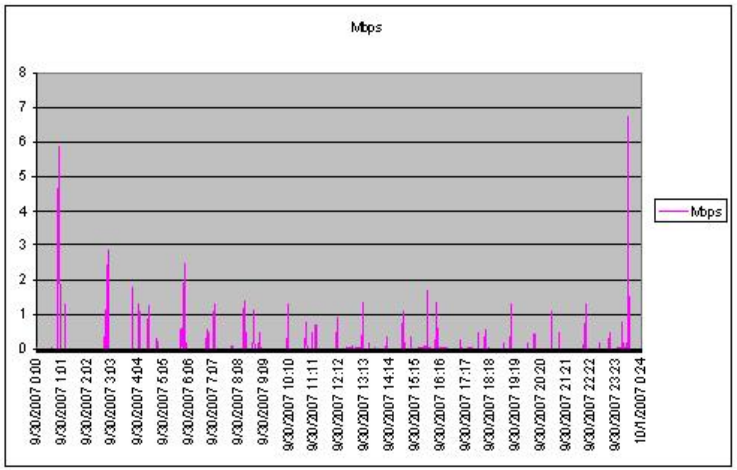
Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

Try It #21!

Our first Excel visualization shows the volume of high-port to high-port traffic over time.

- Using rwcount, create and plot volume of highport.rw data for 1-hour and 1-minute bins.





Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

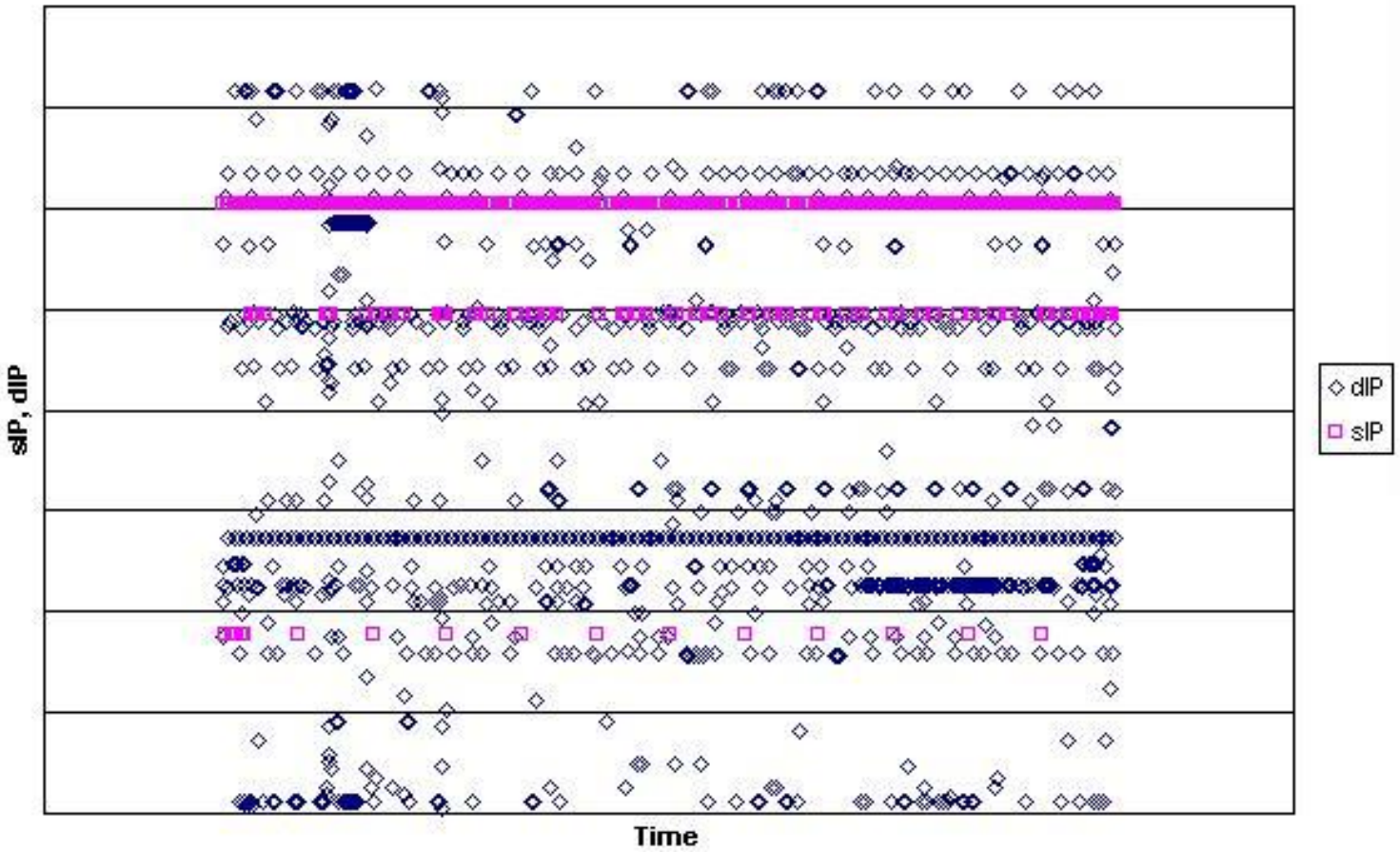
Try It #22!

Our second excel visualization uses scatterplots to try and better understand the data

- Use rwcut to dump out the raw flow records with numeric IP, and import them into Excel
- Create one or more scatter plots using two-variable combinations such as
  - source port vs. dest port
  - source port vs. flow size and dest port vs. flow size
  - source port vs. time and dest port vs. time
  - source IP vs. dest IP
  - source IP and dest IP vs. time

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

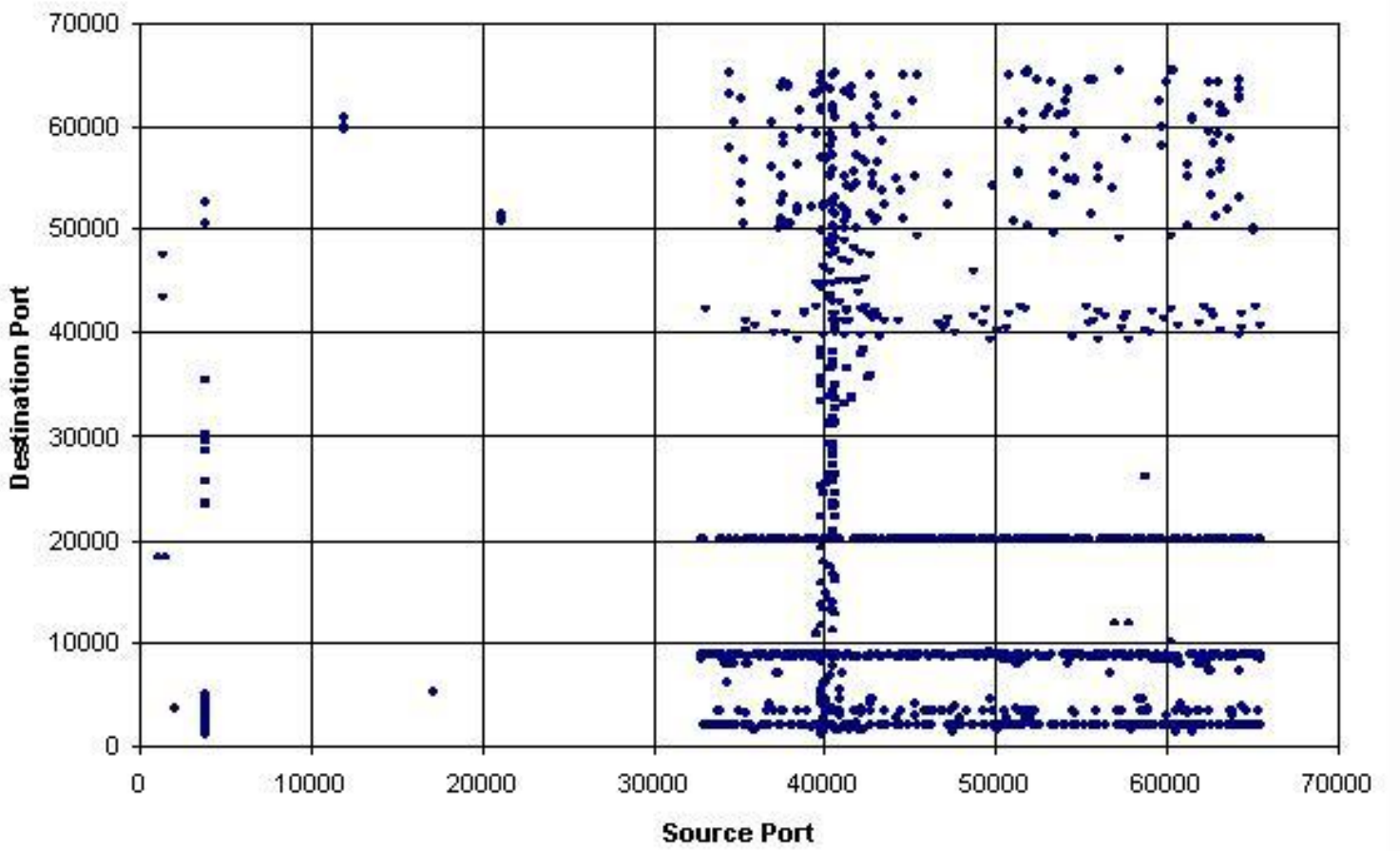
Source IP and Destination IP over time(Existence Plot)



Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

CERT

Source Port vs Destination Port



Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

CERT

Why not stick with Excel?

Excel has some concrete limitations:

- Maximum of 65k rows
- Can't easily automate
- Difficult to regularly update data for a template plot

Gnuplot makes a great complementary tool

- Fully scriptable
- Runs on the server
- Only limits are file sizes
- No GUI

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

What is gnuplot

Command-line interactive data plotting utility

- Originally intended for scientists to visualize mathematical functions
- Now supports many non-interactive uses

Multiple output formats

- Interactive (X-windows)
- Direct to printers
- Many file types

We'll use it to plot text data into a postscript file

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

A minimal gnuplot script

```
#! /usr/bin/gnuplot
set terminal postscript enhanced color solid
set output "temp.ps"
set title "Hello, World"
plot sin(x)
```

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

Try It #23!

Use this script to reproduce high port traffic as in the previous example:

```
#! /usr/bin/gnuplot
set terminal postscript enhanced color solid
set output "hourly.ps"
set title "Hourly High Port Trends"
set datafile separator ","
set timefmt '%m/%d/%Y %H:%M:%S'
set xdata time
mbps(x)=x*8/3600/1000000
plot 'hourly.csv' using 1:(mbps($3)) with lines \
  title "Mbps"
```


Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

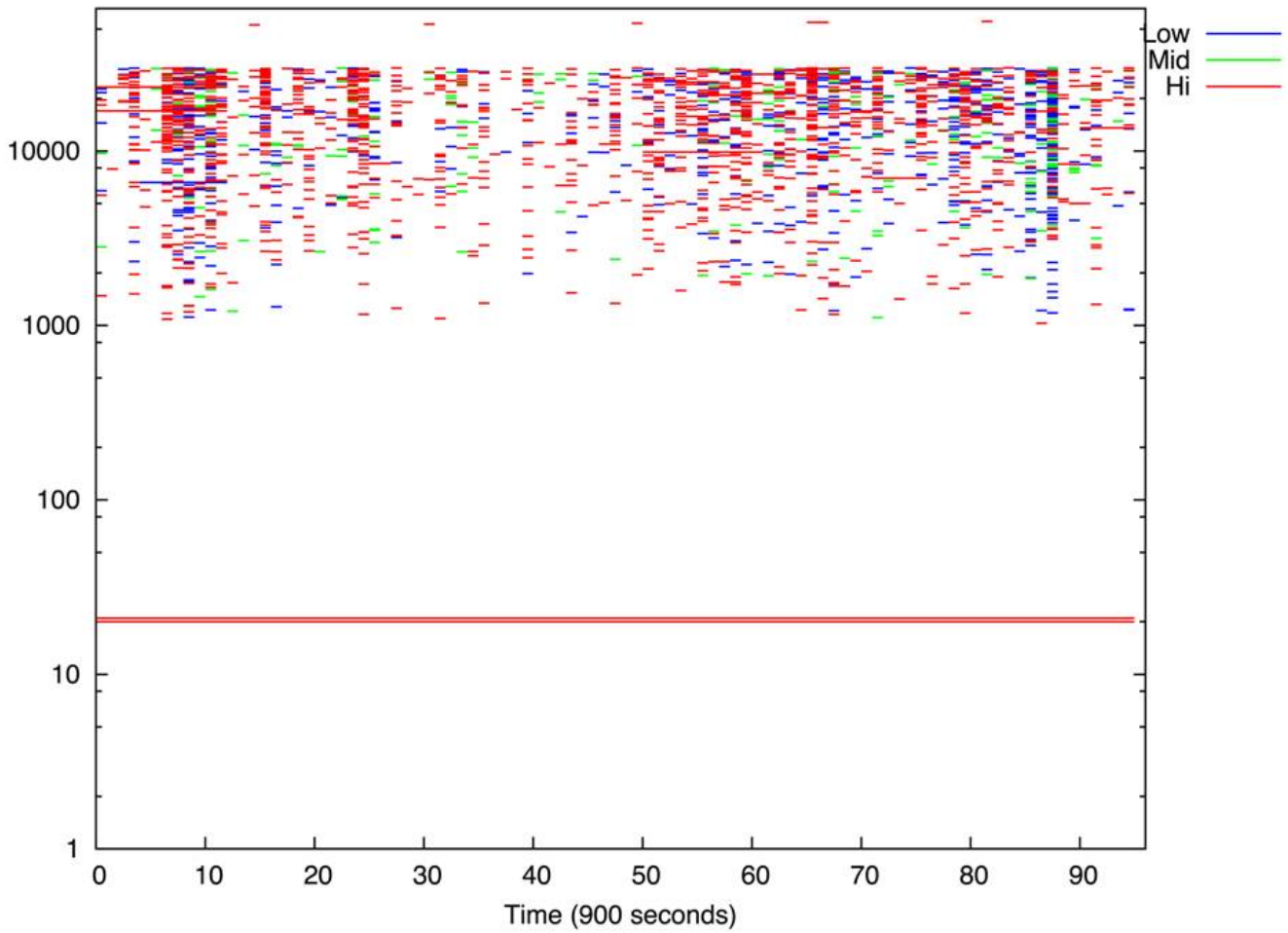
Try It #24!

Use gnuplot to reproduce the results of Excel scatterplot exercise above. Start with the script from the previous example.

- Once you've generated some output, try plotting other column variations.

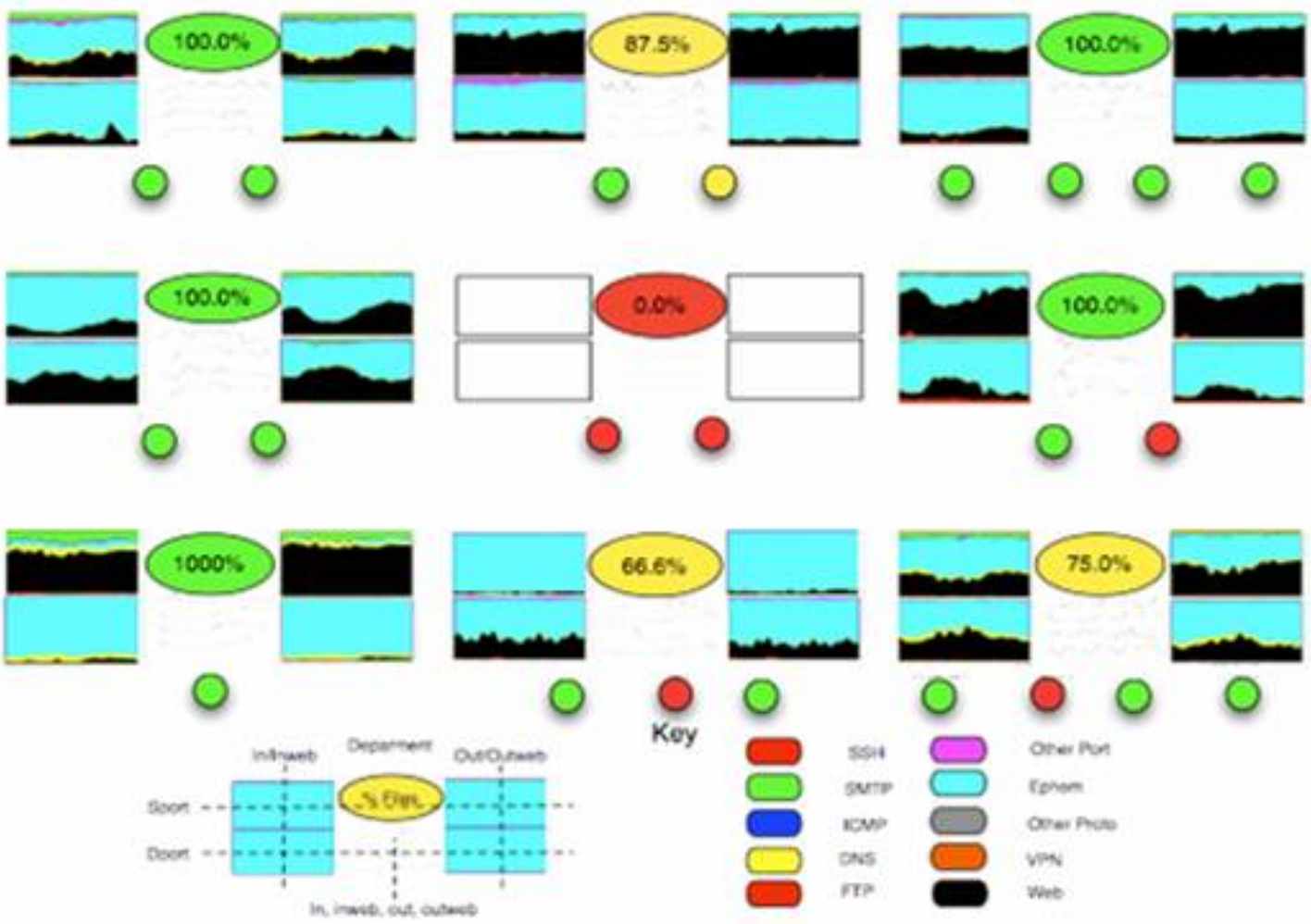
Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

Advanced Graphs: Existence Plots



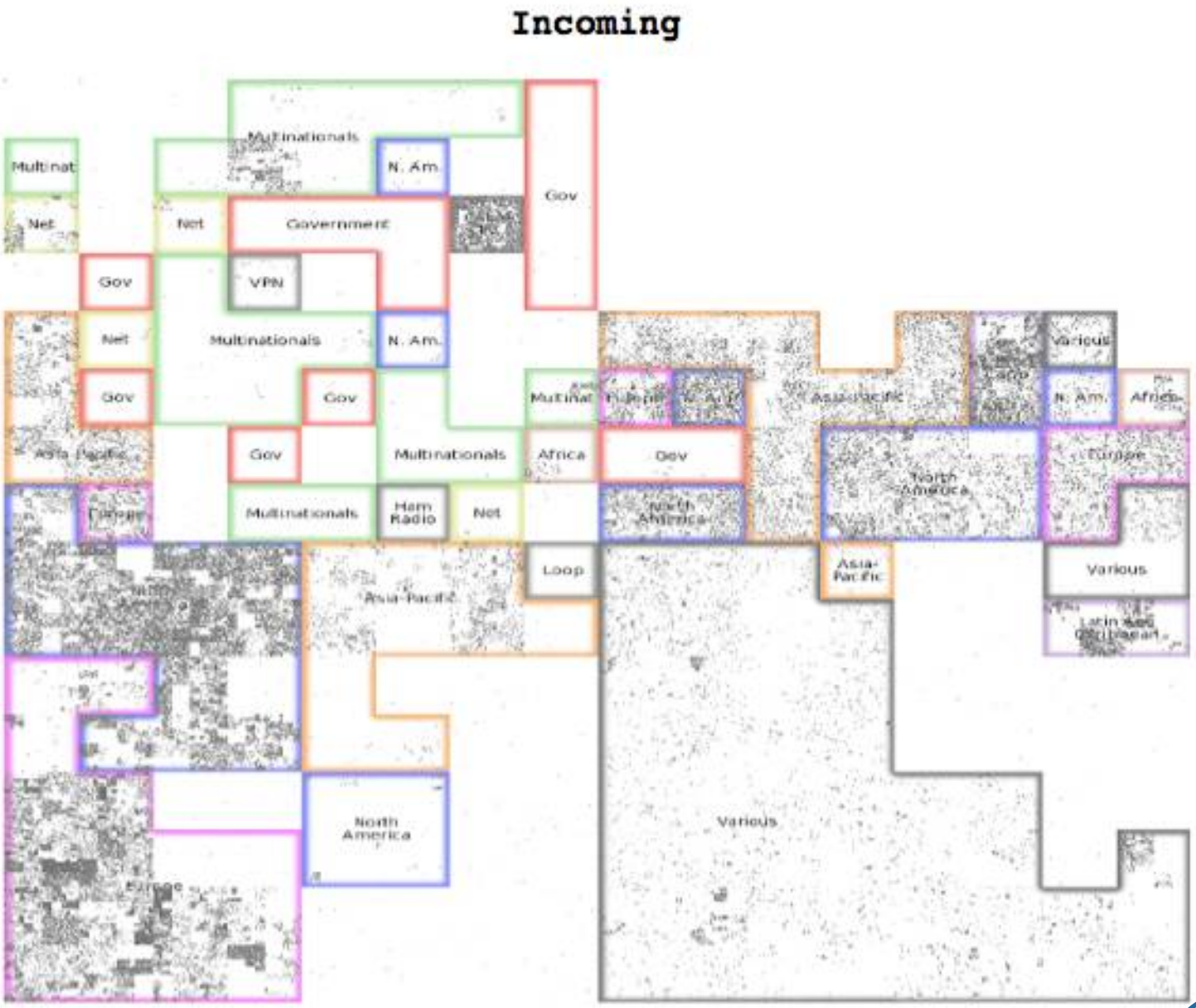
Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

Advanced Graphs: Information Density



Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

Advanced Graphs: Abstract Art



Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

What have we done so far?

Visualization

- Basic graphs
- The Excel Cookbook
- The Gnuplot Cookbook
- Advanced graphs

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

Using the SiLK Tool Suite

Many tools not touched on in this training:

- Packet content: rwptoflow, rwpmatch
- Text to Flow: rwtuc
- Address aggregation: rwnetmask
- Newer and broader tools still under development

Explore the tool set for awareness you need to build of your network

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

Where now?

Explore Your Data

- Rich data source

Using SiLK for Situational Awareness

- Top-N lists (rwfilter + rwstats)
- Odd behavior frequencies (rwfilter+rwcount)
- Drill-down (rwfilter +rw*)
- Build off of examples presented

Script analyses of interest

- Shell script

Create visualization templates

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

What's coming

Tools manipulating higher-level representations:

- Clusters
- Graphs
- Patterns

Expanded data collection Collaboration with other flow analysis efforts

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

SiLK Support

- `mailto:silk-help@cert.org` for bug reports and general inquiries
- FloCon `http://www.cert.org/flocon/` conference on network security analyses with flow data
- `http://tools.netsa.cert.org/` – open source release of SiLK, with documentation
- `http://www.cert.org/netsa` – papers and presentations by CERT NetSA Group

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

Communications Resources

- `silk-help@cert.org`, SiLK tool suite help
- `flocommunity@cert.org`, discussions and announcements for network flow analysts
- FloCon (http://www.cert.org/flocon), Flow analysis conference

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

Other References

- The analysts handbook! Great source of how-to info.
- The SiLK Reference Guide.
- Toolname help (or man toolname)
- http://tools.netsa.cert.org

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

Other Useful Hints

- Use Toolname –help for help and/or analysts handbook (or me)
- Do as much analysis in the binary as possible: it's faster.
- Use sets (or tuples) whenever possible vice going thru the same data multiple times.
- When pulling together by-address, by-protocol, or by-port counts for later processing, use bags instead of rwuniq

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community

Questions?

Introduction
About
Flow
Unix
Beginning Analysis
Basic SiLK Tools
rwfilter
Printing and Sorting Tools
Counting Tools
Other Tools
Advanced
Sets
Bags
Prefix Maps
Unix Scripting
Visualization
Basic Graphs
Excel
Gnuplot
Advanced Graphs
Conclusion
The Community