Tracking Global Threats with the Internet Motion Sensor

Michael Bailey & Evan Cooke

University of Michigan

Timothy Battles

AT&T

Danny McPherson

Arbor Networks

NANOG 32

September 7th, 2004

Introduction

- Networks increasingly subjected to threats that impact the reliability and availability of critical infrastructure.
  - Distributed Denial of Service (DDoS) - *SCO*
  - Fast Moving Worms - *CodeRed, Nimda, Sapphire, Blaster, Witty, Sasser, blah, blah blah …*
- The goal of IMS is to measure and characterize these threats in real-time

Monitoring Approach

- Use routing infrastructure to collect malicious traffic and send it to a box to be analyzed.
- How do you figure out what to pull off? …attacked customer traffic, outbound to Bogon space, inbound to unused addresses …
- Unused space, or ‘blackhole’ is appealing because no production hosts in the unused space. Traffic must be the result of worms, backscatter traffic, scans, or mis-configuration.

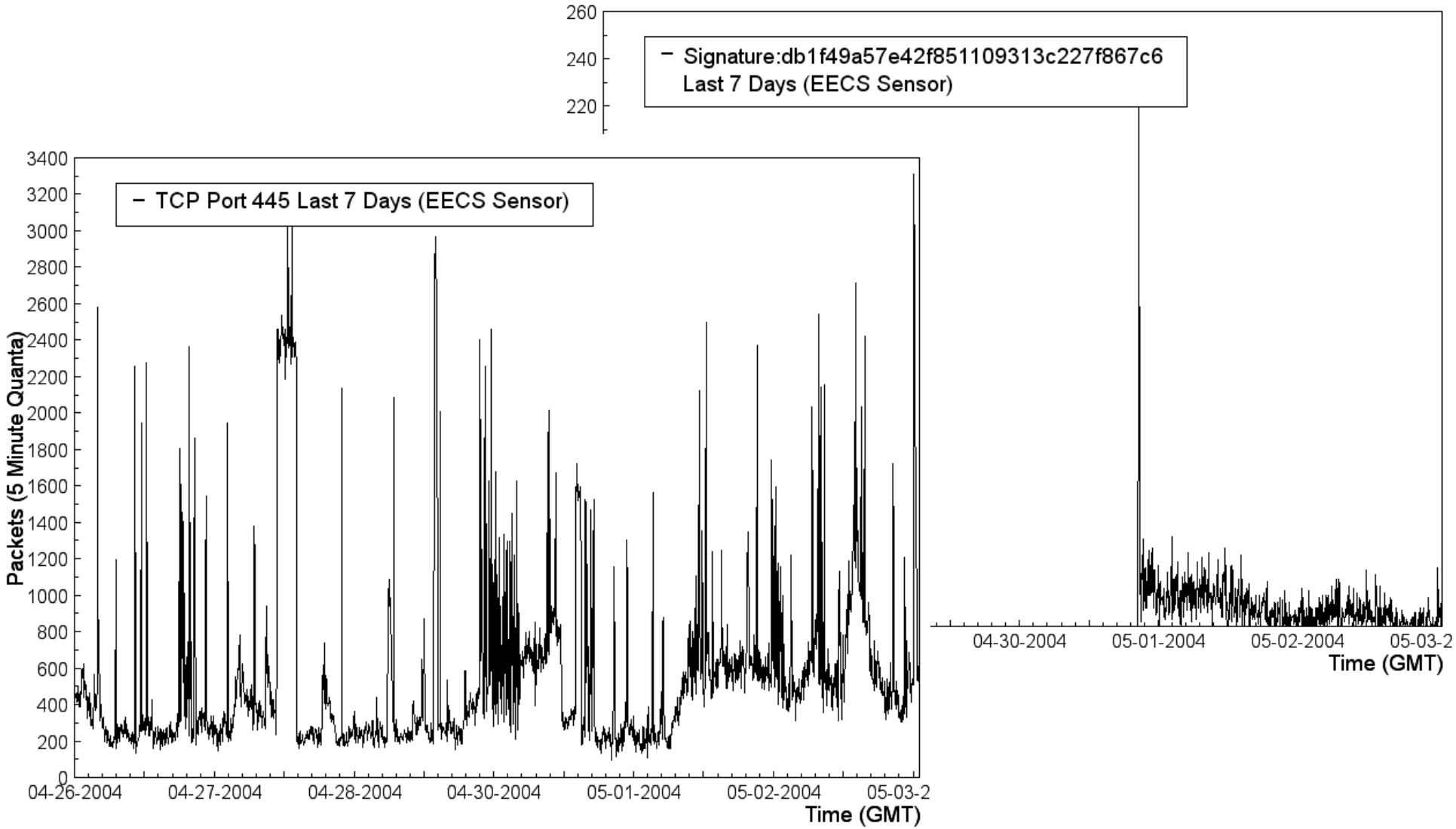
Related Work

- CAIDA
  - http://www.caida.org/analysis/security/
- Arbor
  - http://www.arbor.net/research_researchers.php?
- Team Cymru
  - http://www.cymru.com/Documents/darknet-project.html
- IUCC/IDC Internet Telescope
  - http://noc.ilan.net.il/research/telescope/
- iSink
  - http://www.cs.wisc.edu/~pb/
- Related BGP off ramping techniques (CenterTrack, SinkHoles)

Our Design Goals

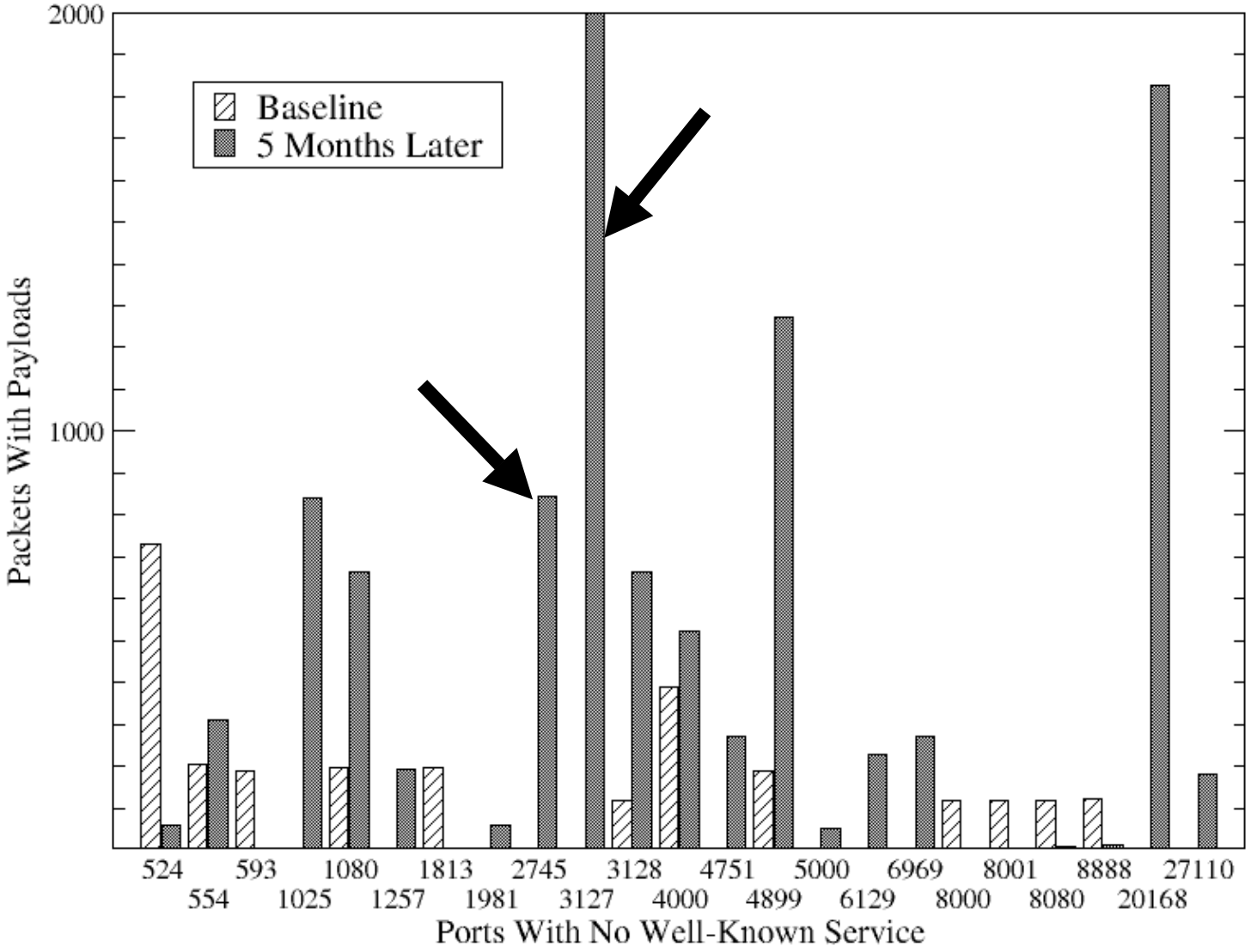
1. Differentiate threats
2. Capture new services
3. Broad Coverage
4. Availability

Packets per 5 minutes of TCP/445 over 7 days at 1 /24

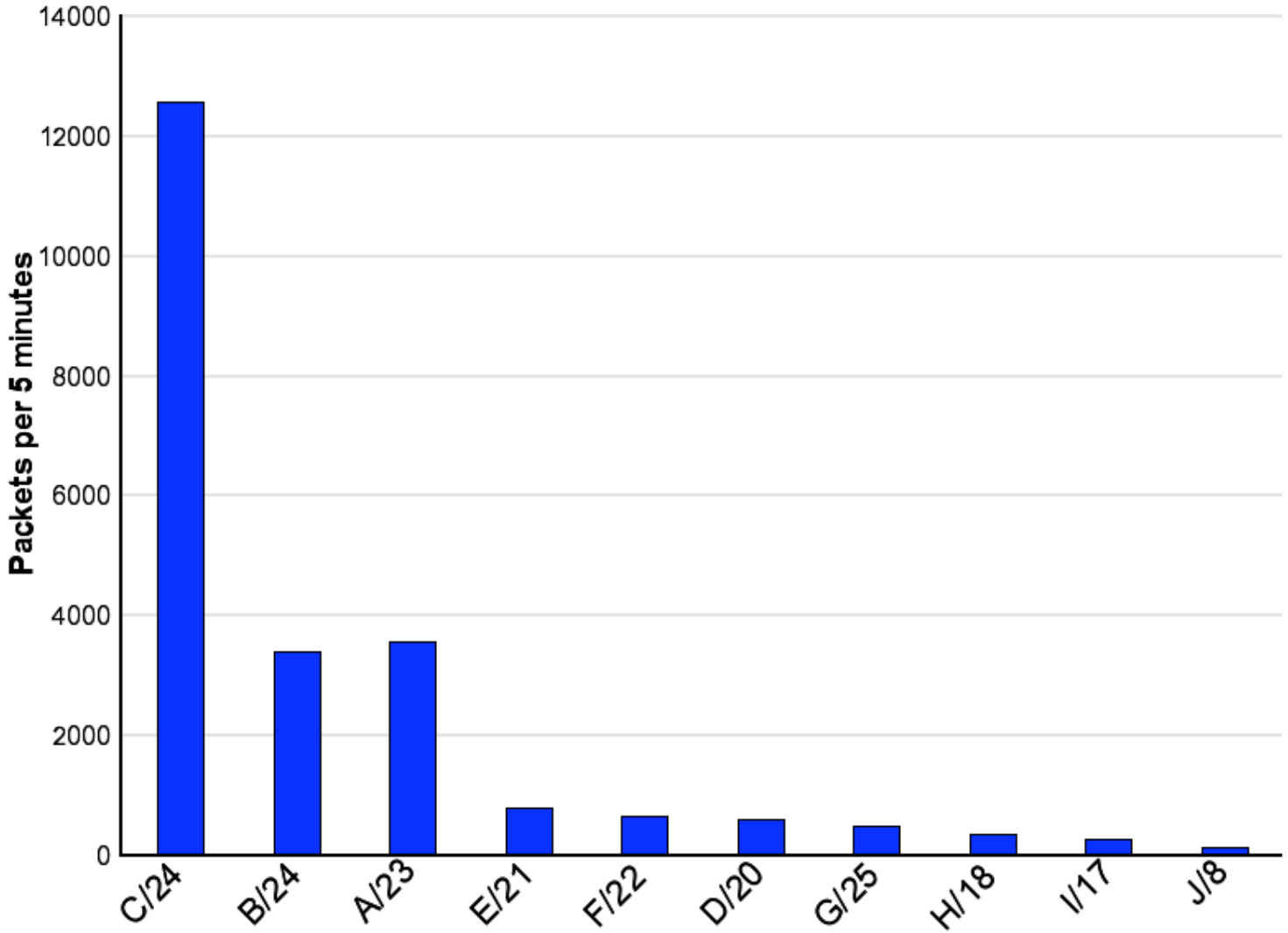


Ports with the biggest changes over a 5 month timeframe

- Significant changes are routine
- Some are more interesting than others.
- Such as the 2745 and 3127, Bagle and MyDoom backdoors

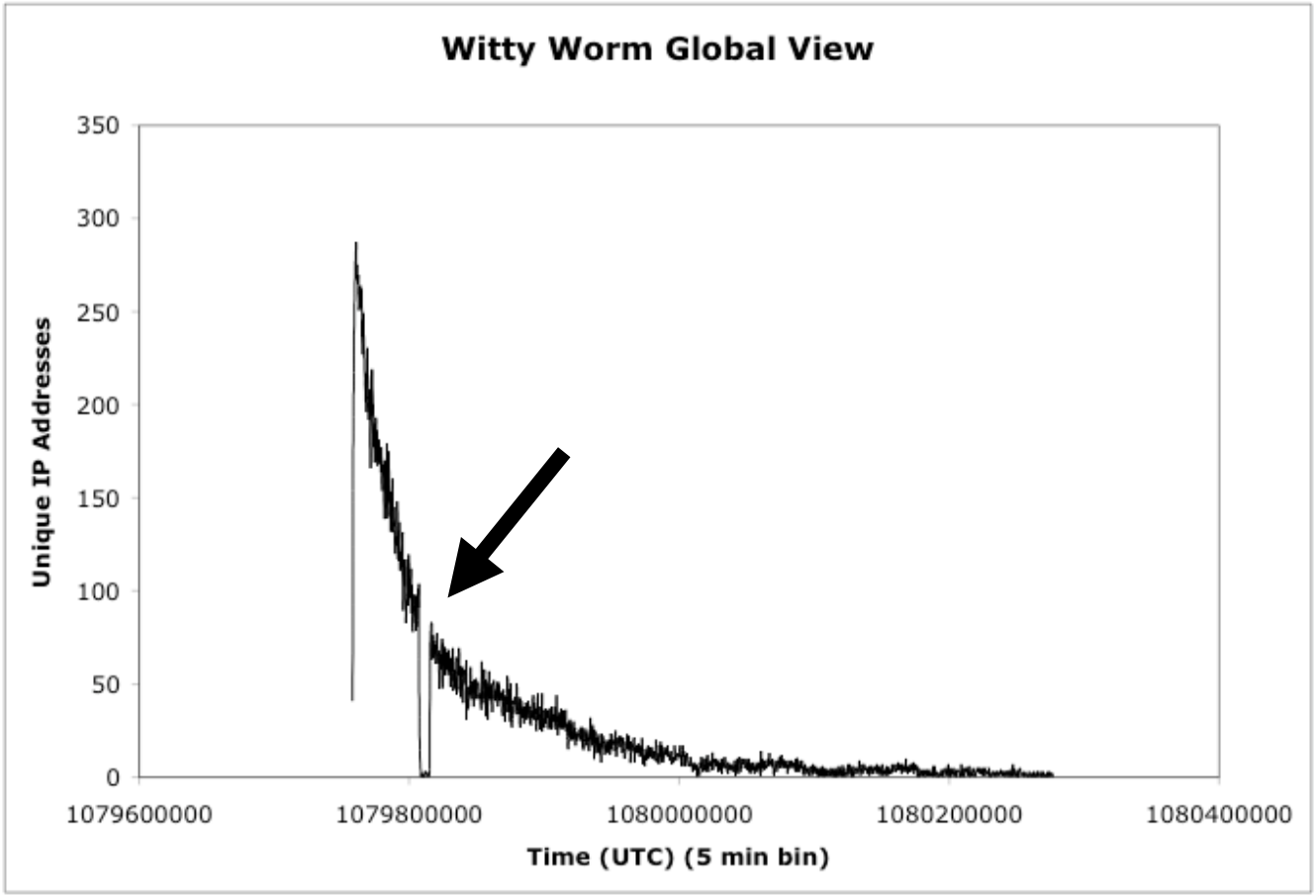


Average packets per 5 Minutes at 10 sensors over 1 week



Unique sources observed at 3 sensors in 5 minute bins

- In 3 different /16, 2 different /8s
- But same upstream provider

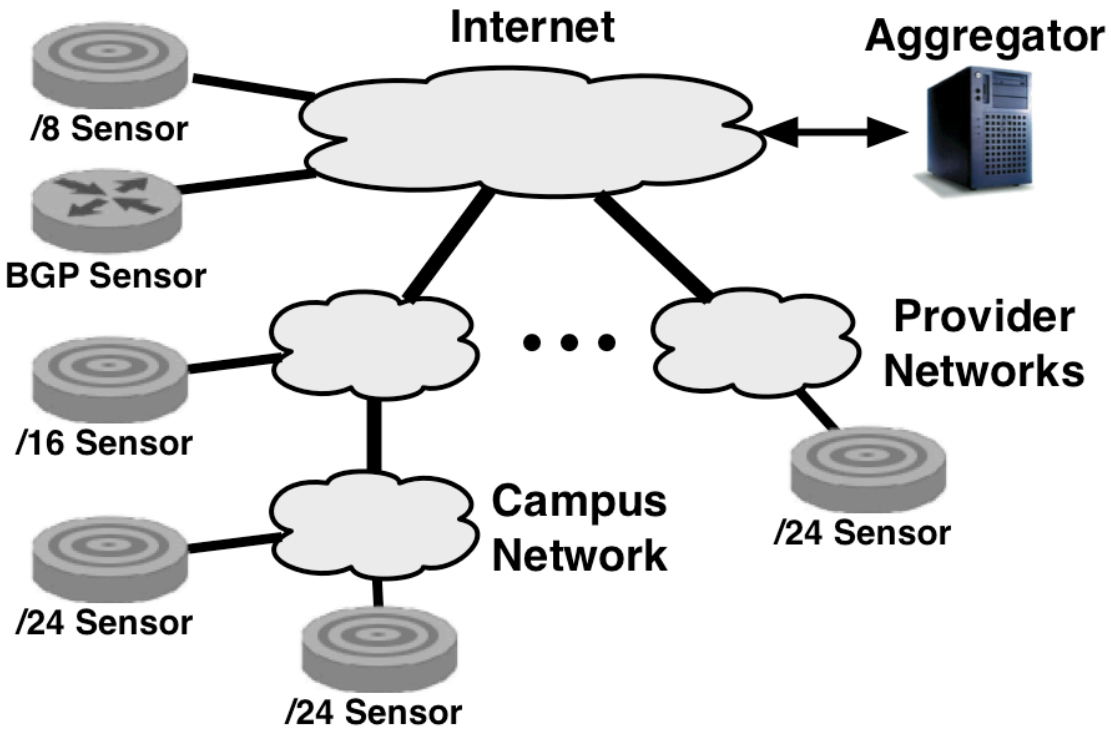


Architecture

- Distributed sensors with address and topology diversity
- Differentiate services by replying to TCP SYN with TCP SYN ACK. Elicits first part of application data.
- Payloads reconstructed and md5 checksum to build signatures.
- Respond to all services

IMS Deployment

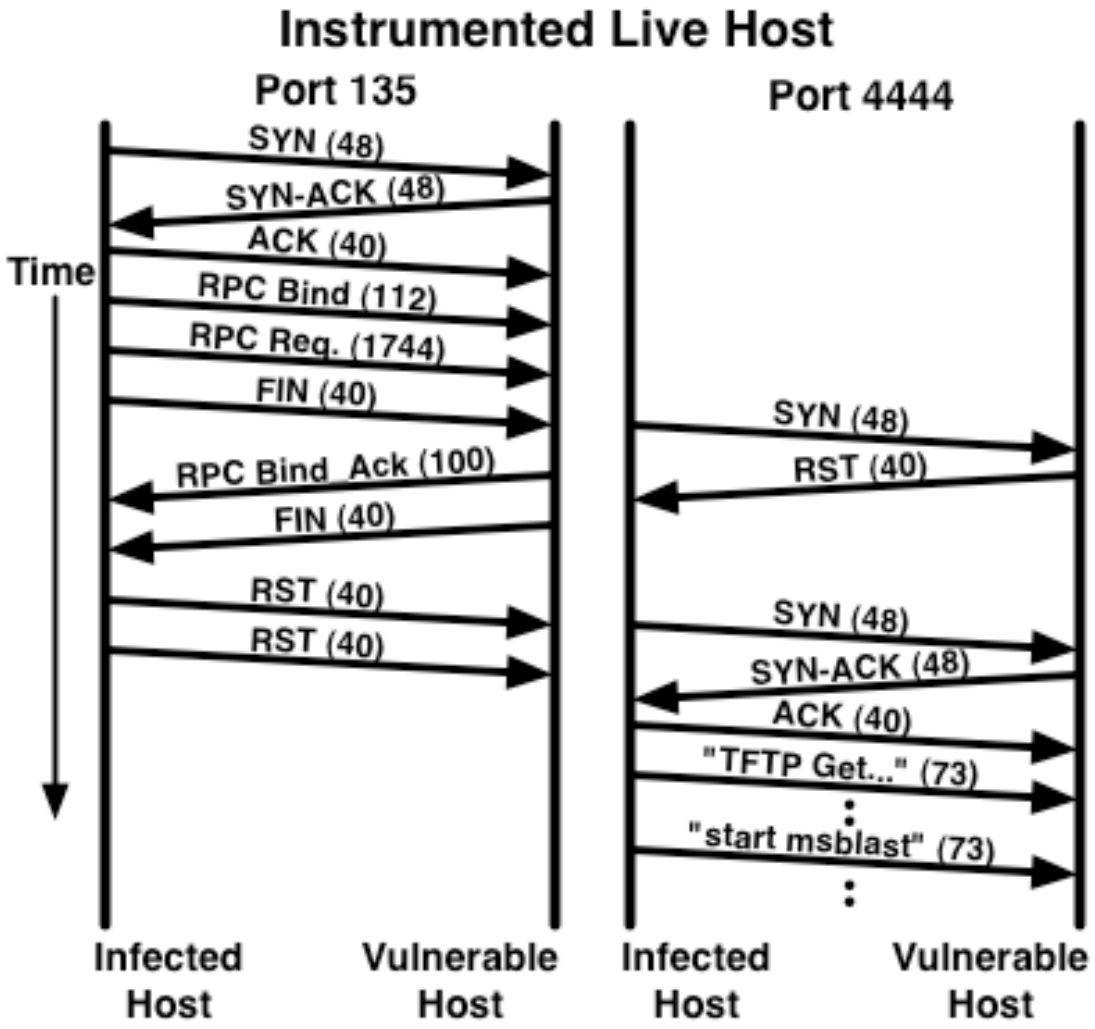
- 28 monitored blocks at 18 networks including
  - Academic networks
  - Service providers
  - Large businesses
- Consists of
  - One /8
  - Three /16
  - Five /16-/24
  - Nineteen /24s
- 3 Continents (5 soon!)



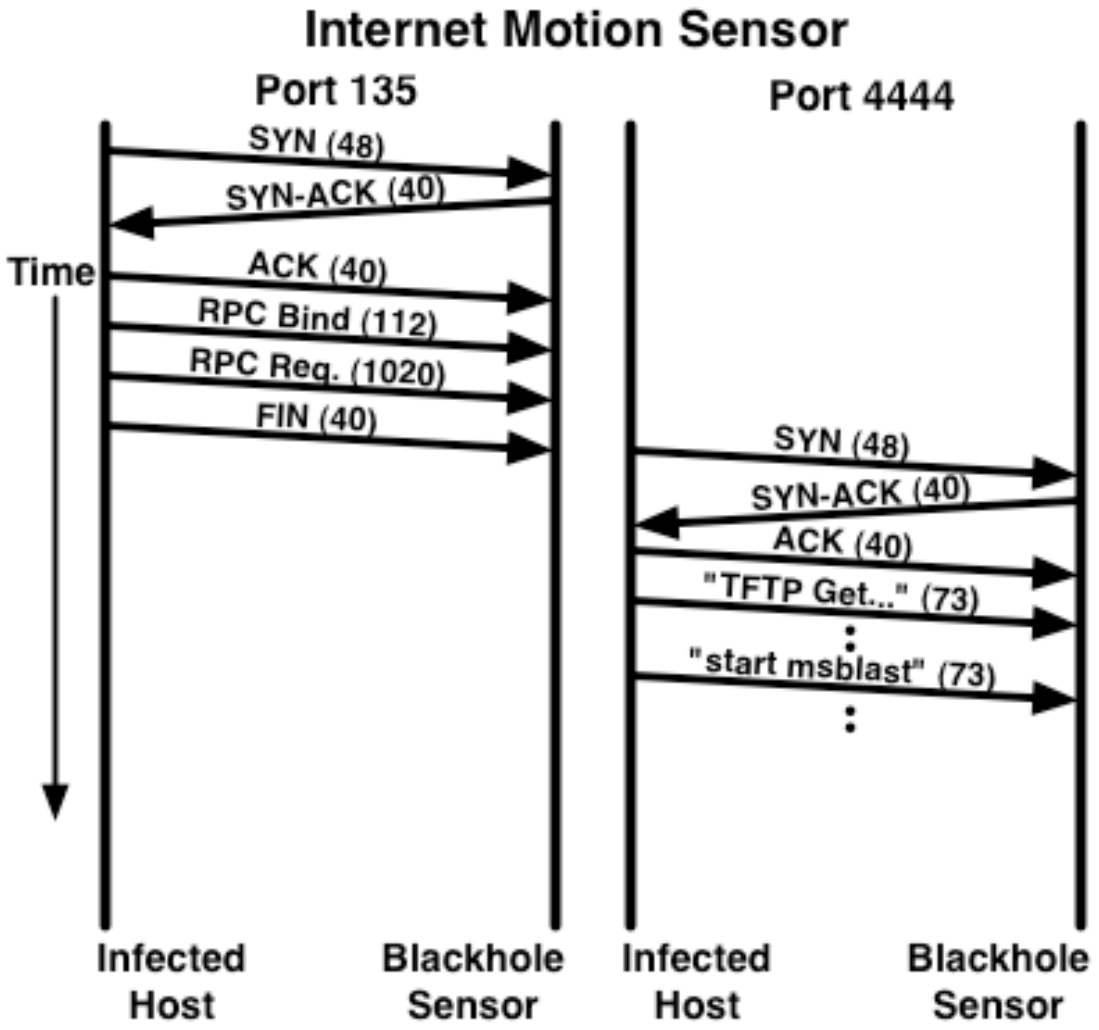
Differentiate Services

- UDP/ICMP are OK passive because we get information in the first packet.
- However, TCP is a problem because no information until handshake
- Solution: Use a lightweight active responder to get the first data packet
- Very simple:
  - Get SYN, Respond with SYN-ACK (no state)

The Blaster Worm - Live Host



The Blaster Worm - IMS

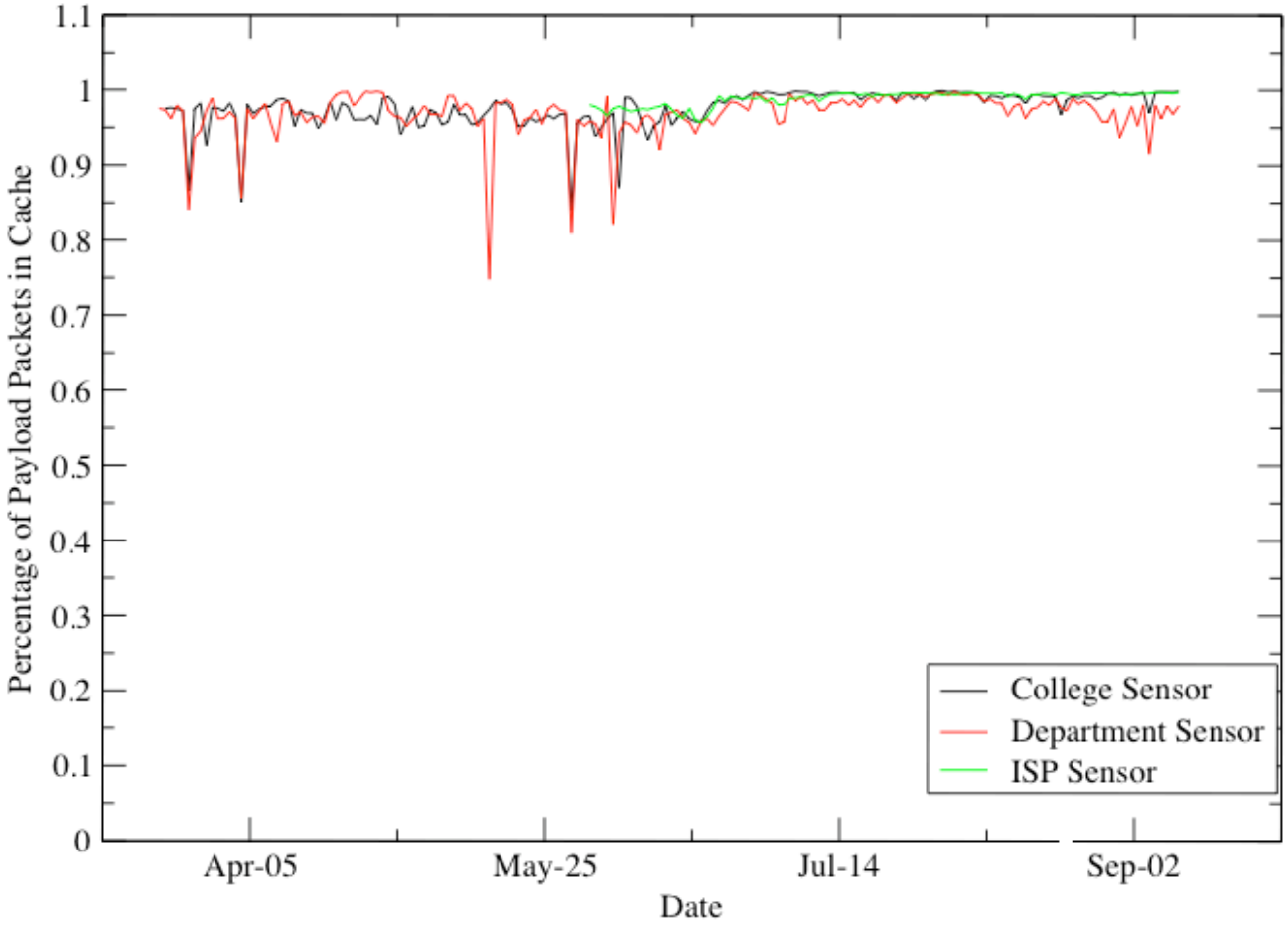


Payload Caching

- Active responder produces lots payload data
- Solution: only store payloads if they are ‘new’
- Implementation: take MD5 hash of payloads and only store payloads which have a unique hash

% of Payload Cache Hits over 5 Months at 3 sensors

- ~95% signature cache hit-rate
- Most payloads have been seen before

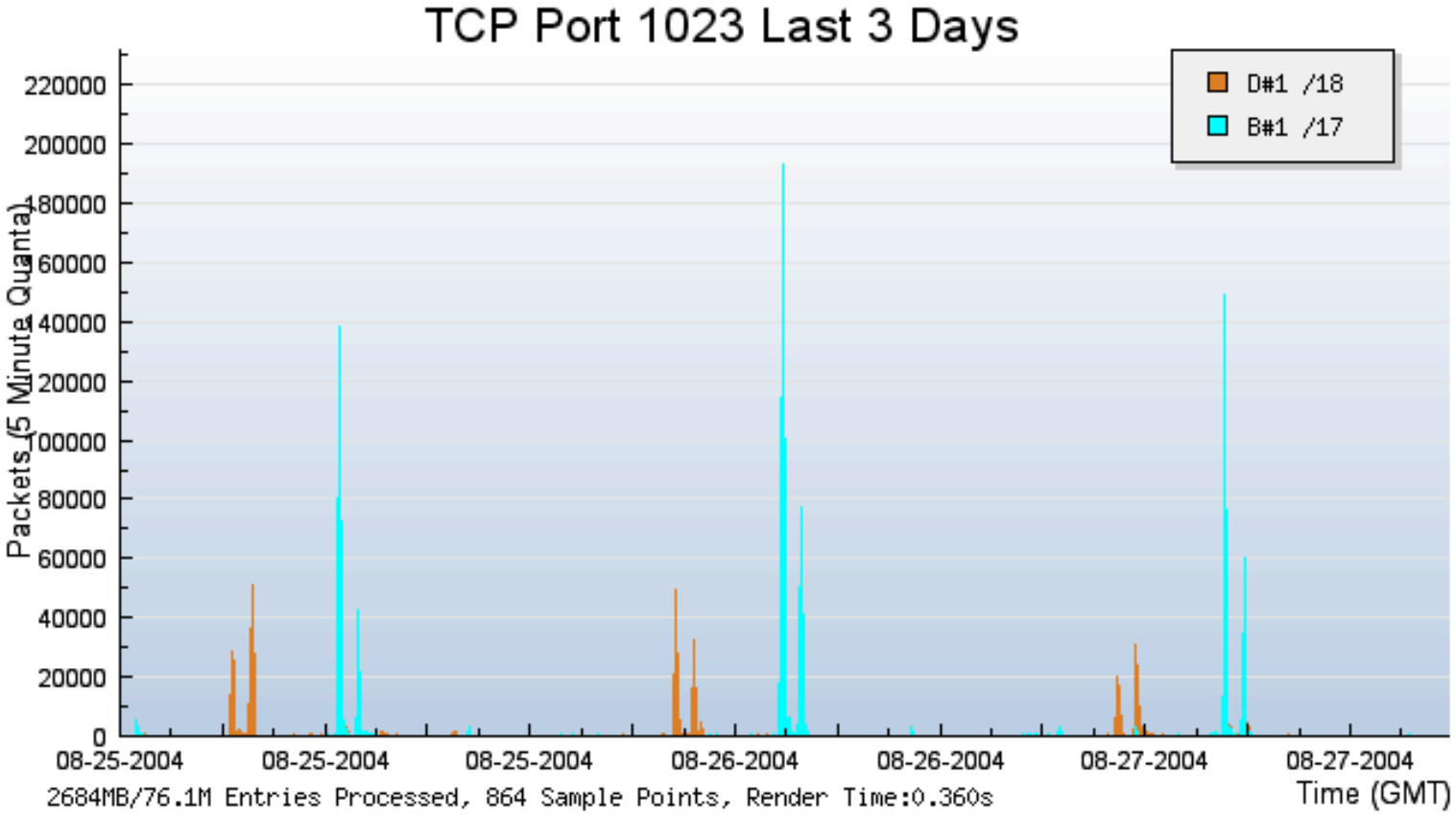


Recent Observations

Worms

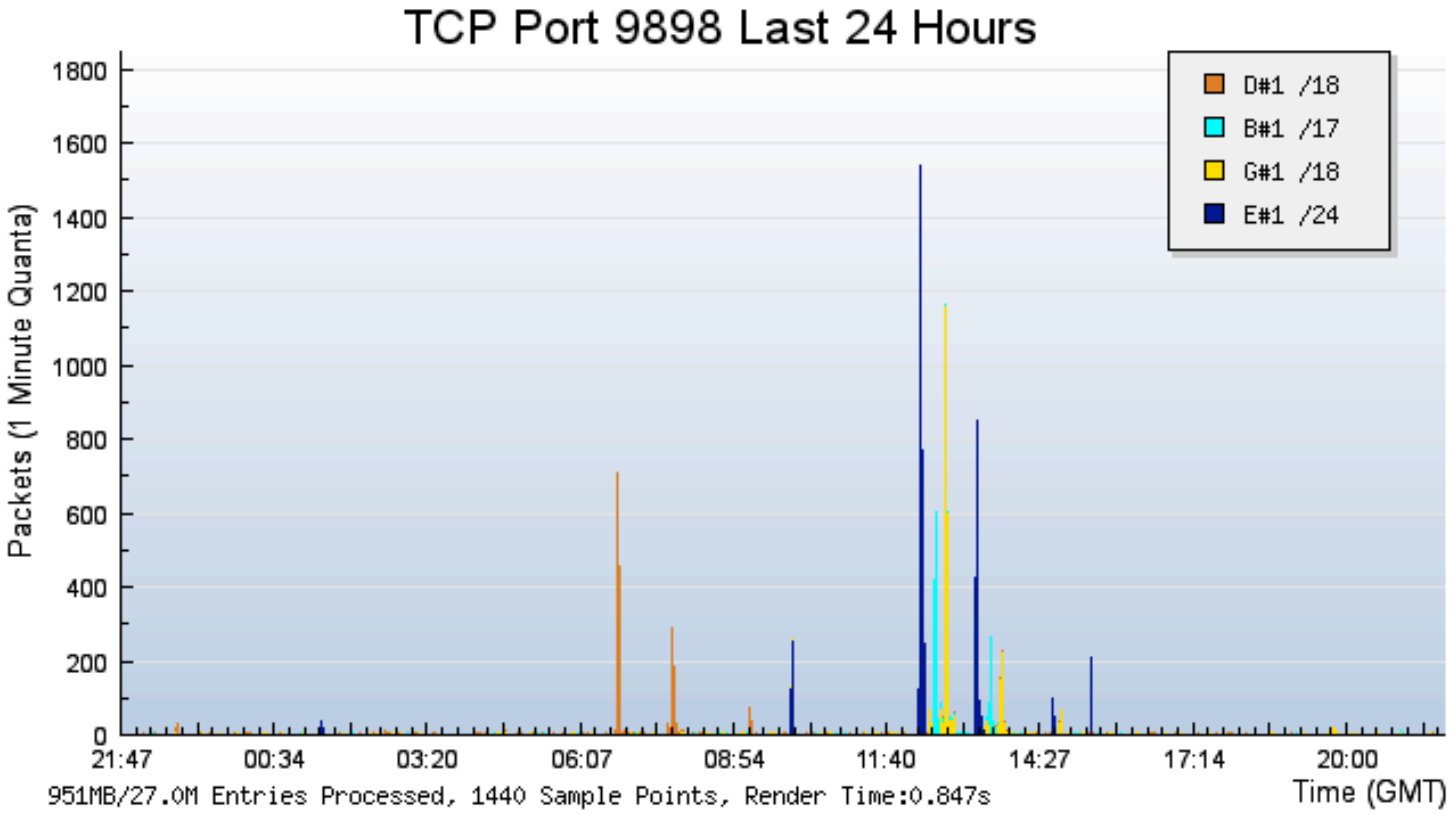
Worm	Sasser	Sasser.e	Dabber.a
Vulnerability	LSASS (MS04-011)	LSASS (MS04-011)	Sasser-FTP
Population	Windows XP Windows 2K	Windows XP Windows 2K	Sasser infected hosts
Scan Port	TCP/445	TCP/445	TCP/5554
Backdoor Port	TCP/5554	TCP/1023	TCP/9898
Release	May	May	May
Who Cares?	First LSASS	Changes backdoor port	Vulnerability hits bugs in a worm backdoor

Packets per 5 minutes on a /17 and a /18 over 3 days for TCP/1023



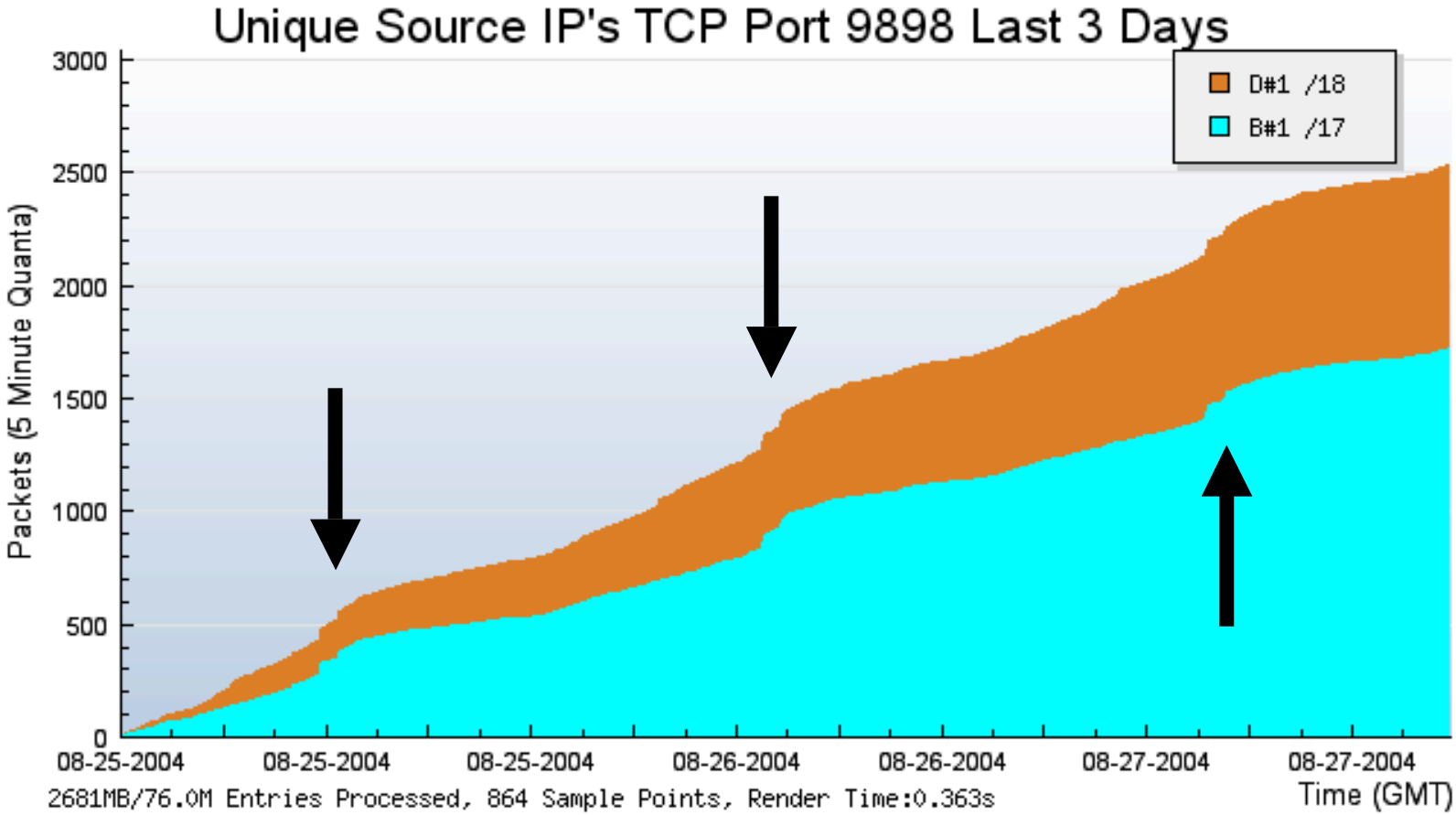
- Large, short lived spikes
- Same shaped graph across (1023, 5554, 9898)
- Nearly all sources in China and Korea

Packets per 1 minute on 4 sensors over 3 days for TCP/9898, normalized by /24



- D -> E -> B -> G (Ordered by /8)
  - 7.8 Hrs ->.4 Hrs ->.25 Hrs
  - 51 /8s -> 2.5 /8s -> 1.5 /8s
- ~6 /8's an Hour

Cumulative Unique Sources on a /17 and a /18 over 3 days for TCP/9898



- small number of hosts are involved (~100)
- the size of the bumps is similar each time
- Hosts dwarfed by background noise

Overlap in Scanning IPs

		/17			/18		
		1023	5554	9898	1023	5554	9898
/17	1023	173					
	5554	142	470				
	9898	168	424	536			
/18	1023	0	0	0	99		
	5554		10	9	94	231	
	9898			14	99	224	280

Signature Analysis

- No signatures captured on 9898/tcp
- 2 unique signatures on port 5554/tcp
- Same 2 unique signatures on port 1023/tcp
- Here are the sigs:
  - e5502ddb7ce4a7ff2176e6455732601c

    00000000 55 53 45 52 20 78 0a |USER x.|

    00000007
  - F623e75af30e62bbd73d6df5b50bb7b5

    00000000 44 |D| 00000001

Getting Involved

- IMS is a public research project funded by various government agencies and corporations
  - http://ims.eecs.umich.edu/people.html
- Got space? We'll send you a box.
  - mailto:ims@umich.edu
- Or you can BETA the software only distribution
- You'll get aggregate views across all other participants, and detailed views on your sensor.

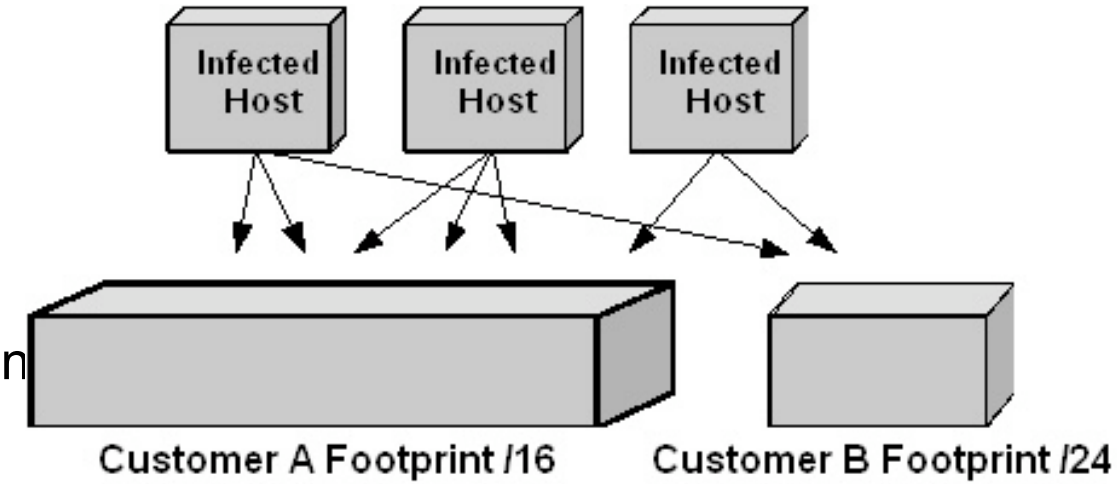
Making Dark Address Space work for you.

Timothy A Battles

AT&T IP Security

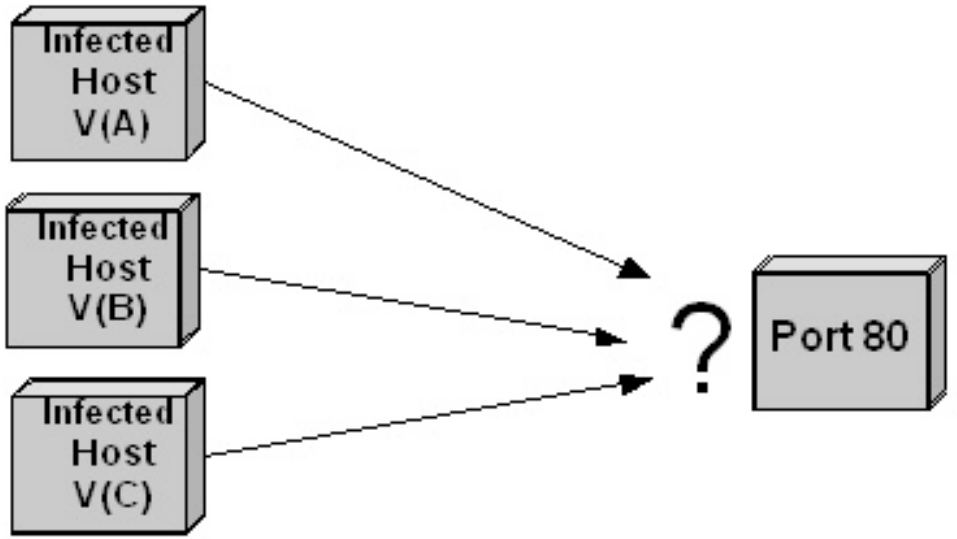
What's the Importance?

- Can provide lists of infected clients to abuse groups.
- Provides baselines based on footprint into measurements for customers receiving white noise.
- Can provide early detection of worms/viruses.



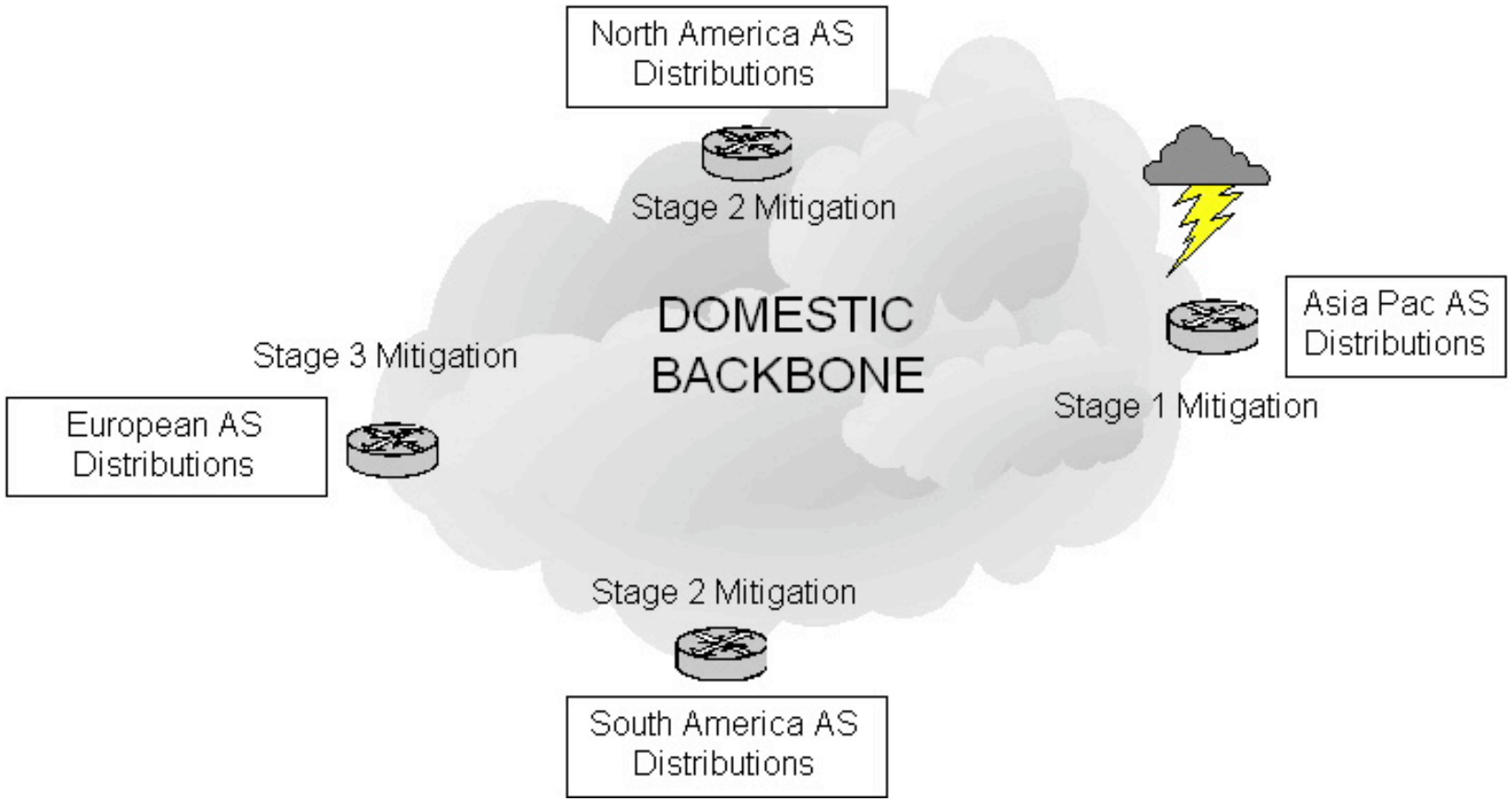
Why do we need the payload?

- Enables detection of new variants operating on known ports.
- Can be used to create new IDS signatures.
- Enables variant tracking independent of associated port.



The important of source address distributions.

- Allows for prioritization



Why do we need it real-time?

- Do we really need to answer this question?

Bringing it all together