# IAB Unwanted Internet Traffic Workshop

## Session 7 - What's in the pipeline, and what should be in the pipeline?

Danny McPherson danny@arbor.net

March 10, 2006

# Keep it simple…

- Central repository and subsequent functions & benefits
- Flow data - detect and remove bot substrate, many other functions as well
- DarkNets

- And, of course, “clue sharing” and associated implications (more “raising the bar” tutorials, etc..). No excuses for operators not employing X, Y & Z.

# Two biggest problems…

- **Route hijacking**
  - Lack of inter-domain policy application
- **Source address validation**
  - Spoofing
  - Reflection & amplification
  - Traceback

- How to begin addressing
  - [Central] up to date repository (RIR/IRR) w/AAAish functions, w/IRRD toolset
  - Vendor support for extremely large (BGP prefix) control and data path filters (need at customer & peering edge)
  - Cost|Risk|Benefit Analysis: Lots of bang for the buck$

# Flow Data

- Used to convey Network & Transport Layer attributes of network transactions transiting or terminating on a network device
  - NetFlow
  - sFlow
  - JFlow
  - IPFIX
- Lots of open source and commercial tools available

- Application
  - DDOS Traceback
  - Botnet Detection
  - Traffic/Peering Analytics
  - Worm Detection
  - Spyware Detection
  - Compliance
  - Misuse
  - Etc..

# Flow-based Anomaly Detection

- Monitor flows on the network and build baselines for what normal behavior looks like:
  - Per interface
  - Per prefix/IP Address
  - Per Transport Layer protocol type, ports/ICMP types/codes
  - Build time-based buckets (e.g., 5 minutes, 30 minutes, 1 hours, 12 hours, day of week, day of month, day of year), could couple with routing or other datasets (e.g., BGP community)

# Flow-based Detection (cont)

- Once baselines are built anomalous activity can be detected
  - Pure **rate-based/statistical** (pps or bps) anomalies may be legitimate or malicious
  - Many **misuse** attacks can be immediately recognized, even without baselines (e.g., TCP SYN or RST floods)
  - **Signatures** can also be defined to identify “interesting” transactional data (e.g., ‘proto udp and port 1434 and 404 octets’ (376 payload) == slammer?)
  - Employ relational databases (perhaps with temporal behavior consideration) and detect zero-day worm, subtle misuse, backdoors, multi-phase propagation, infection, etc..
    - E.g., tcp/80 ->then tcp/9898 -> then *n* within *t* == infection *x*
  - Feeds for known bad entities, botnet controllers, etc..

# Internet Motion Sensor (IMS)

unfortunate collision of acronyms with Internet Multimedia Subsystem

http://ims.eecs.umich.edu

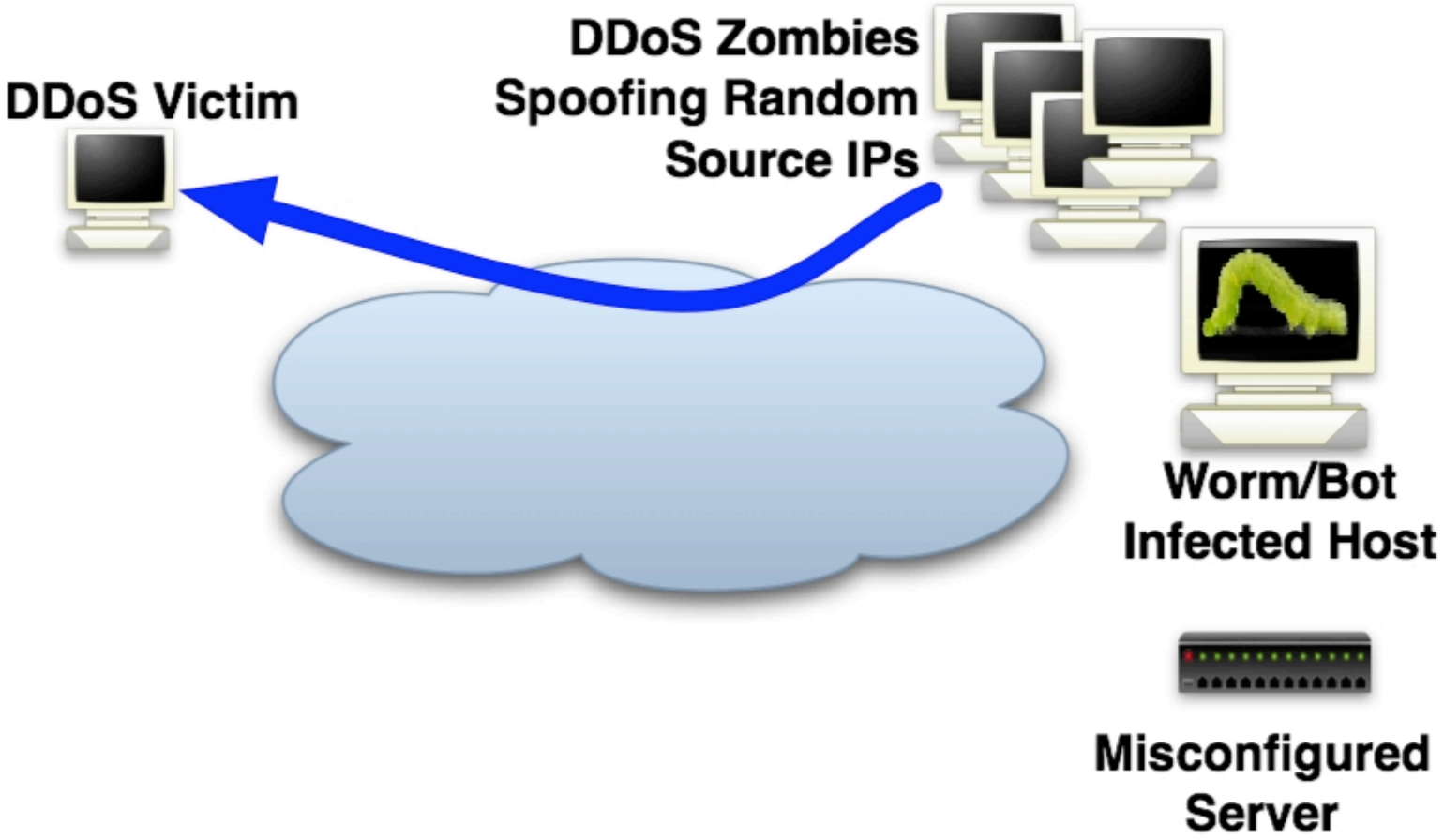
Email: ims@umich.edu

Thanks to Evan Cooke, Michael Bailey,
Farnam Jahanian, Jose Nazario & Dug Song

# About Dark IP Analytics

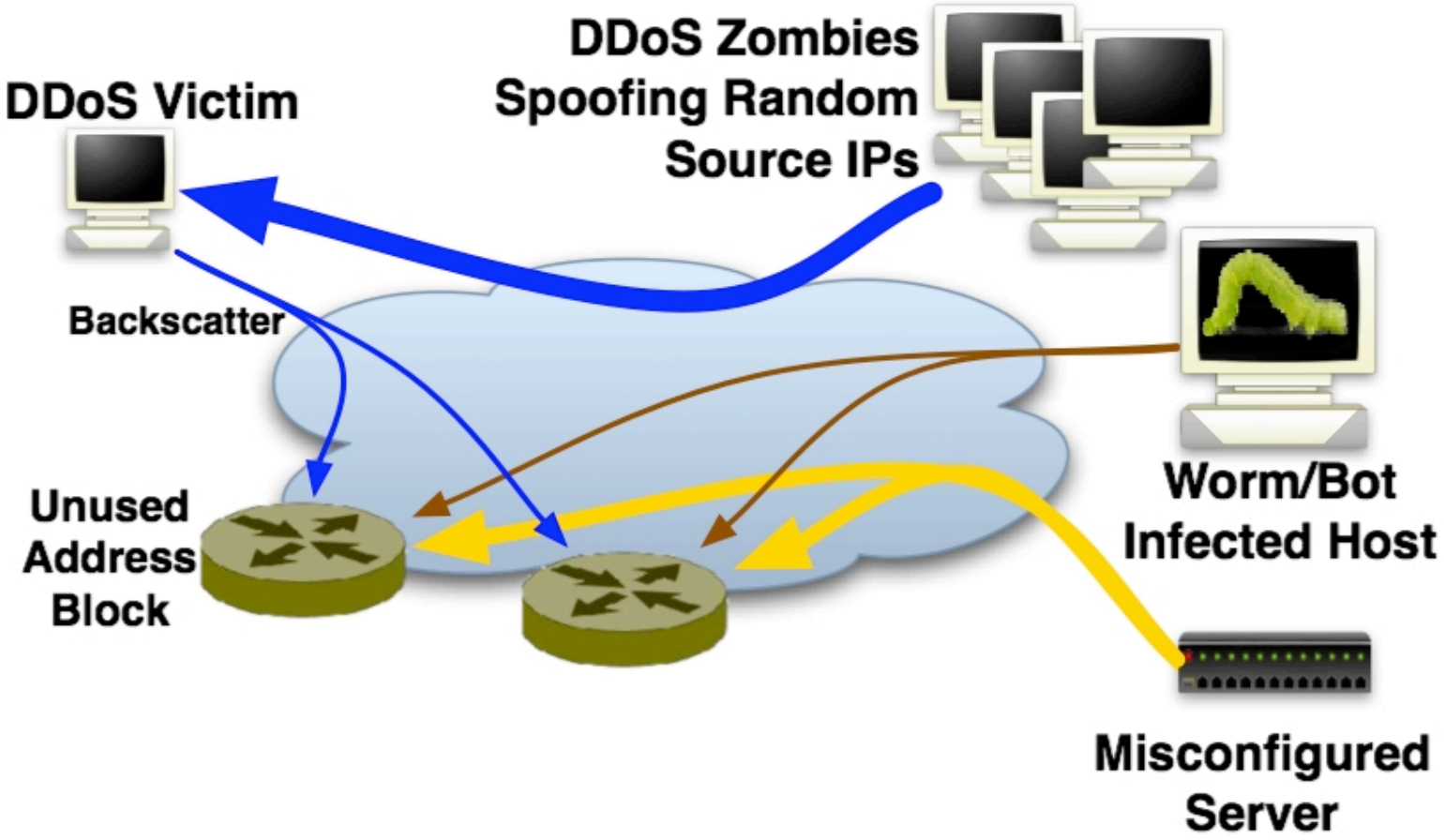
- Significant % of routed Internet address space lacks actual end hosts
  - (IANA->RIR->LIR/SP->HOST)
- Exploit to infer denial of service activity, gauge infected worm population, detect misconfiguration, scanning and other reconnaissance
- Even more intelligent - pick up payload with active responders, coordinate, aggregate & correlate

# IMS Overview



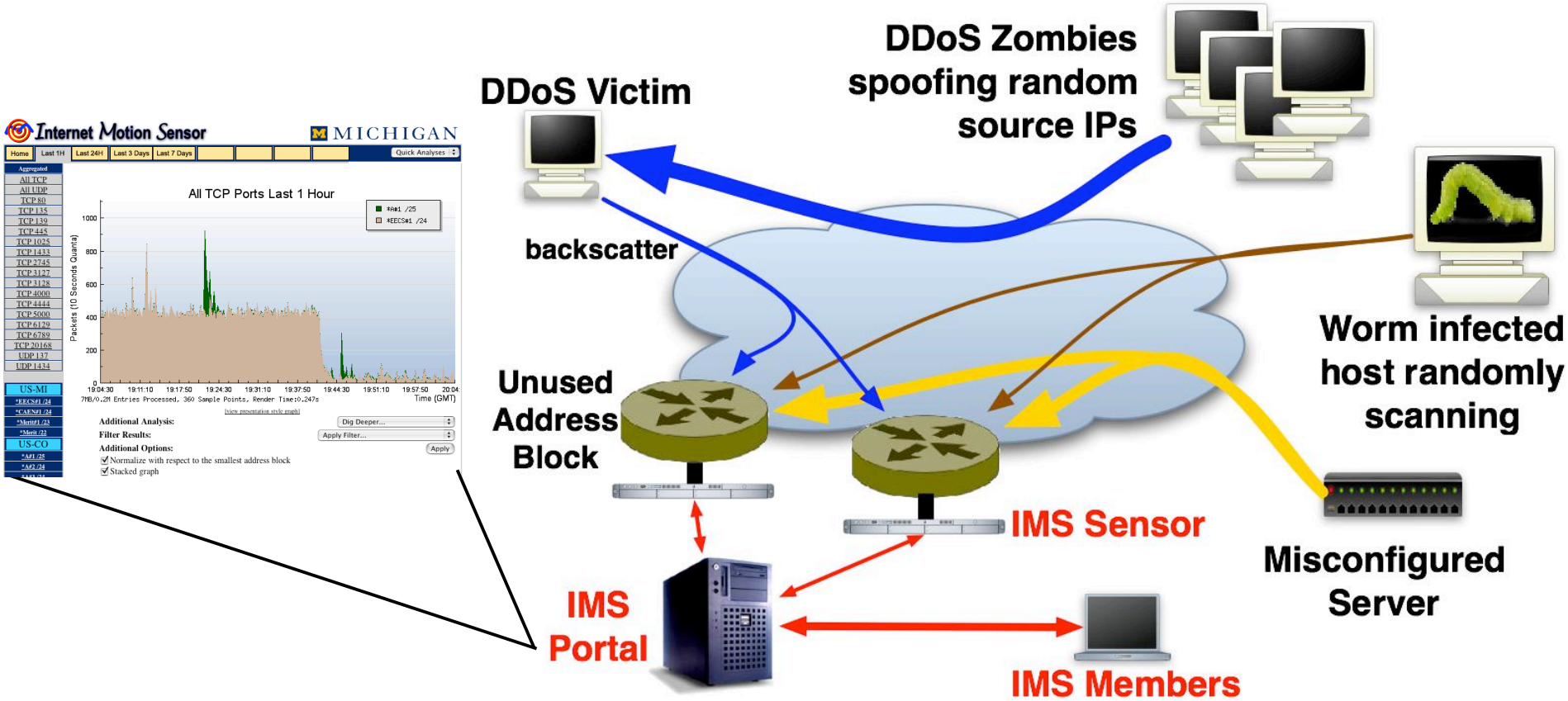

- There is significant malicious and non-productive activity on the Internet today (e.g. DoS, worms, botnets, misconfiguration)

# IMS Overview




- Much of this non-productive traffic is observed by unused addresses

# IMS Overview




- The IMS project monitors these unused address spaces (called ***darknets***) at *providers*, *enterprises*, and *academic institutions* to provide intelligence on global Internet threat activity.

# IMS Deployment

- **17,096,192** IPs monitored
- **1.15%** of routed IPv4 space
- **31** /8 blocks with an IMS sensor
- **21%** of *all* routable /8 blocks have at least one sensor

⇒Tier 1 SPs, Regional ISPs, National ISPs, Large Enterprises, Academic Networks

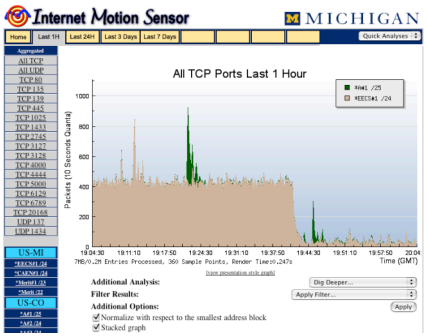
- Expanding IMS (5 continents soon)

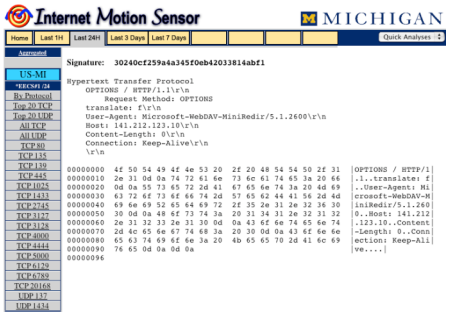
# Operational Value

**IMS Operational Utility:**

- IMS portal used to investigate anomalies:

  *"Anyone seeing an uptick on UDP 5060?"*




- Daily IMS reports provide detailed forensics on infected machines on your network:

| *Source IP* | *TCP Pkts* | *Top Dst Ports* |
|---|---|---|
| *10.0.153.156* | *219602* | *tcp/445:219593 tcp/80:9* |

# IMS Observations

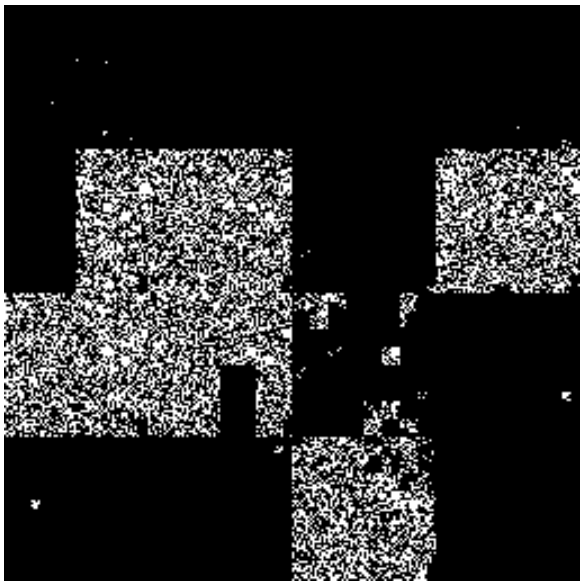
Two major trends observed with IMS:

1. **Attacks are more targeted** (e.g. botnet targeted scanning)

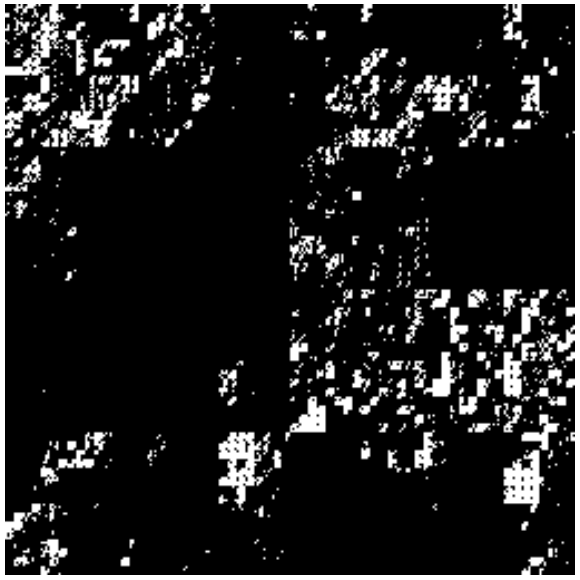
| Bot Command Detected | Δ IMS Detection | Scan Type |
|---|---|---|
| `ipscan r.r.r.r dcom2` | *11 secs* | *Global Random* |
| `ipscan 24.s.s.s dcom2` | - | *Local 24/8 Seq.* |
| `ipscan 69.27.s.s dcom2` | - | *Local 69.27/16 Seq.* |
| `ipscan s.s.s lsass` | *0 secs* | *Local /8 Seq.* |
| `ipscan s.s webdav3` | *0 secs* | *Local /16 Seq.* |

2. **Vulnerability ≠ Threat** (many threats today AgoBot/SDBot/GTBot leverage similar exploits)

# Ubiquitous Darknets



**Distributed Darknets Inside a /16**

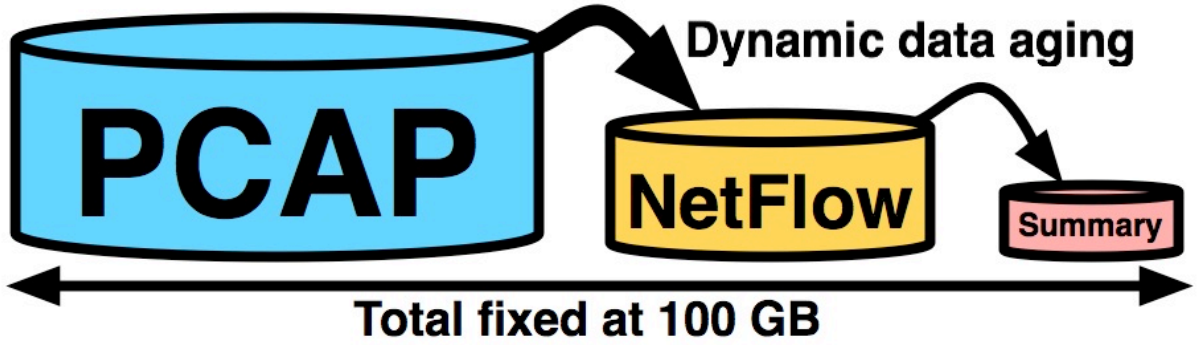


**Infected/Misconfigured Sources in the same /16**

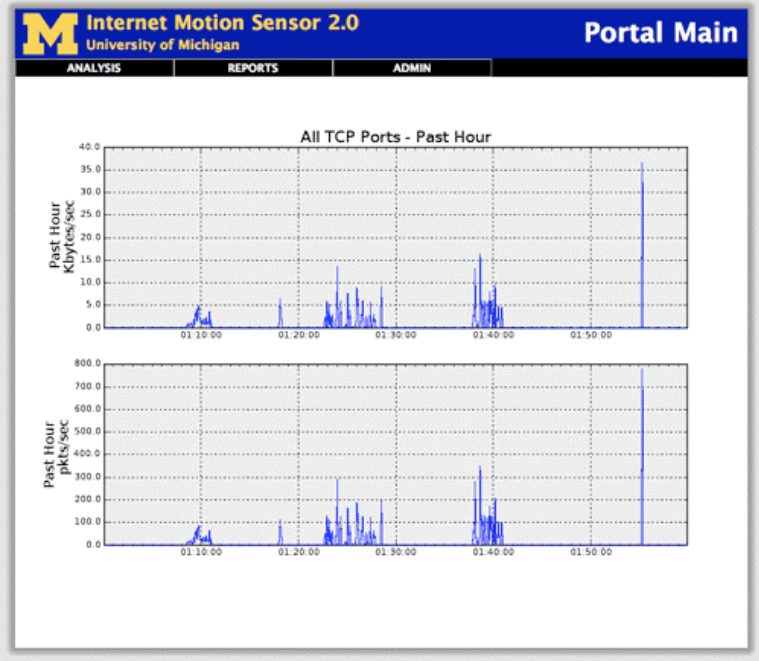
- To catch targeted attacks IMS now supports many noncontiguous darknets within a network
- Data visualization w/2D Quad Charts courtesy of IPMAPS: monkey.org/~phy/ipmaps

# Resource-aware Data Collection

- Each IMS sensor dynamically adjusts its requirements based on resource availability
- Historical data is ***dynamically scaled*** into more compact representations as it ages
- Constructed from the ground up using ***standard formats***: *pcap*, *NetFlow*, *text*



# Infinite Time Queries

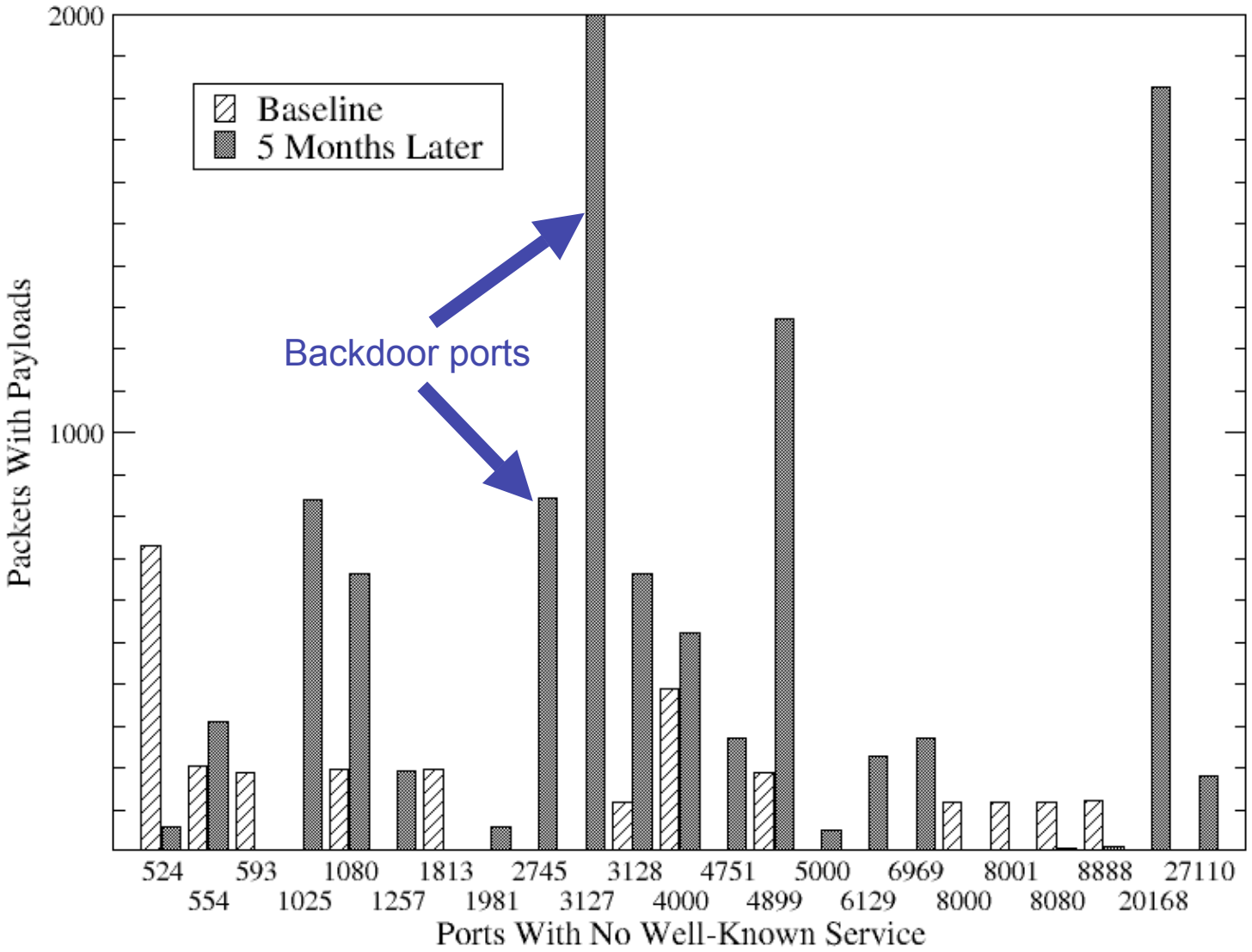



- Ability to query over the entire history of the sensor (hour, day, week, month, year, etc)
- Support for complete ***pcap filter expressions*** (can run over full historical data)

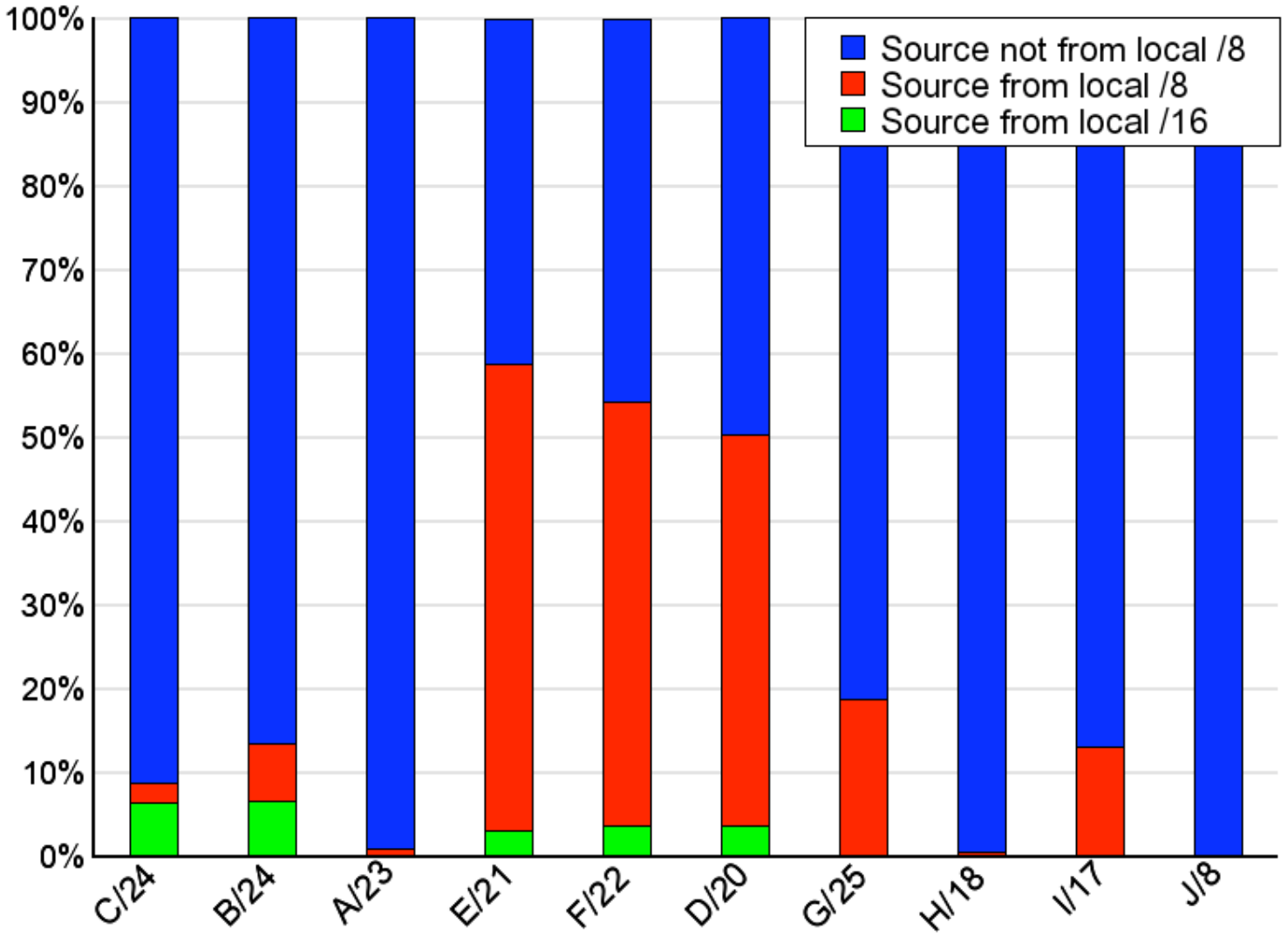
- Can view ***full payload*** for months to years depending on space allocation (~1 year on /24)

# Ports with the biggest changes over a 5 month timeframe

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



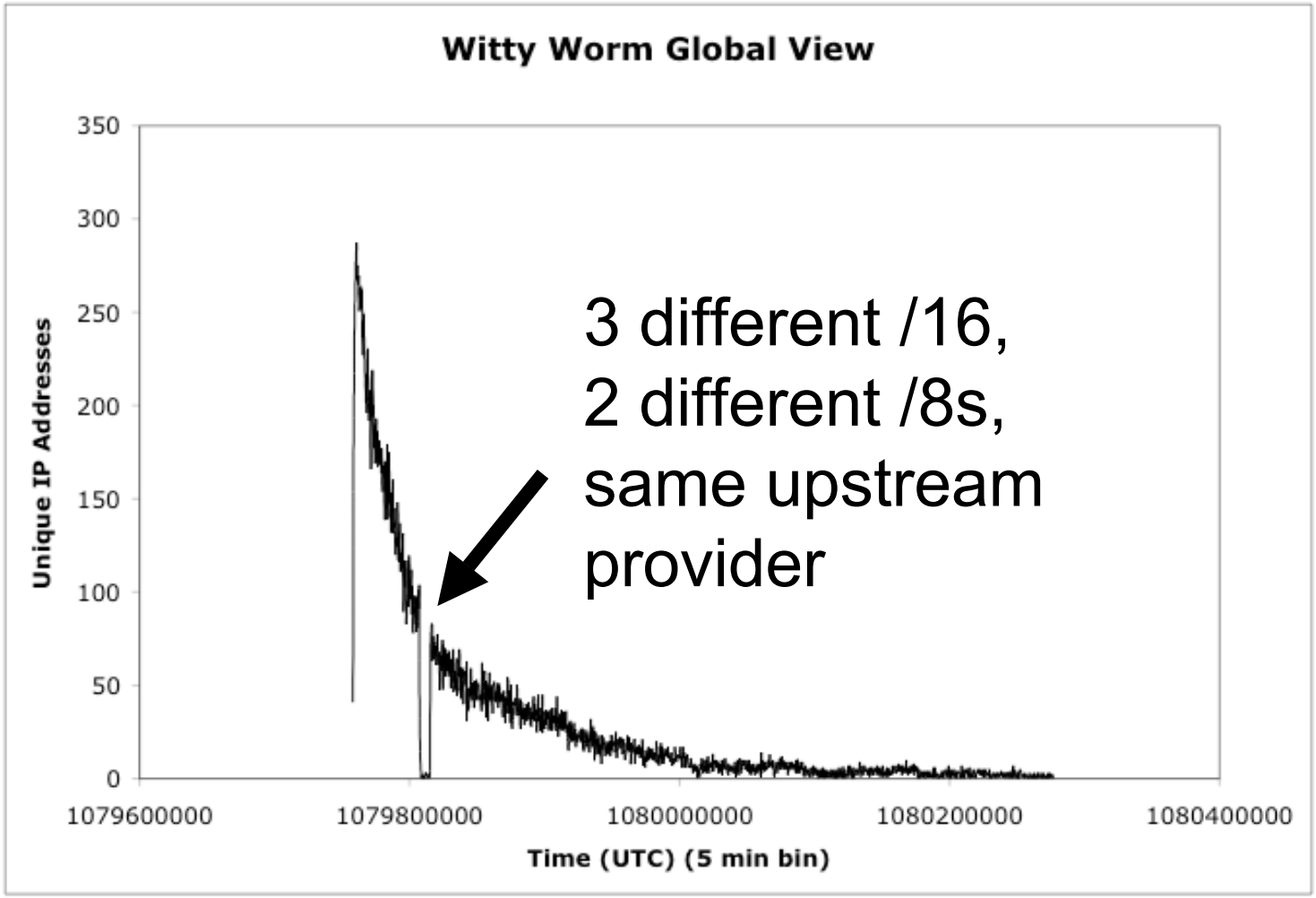
# % of packets from local /16, /8, or global at 10 sensors over 1 week



# 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 |

# Unique SRCs: 3 sensors, 5 minute bins



# 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)

# Flexible Honeypot Responders

- To catch more complex attacks IMS capture architecture now supports different active responders: example: ims.conf

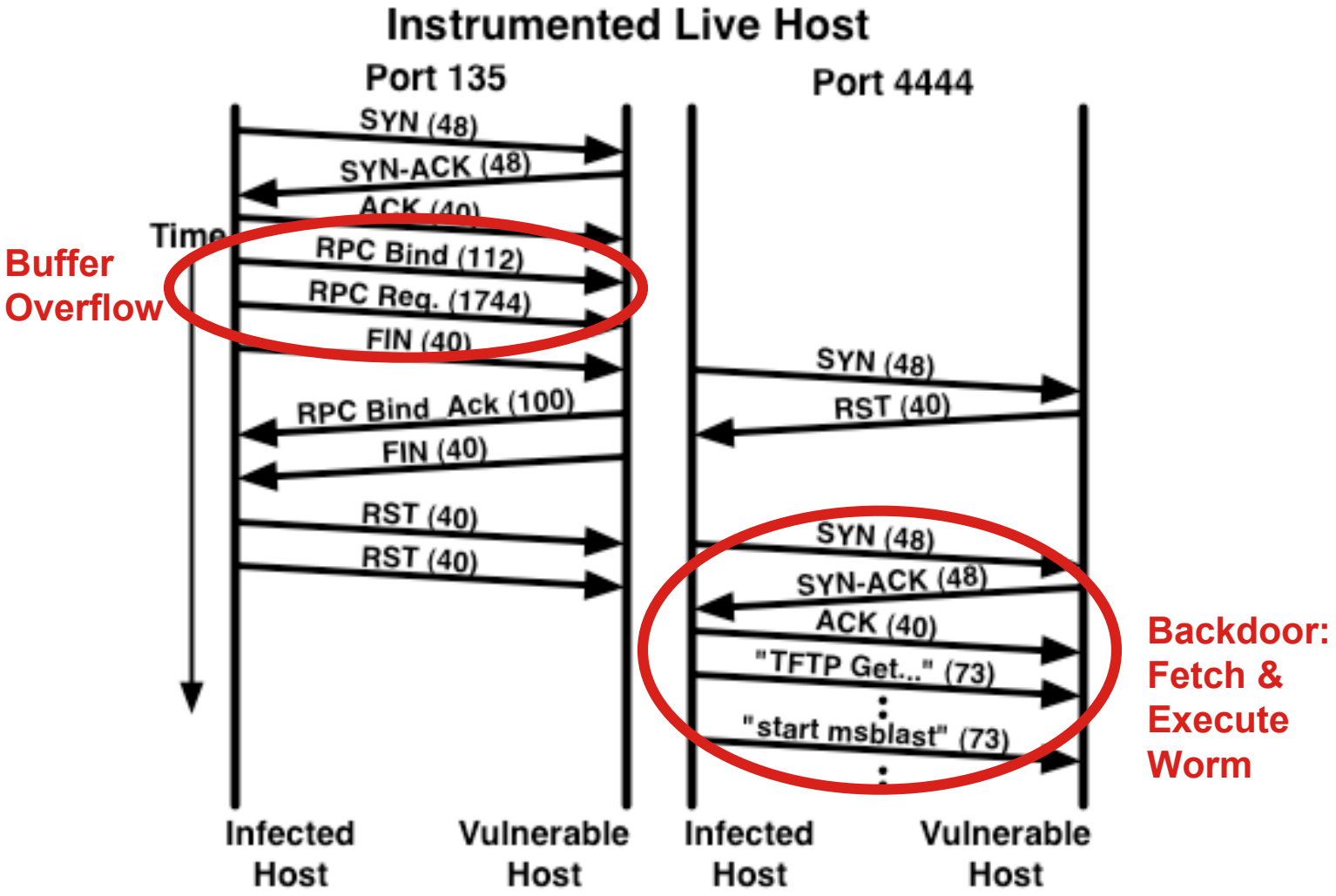
```
# darknet declarations
darknet my16 {
    filter "dst net xxx.xxx.0.0/16"
    responder passive
    capture pcap nf summary
    path /usr/local/ims/data
    size 70 GB

    # sub darknet
    darknet my16-sub-synack24 {
        filter "dst net xxx.xxx.0.0/24"
        responder synack
        capture pcap nf summary
        path /big-disk
        size 10 GB
}}
```

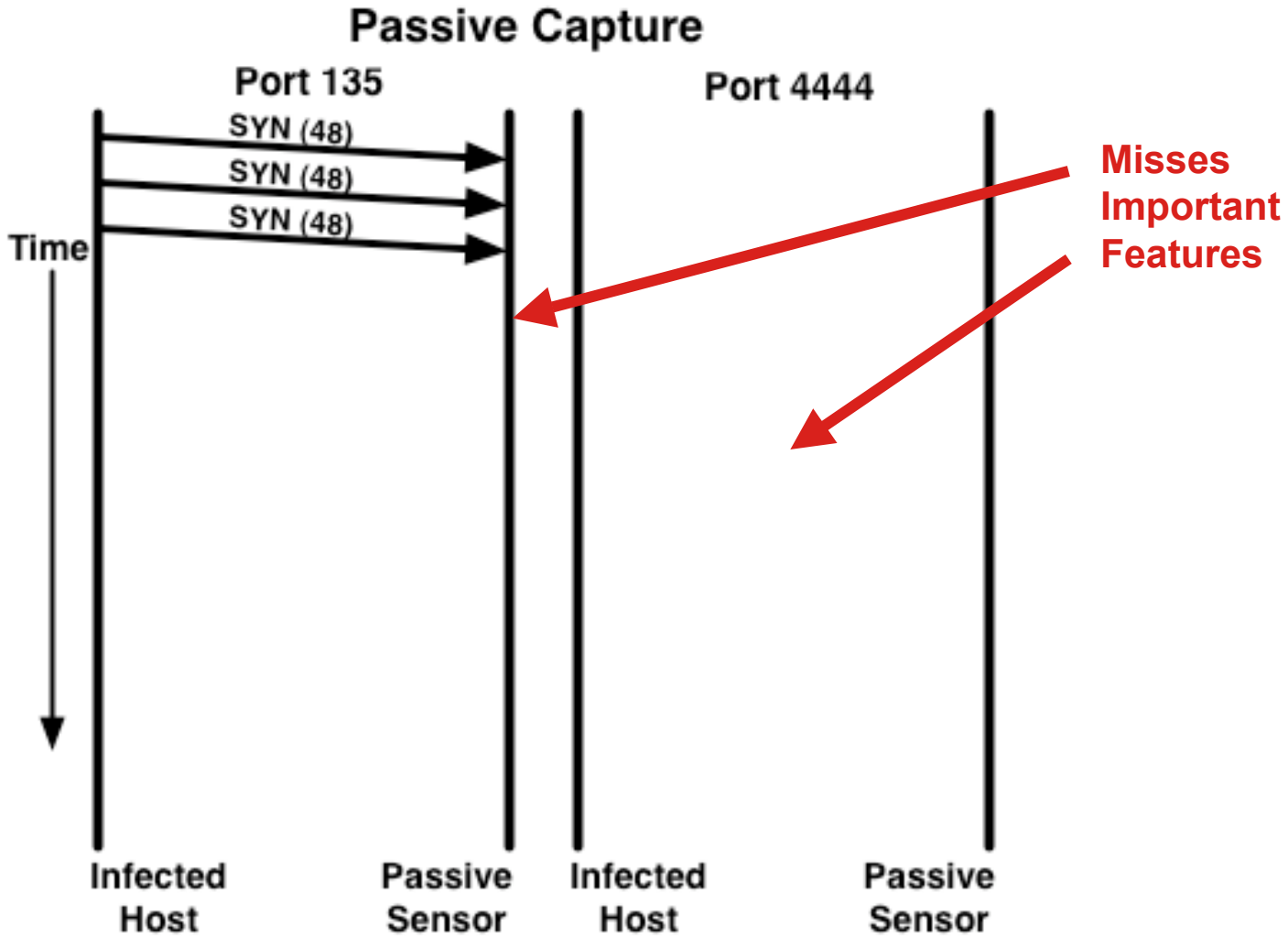
Can have different responder strategies on different darknets or portions of larger darknet

Could integrate w/ honeyd

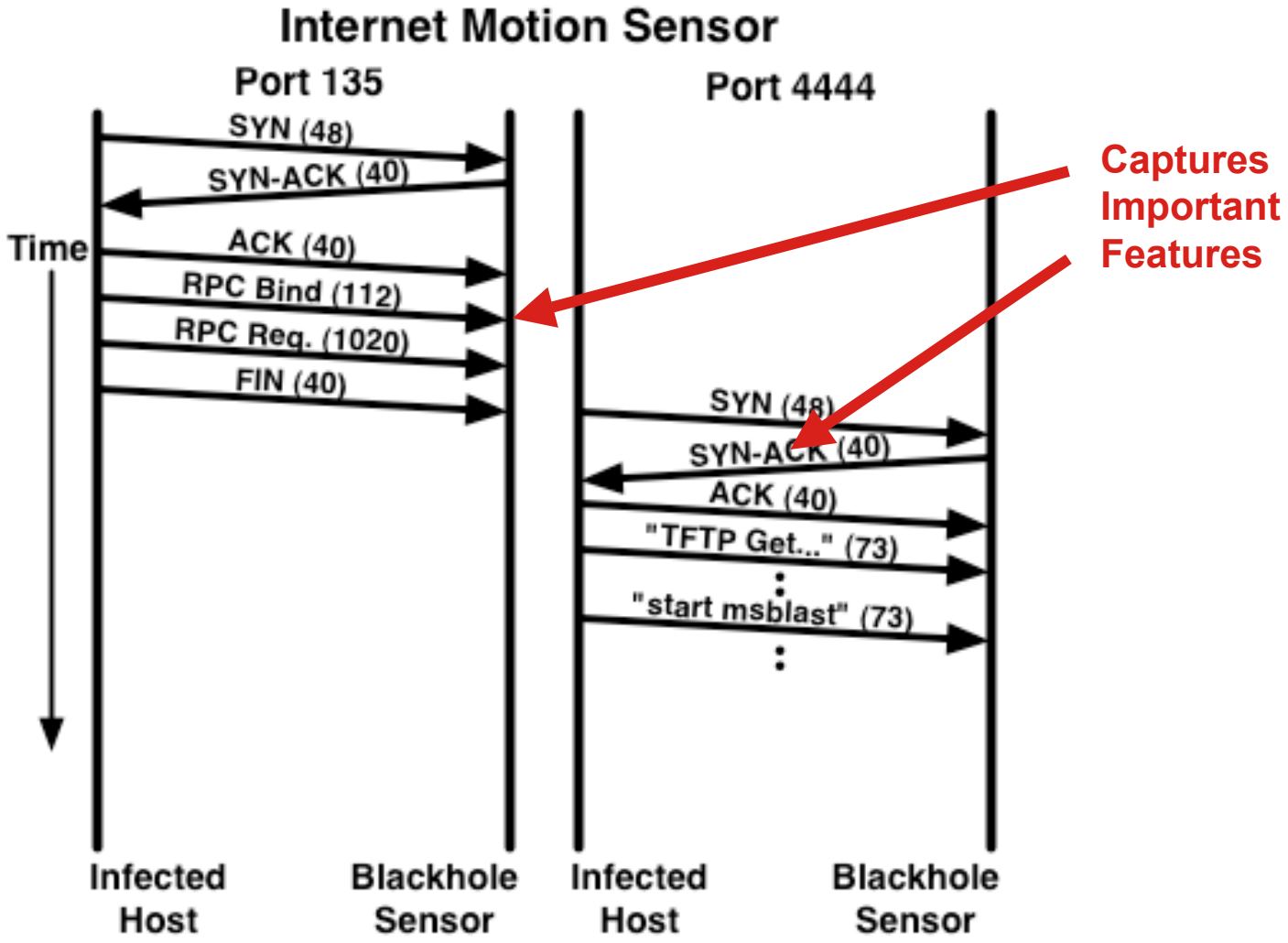
# The Blaster Worm - Live Host



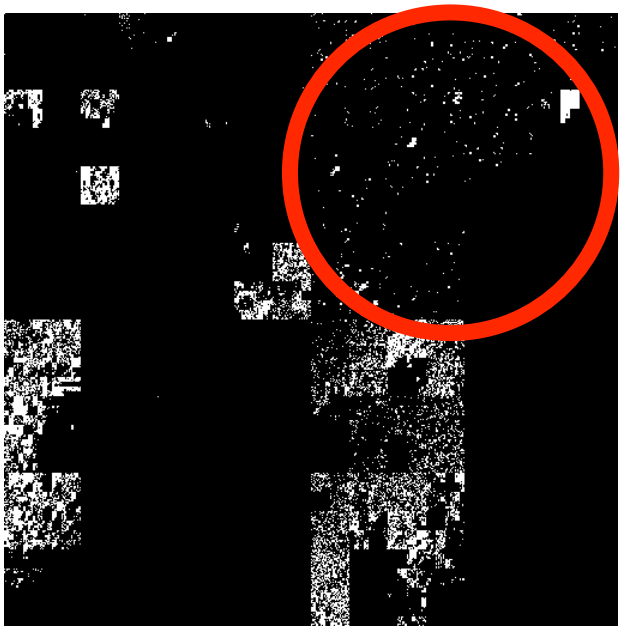
# The Blaster Worm - Passive



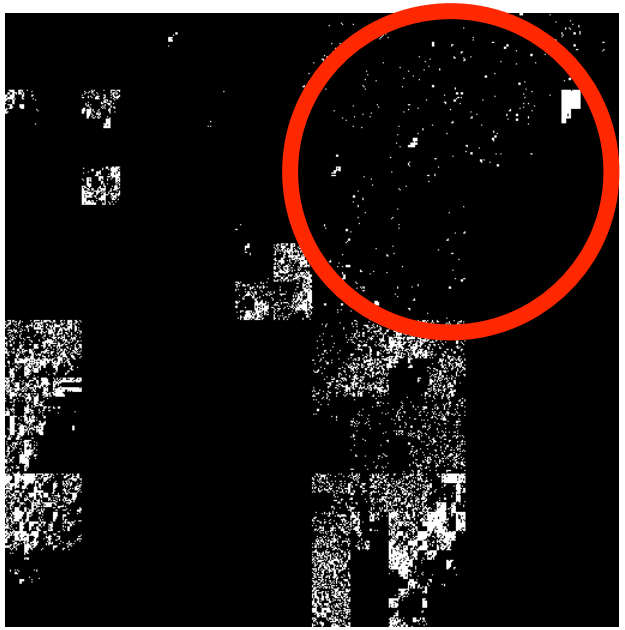
# The Blaster Worm - IMS



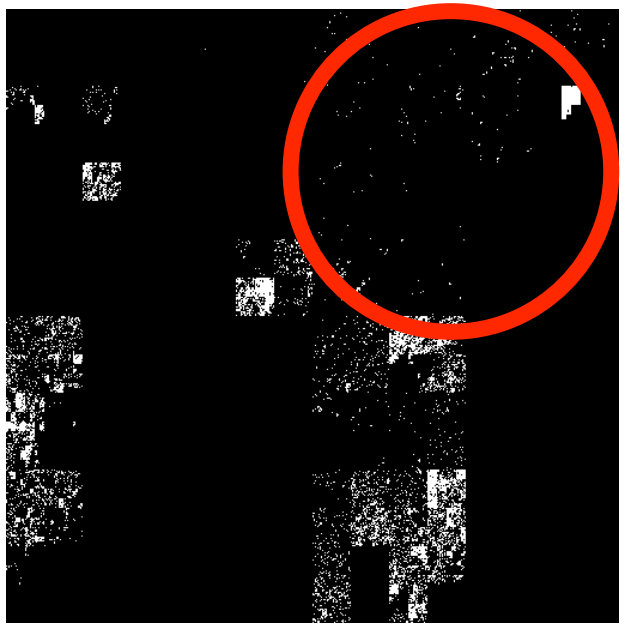
# The Worms



**Slammer**



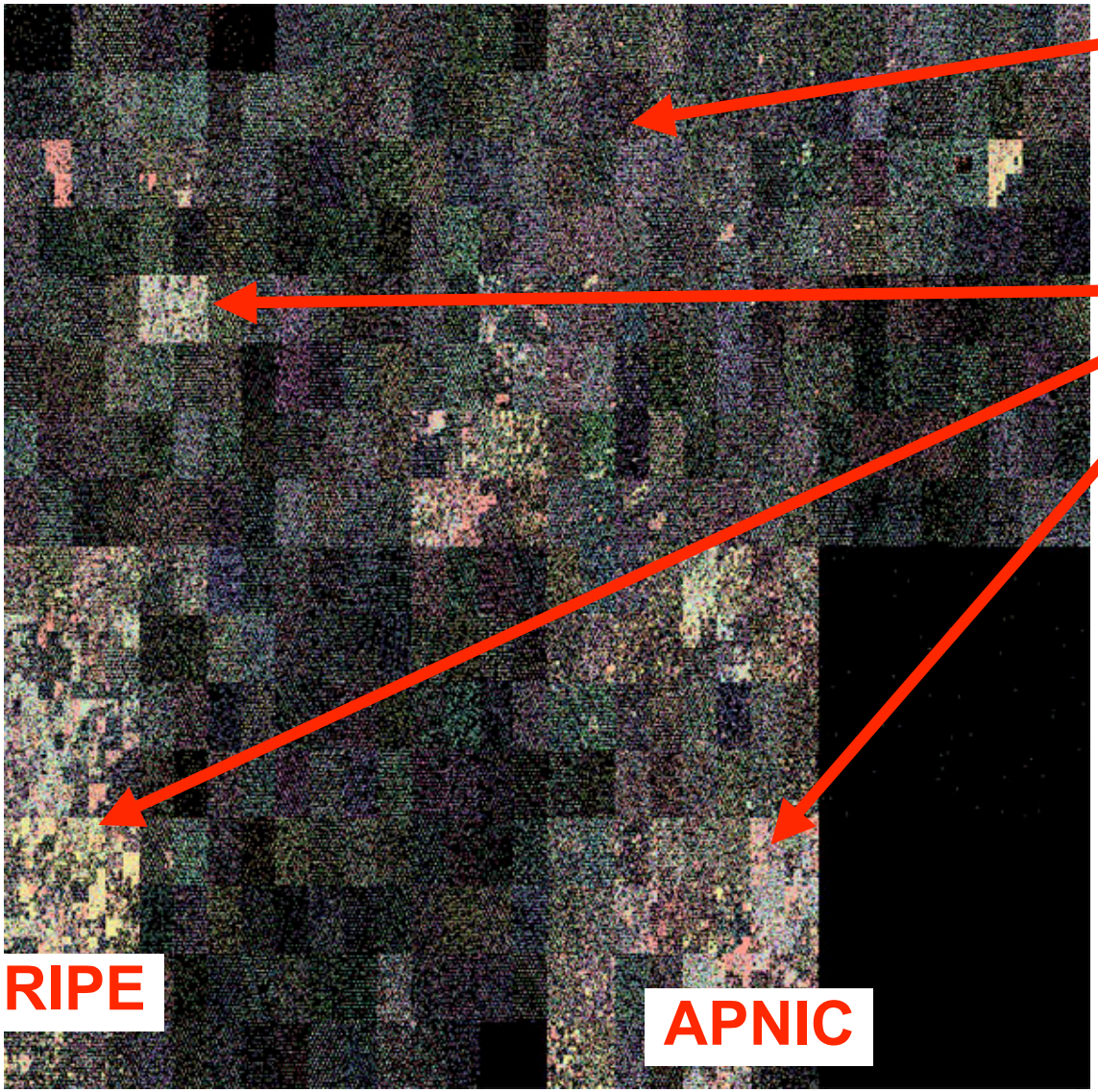
**Blaster**



**CodeRed II**

- Worm infected source addresses highly distributed over IP space
- Although… notice how few sources from Class B allocation space… enterprise egress filtering?

# Example: /8 Darknet

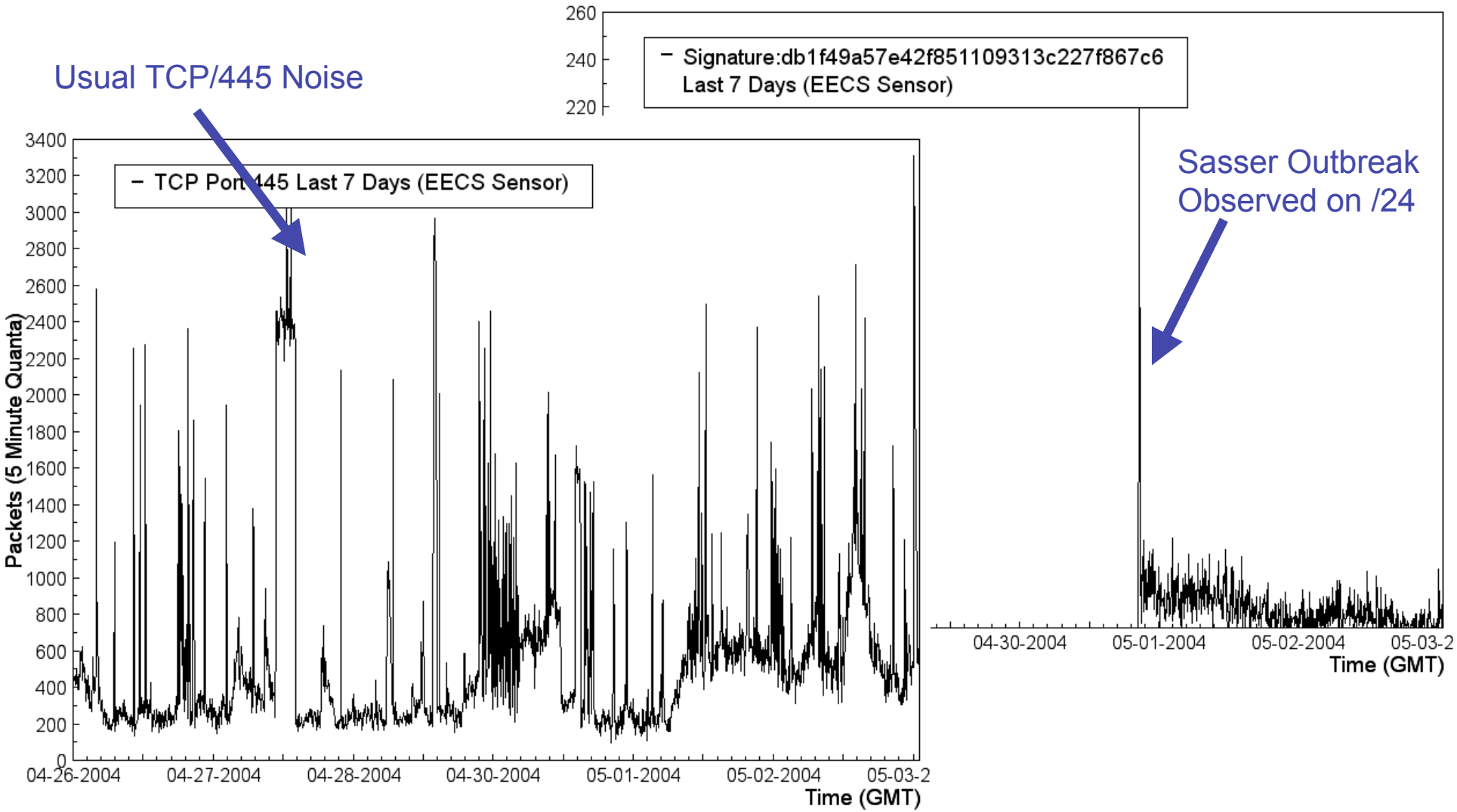



- Visualization provides insight into attack source distributions

# Payload Caching

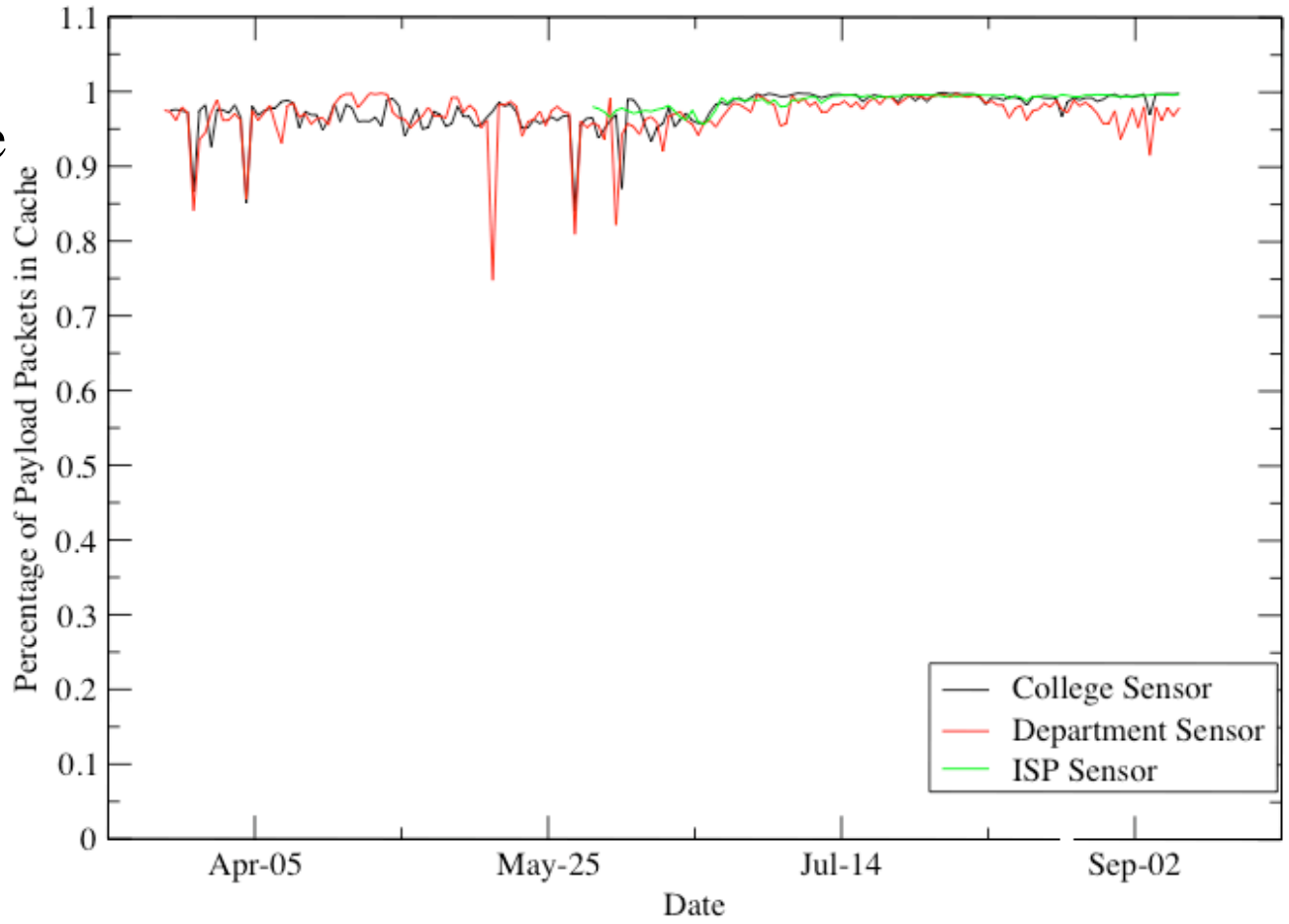
- Active responder produces lots payload data
- Solution: only store payloads if they are ‘new’
- Implementation: take MD5 hash of payloads and only store payloads which have a unique hash

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



# % of Payload Cache Hits over 5 Months at 3 sensors

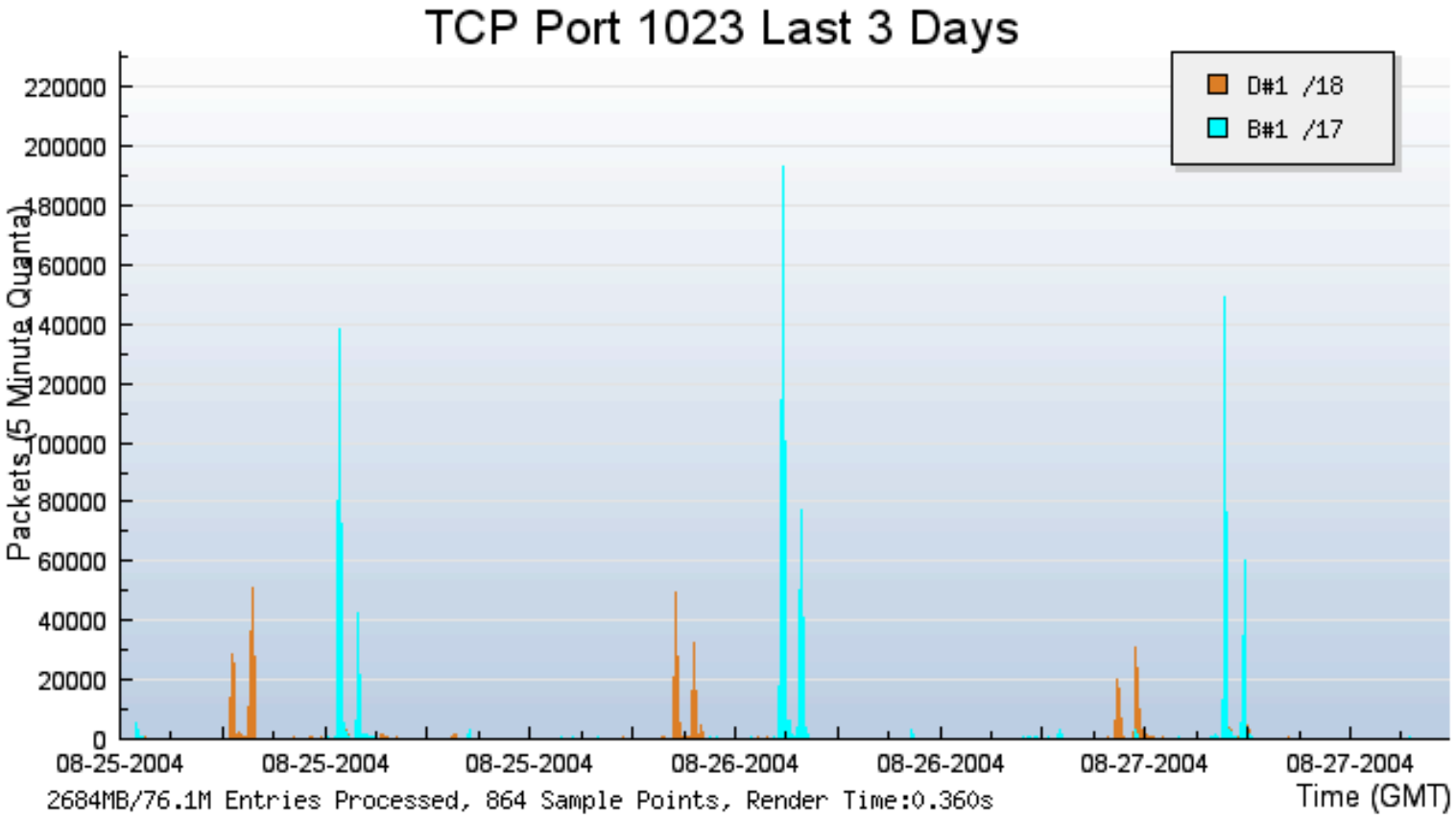
- ~95% signature cache hit-rate
- Most payloads have been seen before



# Worms

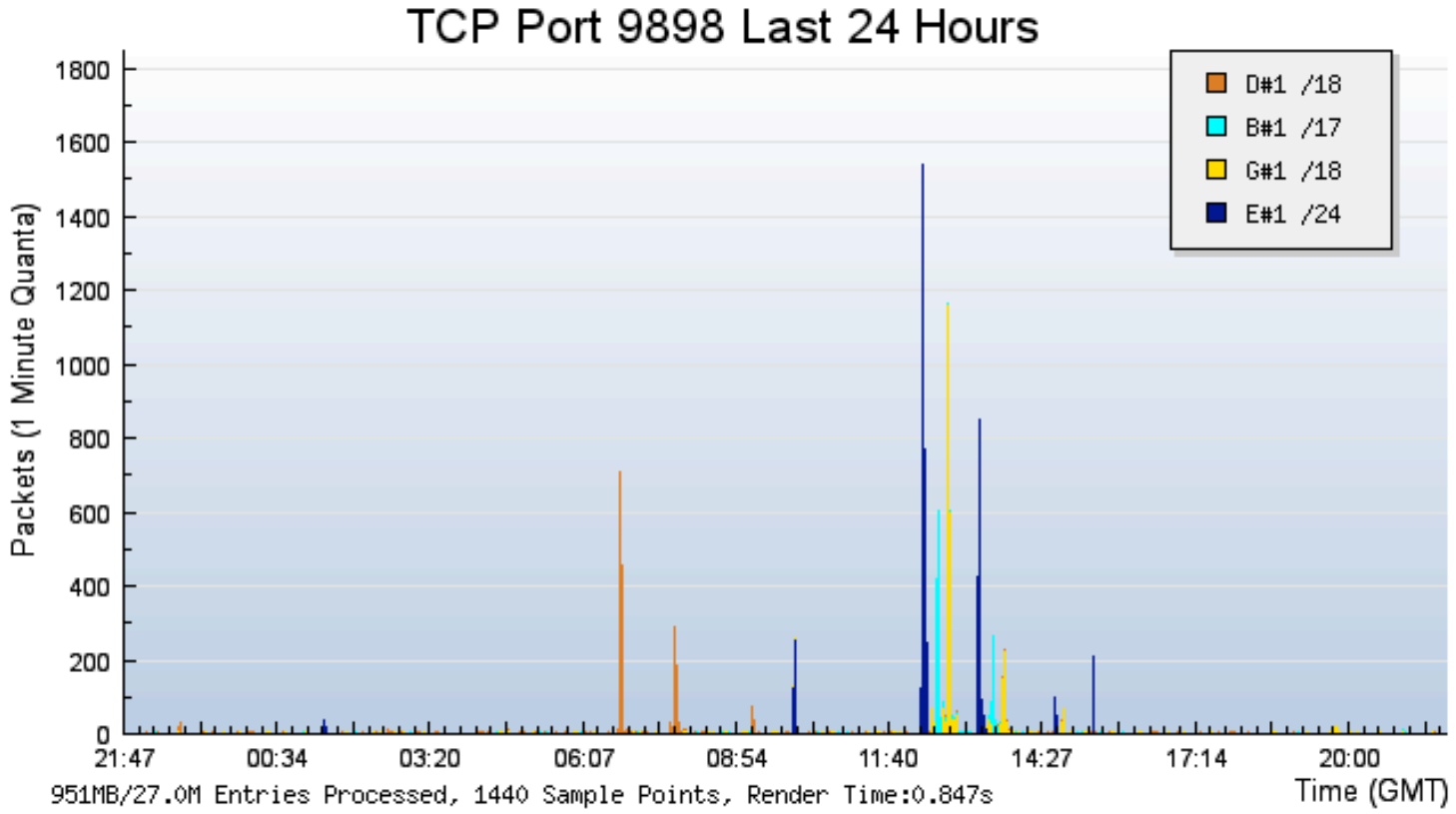
| Worm | Sasser | Sasser.e | Dabber.a |
|---|---|---|---|
| Vulnerability | LSASS (MS04-011) | LSASS (MS04-011) | Sasser-FTP |
| Population | Windows XP Windows 2K | Windows XP Windows 2K | Sasser infected hosts |
| Scan Port | TCP/445 | TCP/445 | TCP/5554 |
| Backdoor Port | TCP/5554 | TCP/1023 | TCP/9898 |
| 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



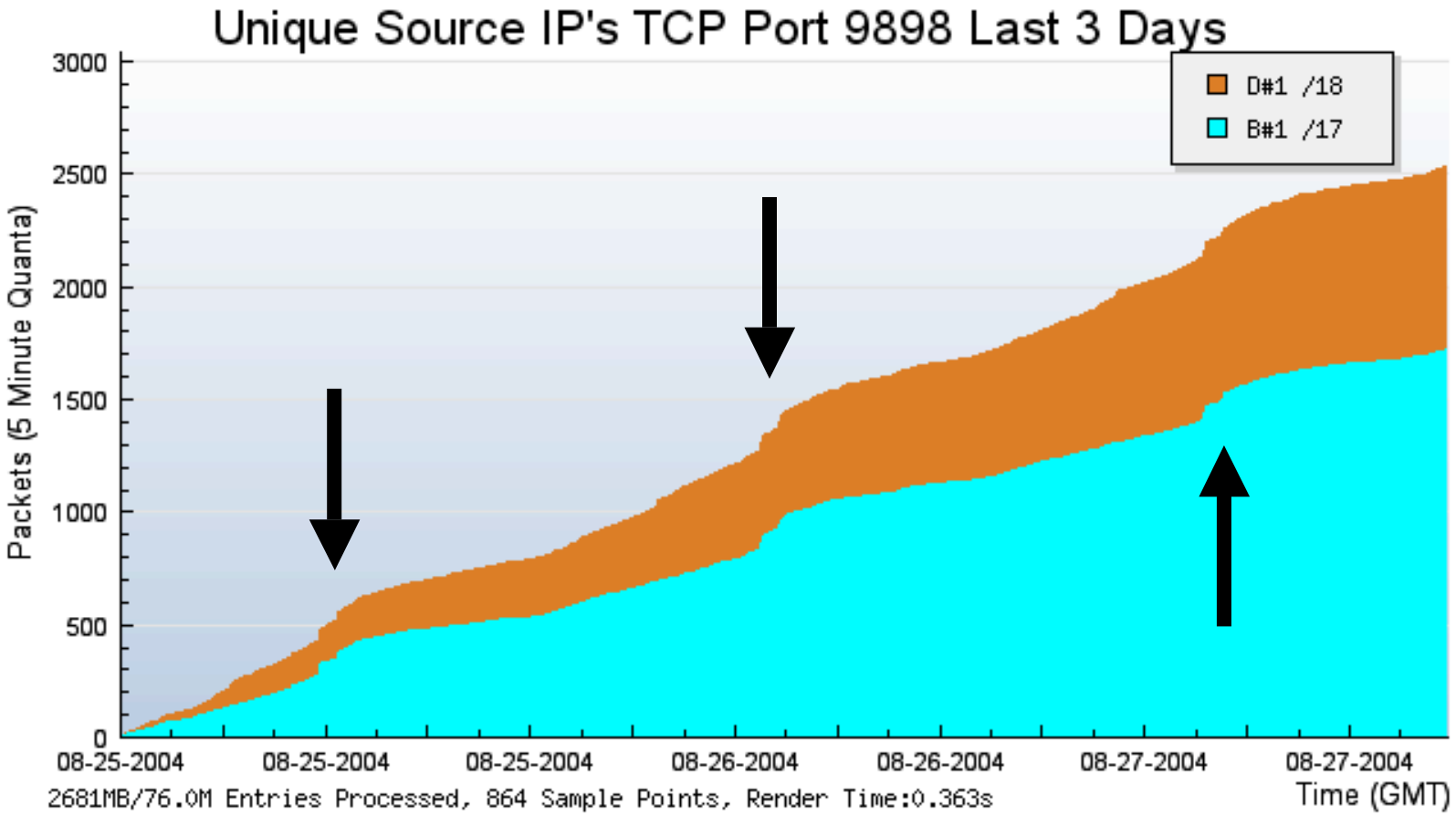

- Large, short lived spikes
- Same shaped graph across (1023, 5554, 9898)
- Nearly all sources in China and Korea

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



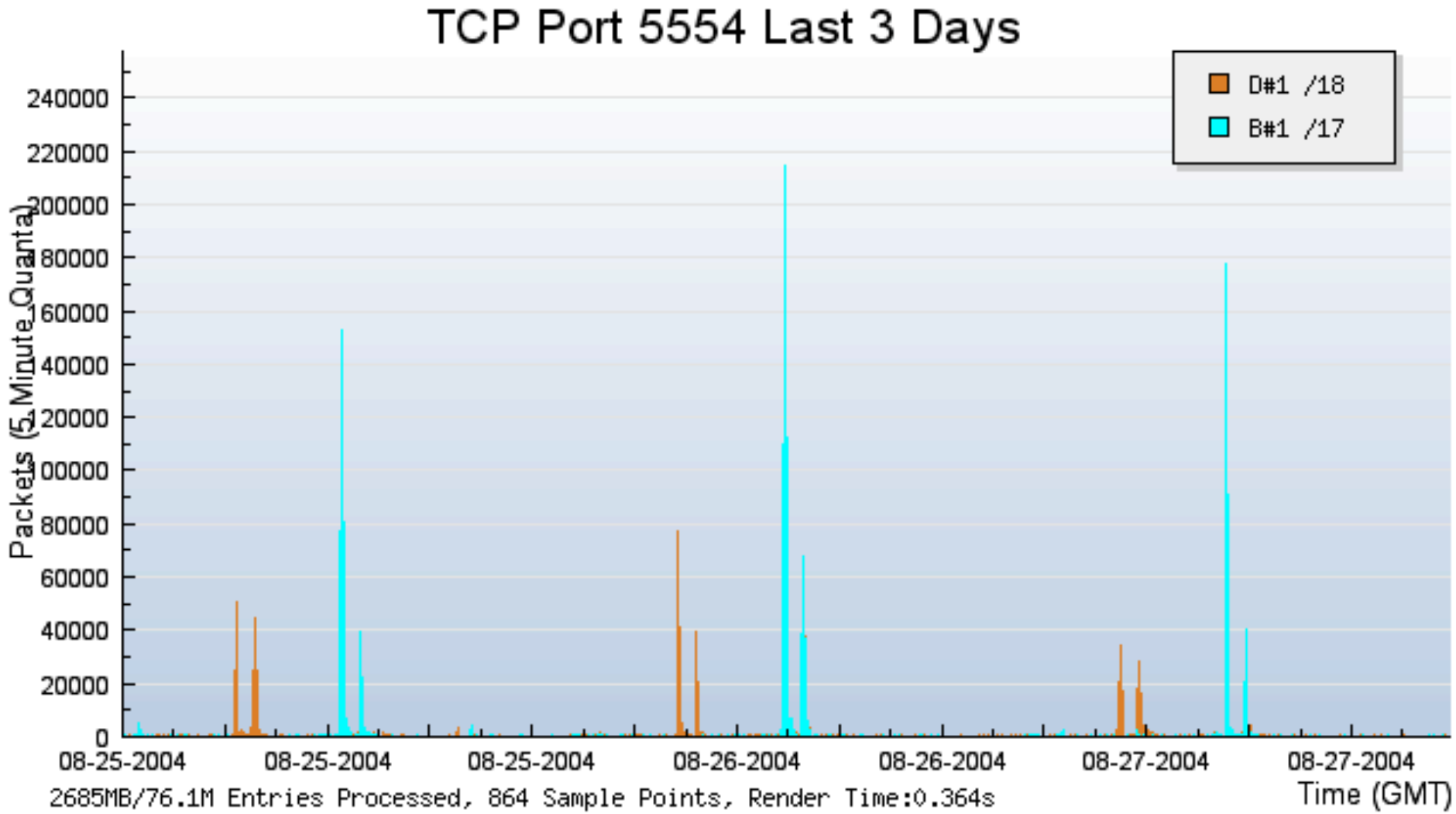

- D -> E -> B -> G (Ordered by /8)
- ~6 /8's an Hour

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

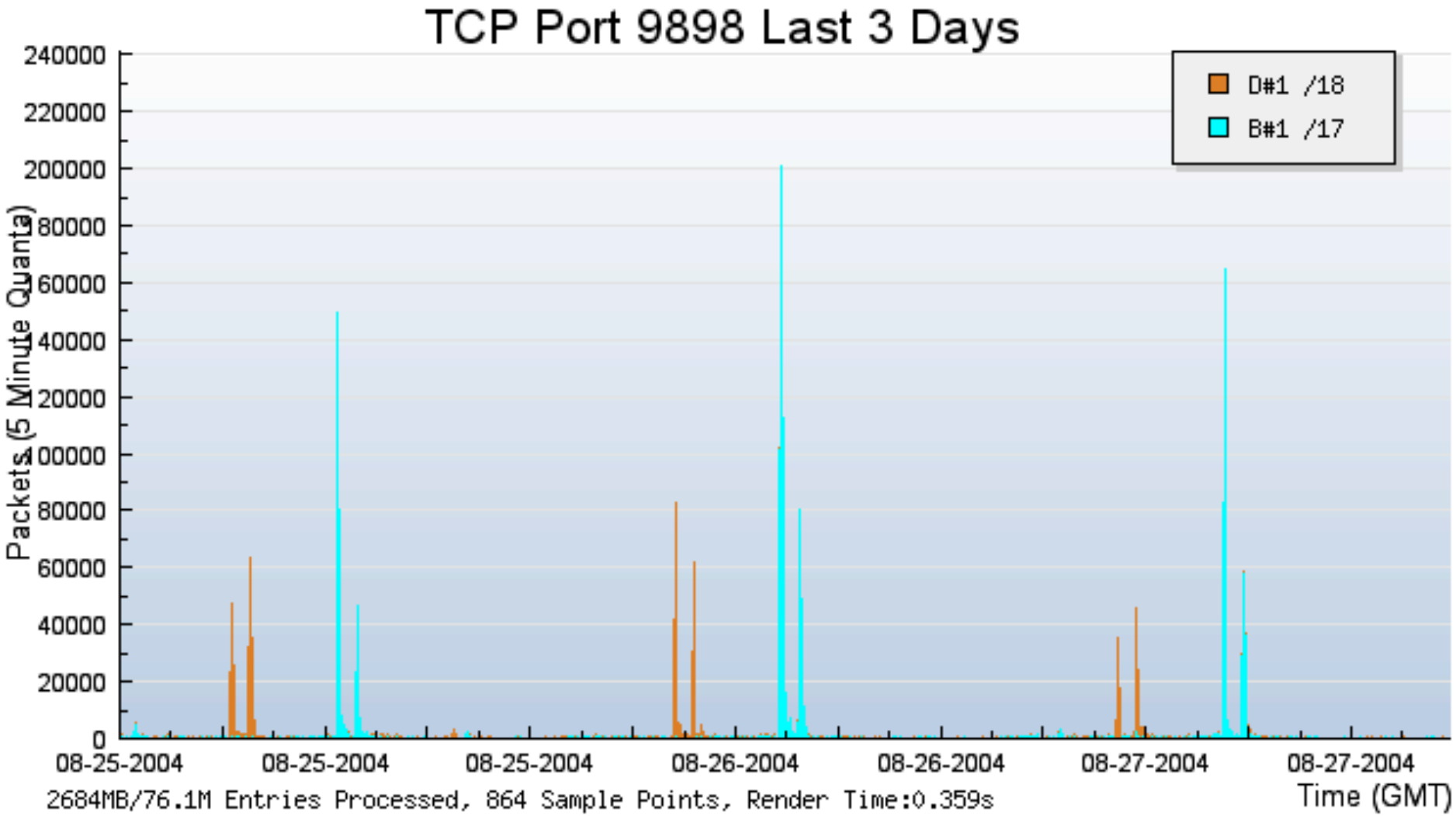



- small number of hosts are involved (~100)
- the size of the bumps is similar each time
- Hosts dwarfed by background noise

# Packets per 5 minutes on a /17 and a /18 over 3 days for TCP/5554



# Packets per 5 minutes on a /17 and a /18 over 3 days for TCP/9898



# 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

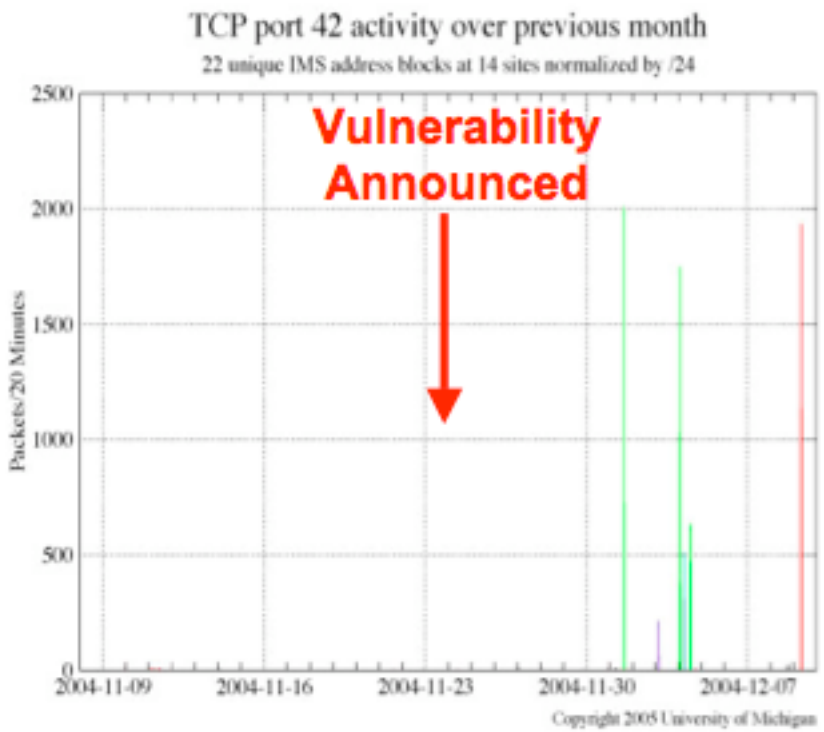
    00000000 55 53 45 52 20 78 0a |USER x.|

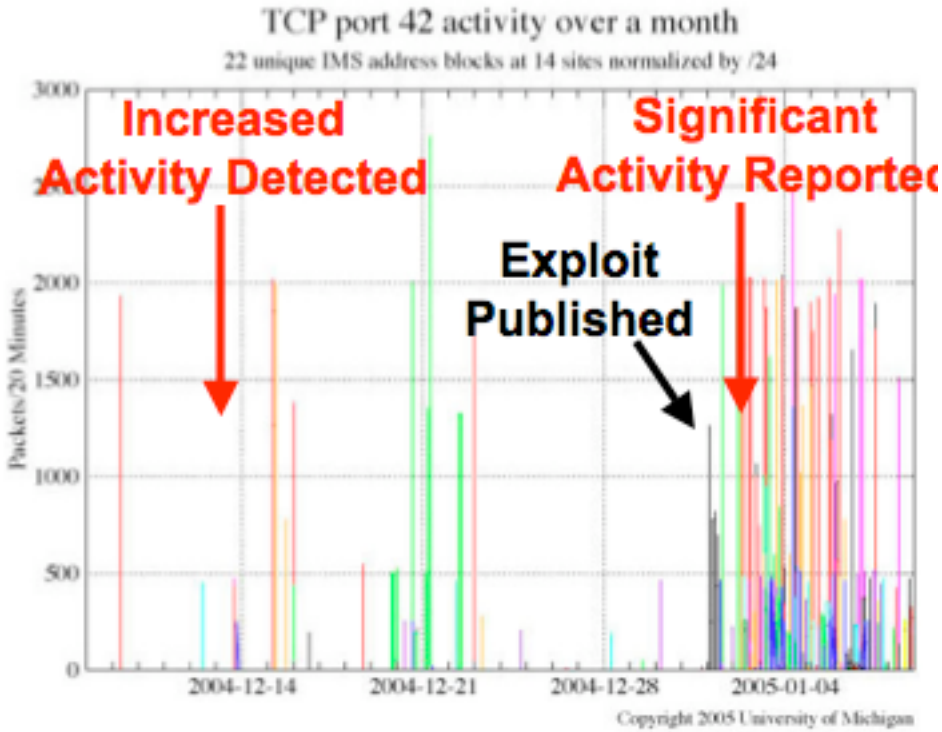
    00000007
  - F623e75af30e62bbd73d6df5b50bb7b5

    00000000 44 |D| 00000001

# TCP 42 Activity

- November 24, 2004 vulnerability announced on remotely exploited overflow in the WINS server component of Microsoft Windows
- December 2004 an increase in activity to TCP/42 was detected
- January 2005 news of significant amounts of increased activity on TCP/42 was noted in multiple reports



# TCP 42 Payloads

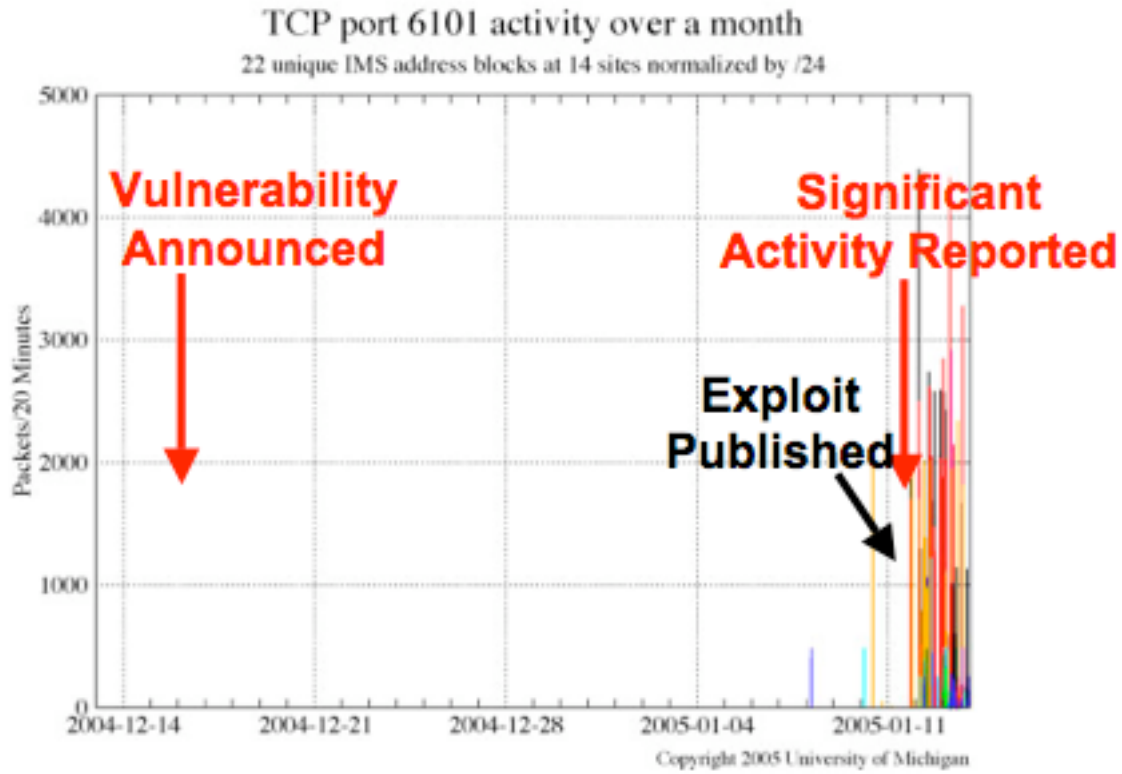
- Captured live payloads that match byte-for-byte with template exploit code
- Same exploit is being used to reinject many different payloads (same exploit with very different shellcode)

```
00000000  00 03 0d 4c 77 77 ff 77  05 4e 00 3c 01 02 03 04  |...Lww.w.N.<....|
00000010  6c f4 3d 05 00 02 4e 05  00 02 4e 05 00 02 4e 05  |l.=...N...N...N.|
00000020  00 02 4e 05 00 02 4e 05  00 02 4e 05 00 02 4e 05  |..N...N...N...N.|
00000030  00 02 4e 05 90 01 4e 05  90 00 4e 05 90 00 4e 05  |..N...N...N...N.|
00000040  90 00 4e 05 90 00 4e 05  90 00 4e 05 90 00 4e 05  |..N...N...N...N.|
00000050  90 03 4e 05 90 00 4e 05  90 01 4e 05 90 00 4e 05  |..N...N...N...N.|
00000060  90 00 4e 05 90 00 4e 05  90 00 4e 05 90 00 4e 05  |..N...N...N...N.|
00000070  90 00 4e 05 90 03 4e 05  90 00 4e 05 90 01 4e 05  |..N...N...N...N.|
00000080  90 00 4e 05 90 00 4e 05  90 00 4e 05 90 00 4e 05  |..N...N...N...N.|
00000090  90 00 4e 05 90 00 4e 05  90 03 4e 05 90 00 4e 05  |..N...N...N...N.|
000000a0  90 01 4e 05 90 00 4e 05  90 00 4e 05 90 00 4e 05  |..N...N...N...N.|
000000b0  90 00 4e 05 90 00 4e 05  90 00 4e 05 90 03 4e 05  |..N...N...N...N.|
000000c0  90 00 4e 05 90 01 4e 05  90 00 4e 05 90 00 4e 05  |..N...N...N...N.|
000000d0  90 00 4e 05 90 00 4e 05  90 00 4e 05 90 00 4e 05  |..N...N...N...N.|
000000e0  90 03 4e 05 90 00 4e 05  90 01 4e 05 90 00 4e 05  |..N...N...N...N.|
000000f0  90 00 4e 05 90 00 4e 05  90 00 4e 05 90 00 4e 05  |..N...N...N...N.|
```

- Evidence suggests attacks are from manual activity and not automated worm
- However, vulnerability is wormable
- http://ims.eecs.umich.edu/reports/port42

# TCP 6101 Activity

- December 16, 2004 iDEFENSE Announces Buffer Overflow vulnerability in Veritas Backup Agent
- January 11, 2005 Hat-Squad publishes exploit code
- January 11, 2005 IMS Detects activity on TCP/6101



# TCP 6101 Payloads

- Captures live payloads that match byte for byte with template exploit code:

```
0000  02 00 32 00 90 90 90 90 31 f6 c1 ec 0c c1 e4 0c   ..2.....1.......
0010  89 e7 89 fb 6a 01 8b 74 24 fe 31 d2 52 42 c1 e2   ....j..t$.1.RB..
0020  10 52 57 56 b8 ff 50 11 40 c1 e8 08 ff 10 85 c0   .RWV..P.@.......
0030  79 07 89 dc 4e 85 f6 75 e1 ff e7 90 90 90 90 90   y...N..u........
0040  90 90 90 90 90 90 90 90 a1 ff 42 01 90 90 90 90   ..........B.....
0050  90 90 90 90 90 90 90 90 90 90 90 90 90 90 90 90   ................
0060  90 90 90 90 90 90 90 90 90 90 90 90 90 90 90 90   ................
0070  90 90 90 90 90 90 00 31 2e 31 2e 31 2e 31 2e 31   .......1.1.1.1.1
0080  2e 31 00 eb 80                                    .1...
```

- Evidence suggests attacks are from manual tools and not automated worm
- Vulnerability is wormable
- Both port 42 & 6101 were zero-day threats!  Exploits released and same day attacks began
- http://ims.eecs.umich.edu/reports/port6101

# References

- Check out IMS site
- Check out Arbor site or email me…
- Lots of references and research papers (e.g., worm04, sruti, etc..) on detecting & distributing botnets, building darknets, implications on sensor placement, etc.., should be easy to find..

# Thanks!

danny@arbor.net