# Decoding Anonymized Packet Data

Finding useful patterns and trends in anonymized packet data.

Shawn McCaffrey
Shrikant Pandhare
Paul Pasquale
David Rennicker

## Agenda

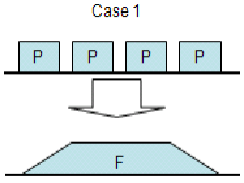
- OS Profiling
- Media Profiling
- Flow Analysis

## OS Profiling - Limited Packet Data vs. Flow Data

- Limited Packet Data
  - Only captures parts of packet
  - Drops payload
  - Necessary fields
    - Source, destination IP
    - Source, destination ports
    - Size
    - Time

## OS Profiling - Limited Packet Data vs. Flow Data

- Flow Data
  - A flow is a set of packets related closely in time
  - Our flows are defined by five attributes
    - Source IP
    - Destination IP
    - Source Port
    - Destination Port
    - Protocol in IP
  - Flows are generated by sensors on the network




* Information and diagram from 9/15 lecture slides.

## OS Profiling - Ephemeral Ports

- Ephemeral Ports
  - Temporary client-side ports used when opening up a TCP connection.
  - The range of ephemeral ports are usually a default setting in the operating system's TCP stack.
    - Windows XP and older: 1024 – 4999
    - Windows Vista and Server: 49152 - 65535
    - Apple Mac OS: 49152 - 65535
    - Linux (2.4 kernel): 32768 – 61000
    - Free BSD: 1024 – 65535 (Assigned Randomly)
    - Solaris: 32768 – 65535

http://www.cnftp.com/nctfpd/doc/misc/ephemeral_ports.html
http://blogs.msdn.com/drnick/archive/2008/09/19/ephemeral-port-limits.aspx

## OS Profiling - More Fingerprinting Options

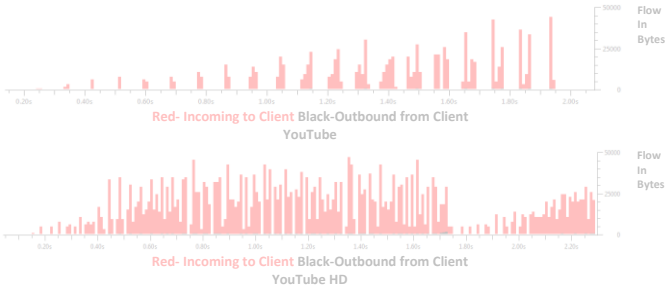
- pof – passive os fingerprinting tool
  http://lcamtuf.coredump.cx/pof.shtml
  - determines operating system based on traffic intercepted on the wire, based on behavior, flags, ttl etc.

- nmap – active os fingerprinting
  http://nmap.org/
  - determines operating system by sending various packets to the host and observing the response

# Media Profiling - Video on Demand

**YouTube/YouTube HD**

- Sends traffic using HTTP over port 80.
- Delivers the file as opposed to streaming the file




# Media Profiling - Video on Demand

**Hulu/Hulu HD**

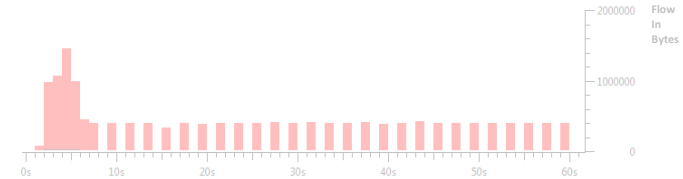
- Server delivers traffic using port 1935, Flash using RTMP.
- Streams the file as opposed to delivers the file.

# Media Profiling - Video on Demand

## Discovery

- Server delivers traffic using port 80, to a proprietary player.
- Streams the file as opposed to delivers the file.

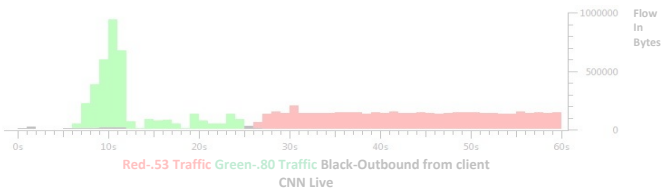
| Address A | Port A | Address B | Port B | Packets | Bytes | Packets A->B | Packets A<-B | Rel Start | Duration |
|---|---|---|---|---|---|---|---|---|---|
| 71.199.97.123 | 4295 | 8.15.32.17 | http | 5912 | 6127776 | 1777 | 4135 | 1.585732 | 76.0053 |
| 71.199.97.123 | 4292 | 8.15.32.17 | http | 6300 | 6352283 | 1836 | 4464 | 0 | 77.435 |
| 71.199.97.123 | 4294 | 8.15.32.17 | http | 6747 | 6961128 | 2008 | 4739 | 1.585114 | 75.9248 |




# Media Profiling - Live Video Services

## CNN Live Video

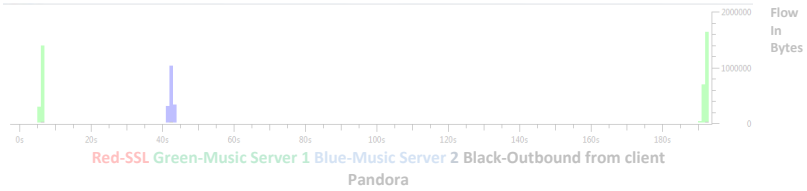
- First uses RTSP(TCP 554) to deliver content.
- Then switches to UDP to continue delivering the content.



# Media Profiling - Audio on Demand

**Pandora**

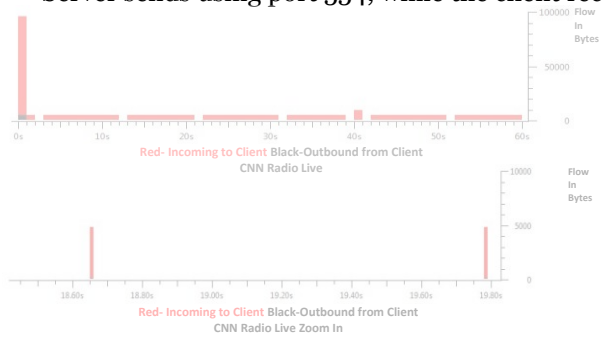
- SSL login before song plays
- Server sends using port 554
- Song sent as a file delivery
- Each new song is a new TCP connection.




# Media Profiling - Audio Live

**CNN Live**

- Server sends using port 554, while the client receives on port 1257.



## Profiling and Flow Data

- The majority of our techniques are only useful for limited packet capture data
- Trying to find media services in flow data is harder
  - Shape isn't present
  - Sampled data means transfer rate has to be extrapolated

## Media Profiling - Pandora

- Found a number of flows from a source 107.133.229.221 serving using port 58.190.251.249.
- All flows were around 2MB to 5MB.
- We then saw conversations that looked like…

# Media Profiling - Pandora

| Source | Destitination | Source Port | Destination Port | Start | End | Seconds | Size in MB |
|---|---|---|---|---|---|---|---|
| 107.133.229.221 | 58.190.251.249 | 554 | 2185 | 5:10:06 AM | 5:10:19 AM | 00:13.0 | 4.43 |
| 107.133.229.221 | 58.190.251.249 | 554 | 2305 | 5:15:38 AM | 5:16:12 AM | 00:34.0 | 3.15 |
| 107.133.229.221 | 58.190.251.249 | 554 | 2333 | 5:16:29 AM | 5:16:37 AM | 00:08.0 | 2.03 |

- The user could have listened to a song at 5:10, listened to some of the next song and then vetoed it, then started a new song.

# Radio Paradise

| sIP | dIP | sPort | dPort | min_sTime | max_eTime | Duration | KBits/s |
|---|---|---|---|---|---|---|---|
| 112.7.36.105 | 119.219.7.30 | 554 | 1378 | 9/9/08 3:32 | 9/9/08 5:51 | 8372 | 80 |
| 112.7.36.105 | 118.136.22.19 | 554 | 4231 | 9/9/08 4:54 | 9/9/08 5:53 | 3547 | 79 |
| 112.7.36.105 | 118.136.33.37 | 554 | 1269 | 9/9/08 5:29 | 9/9/08 5:53 | 1446 | 71 |
| 112.7.36.105 | 58.190.66.237 | 554 | 1045 | 9/9/08 2:16 | 9/9/08 4:35 | 8328 | 93 |
| 112.7.36.105 | 58.190.66.237 | 554 | 1052 | 9/9/08 4:36 | 9/9/08 5:53 | 4645 | 79 |
| 112.7.36.105 | 97.224.22.148 | 554 | 50331 | 9/9/08 1:57 | 9/9/08 5:53 | 14190 | 87 |
| 112.7.36.105 | 97.224.55.150 | 554 | 1312 | 9/9/08 3:04 | 9/9/08 5:53 | 10142 | 82 |
| 112.7.36.105 | 118.136.32.217 | 554 | 4048 | 9/9/08 2:22 | 9/9/08 5:53 | 12676 | 85 |
| 112.7.36.105 | 118.137.227.20 | 554 | 1507 | 9/9/08 2:29 | 9/9/08 5:53 | 12202 | 89 |
| 112.7.36.105 | 118.138.62.206 | 554 | 4611 | 9/9/08 5:06 | 9/9/08 5:53 | 2820 | 71 |
| 112.7.36.105 | 119.219.160.39 | 554 | 2258 | 9/9/08 4:53 | 9/9/08 5:53 | 3640 | 75 |
| 112.7.36.105 | 58.190.231.127 | 554 | 49190 | 9/9/08 1:42 | 9/9/08 2:45 | 3757 | 106 |
| 112.7.36.105 | 118.136.190.161 | 554 | 1203 | 9/9/08 2:37 | 9/9/08 5:53 | 11794 | 86 |
| 112.7.36.105 | 118.138.207.155 | 554 | 2548 | 9/9/08 3:51 | 9/9/08 5:53 | 7348 | 86 |
| 112.7.36.105 | 119.219.124.133 | 554 | 1651 | 9/9/08 0:43 | 9/9/08 1:18 | 2111 | 98 |

Lots of ~100Kbps connections on Port 554 – durations of a few hours!

# Another Video Protocol?

| sIP | dIP | sPort | dPort | min_sTime | max_eTime | Duration | KBits/s |
|---|---|---|---|---|---|---|---|
| 57.49.241.85 | 97.224.79.23 | 554 | 57200 | 9/9/08 4:29 | 9/9/08 5:00 | 1836 | 488 |
| 57.49.241.85 | 118.136.163.193 | 554 | 1220 | 9/9/08 4:34 | 9/9/08 5:02 | 1688 | 482 |
| 57.49.241.86 | 119.219.244.47 | 554 | 2566 | 9/9/08 4:58 | 9/9/08 5:02 | 261 | 470 |
| 57.49.241.86 | 119.219.182.229 | 554 | 43715 | 9/9/08 4:34 | 9/9/08 5:01 | 1641 | 488 |
| 57.49.241.87 | 118.136.36.212 | 554 | 1329 | 9/9/08 4:35 | 9/9/08 5:02 | 1626 | 477 |
| 57.49.241.88 | 118.201.77.155 | 554 | 53656 | 9/9/08 4:51 | 9/9/08 4:57 | 363 | 482 |
| 57.49.241.88 | 97.224.215.251 | 554 | 2257 | 9/9/08 4:39 | 9/9/08 4:51 | 722 | 439 |
| 57.49.241.88 | 118.136.165.221 | 554 | 50750 | 9/9/08 5:08 | 9/9/08 5:17 | 537 | 415 |
| 57.49.241.88 | 119.219.246.224 | 554 | 53857 | 9/9/08 4:51 | 9/9/08 5:21 | 1791 | 448 |
| 57.49.241.89 | 119.219.77.51 | 554 | 1431 | 9/9/08 4:46 | 9/9/08 4:46 | 17 | 298 |
| 57.49.241.89 | 97.225.234.153 | 554 | 2106 | 9/9/08 4:35 | 9/9/08 4:36 | 84 | 443 |

# Conclusion

- Media services have distinct patterns
- Not standardized – so many different profiles
- In flow data, hard to make accurate predictions
  - Better analysis done at the packet level

# Questions?