Transport Layer Identification of P2P Traffic

by T. Karagiannis, A. Broido, M. Faloutsos, and k. claffy

Hadrien Croubois

Computer Science Department of the ENS de Lyon

May 14, 2012

ENS DE LYON

ENS DE LYON

ENS DE LYON

ENS DE LYON

- Data captured at an OC-48 link of a Tier 1 US ISP connection (2,48 Gbits/s)

- Data captured at an OC-48 link of a Tier 1 US ISP connection (2,48 Gbits/s)

ENS DE LYON

- Data captured at an OC-48 link of a Tier 1 US ISP connection (2,48 Gbits/s)

- 4 datasets, from May 2003 to April 2004 (60-122 minutes each)

ENS DE LYON

Datasets description

D09 - D10 : 44 bytes for each packet

Datasets description

D09 - D10 : 44 bytes for each packet

- IP & TCP/UDP headers

ENS DE LYON

Datasets description

D09 - D10 : 44 bytes for each packet

- IP & TCP/UDP headers
- 4 bytes of payload

ENS DE LYON

Datasets description

D09 - D10 : 44 bytes for each packet

- IP & TCP/UDP headers
- 4 bytes of payload

D11 - D13 : 58 bytes for each packet

ENS DE LYON

Datasets description

D09 - D10 : 44 bytes for each packet

- IP & TCP/UDP headers
- 4 bytes of payload

D11 - D13 : 58 bytes for each packet

- 16 bytes of TCP/UDP payload

ENS DE LYON

ENS DE LYON

Method Description

- Identification of P2P traffic based on characteristic bit string in packet payload.

Method Description

- Identification of P2P traffic based on characteristic bit string in packet payload.

P2P protocol	*String*	*Trans. prot.*	*Def. ports*
eDonkey2000	0xe319010000 0xc53f010000	TCP/UDP	4661-4665
Fasttrack	"Get /.hash" 0x270000002980	TCP UDP	1214
BitTorrent	"0x13Bit"	TCP	6881-6889
Gnutella	"GNUT", "GIV" "GND"	TCP UDP	6346-6347

ENS DE LYON

M1 : Check source/destination port with table

M1 : Check source/destination port with table

- Port matches → Flow tagged as P2P

ENS DE LYON

M1 : Check source/destination port with table

- Port matches → Flow tagged as P2P

M2 : Check the payload of each packet with table

M1 : Check source/destination port with table

- Port matches → Flow tagged as P2P

M2 : Check the payload of each packet with table

- String matches → Flow tagged as P2P

ENS DE LYON

M1 : Check source/destination port with table

- Port matches → Flow tagged as P2P

M2 : Check the payload of each packet with table

- String matches → Flow tagged as P2P
- No packet matches → Flow tagged as non-P2P

ENS DE LYON

M1 : Check source/destination port with table

- Port matches → Flow tagged as P2P

M2 : Check the payload of each packet with table

- String matches → Flow tagged as P2P
- No packet matches → Flow tagged as non-P2P

M3 : For P2P flow identified at step M2, record sources & destination IP

ENS DE LYON

M1 : Check source/destination port with table

- Port matches → Flow tagged as P2P

M2 : Check the payload of each packet with table

- String matches → Flow tagged as P2P
- No packet matches → Flow tagged as non-P2P

M3 : For P2P flow identified at step M2, record sources & destination IP

- For all non P2P flows that contain one of these IP
  → Flow tagged as possible-P2P

ENS DE LYON

M1 : Check source/destination port with table

- Port matches → Flow tagged as P2P

M2 : Check the payload of each packet with table

- String matches → Flow tagged as P2P
- No packet matches → Flow tagged as non-P2P

M3 : For P2P flow identified at step M2, record sources & destination IP

- For all non P2P flows that contain one of these IP
  → Flow tagged as possible-P2P

To minimize false positives, FTP, SSL, DNS & online gaming flows are excluded from M3

ENS DE LYON

Limitations

- HTTP requests : P2P protocols using HTTP requests are not identified

Limitations

- HTTP requests : P2P protocols using HTTP requests are not identified
- Encryption : encrypted payload is not identified

ENS DE LYON

Limitations

- HTTP requests : P2P protocols using HTTP requests are not identified
- Encryption : encrypted payload is not identified
- Other P2P protocols : unreferenced P2P protocols are not identified

ENS DE LYON

Limitations

- HTTP requests : P2P protocols using HTTP requests are not identified
- Encryption : encrypted payload is not identified
- Other P2P protocols : unreferenced P2P protocols are not identified
- Unidirectional trace : acknowledgement stream of a P2P download is not always visible because of asymmetric routing

ENS DE LYON

ENS DE LYON

- The non payload method only examines packet headers to detect P2P flow.

- The non payload method only examines packet headers to detect P2P flow.
- As only {IP, port} pairs are the only available, two heuristics, based on the observation of P2P connection patterns, are used

ENS DE LYON

TCP/UDP IP pairs heuristic

- Most P2P protocols use both TCP and UDP protocols

ENS DE LYON

TCP/UDP IP pairs heuristic

- Most P2P protocols use both TCP and UDP protocols
- Other applications using both TCP and UDP protocols are rare and use specific ports

ENS DE LYON

TCP/UDP IP pairs heuristic

- Most P2P protocols use both TCP and UDP protocols
- Other applications using both TCP and UDP protocols are rare and use specific ports

TCP/UDP IP pairs heuristic

{IP,port} using both TCP and UDP protocols (whose ports are not in the exclude list) are considered as P2P traffic

ENS DE LYON

Excluded ports for TCP/UDP IP pairs heuristic

Ports	Applications
135,137,139,445	NETBIOS
53	DNS
123	NTP
500	ISAKMP
554,7070,1755,6970,5000,5001	streaming
7000,7514,6667	IRC
3531	p2pnetworking.exe

ENS DE LYON

{IP,port} pairs heuristic

{IP,port} pairs heuristic

- IPs for which the number of distinct connected IPs is equal to the number of distinct connected ports are considered P2P hosts
- IPs for which the difference between connected IPs and ports is large (e.g., larger than 10) are considered non P2P hosts

ENS DE LYON

False positives

- Mail

False positives

- Mail
- DNS

ENS DE LYON

False positives

- Mail
- DNS
- Gaming

ENS DE LYON

False positives

- Mail
- DNS
- Gaming
- Malware

ENS DE LYON

False positives

- Mail
- DNS
- Gaming
- Malware
- Other heuristics (One-packet pairs, MSN messenger server . . .)

ENS DE LYON

ENS DE LYON

Conclusion

- Easy to understand, efficient method

ENS DE LYON

Conclusion

- Easy to understand, efficient method
- General method (not specific to some P2P protocols, unaffected by encryption)

ENS DE LYON

Conclusion

- Easy to understand, efficient method
- General method (not specific to some P2P protocols, unaffected by encryption)
- Doesn't need to look at payload

ENS DE LYON

- Any questions ?

ENS DE LYON