SECURITY MONITORING OF HTTP TRAFFIC USING EXTENDED FLOWS

Thursday 27th August, 2015

Martin Husák
Petr Velan
Jan Vykopal

CSIRT-MU

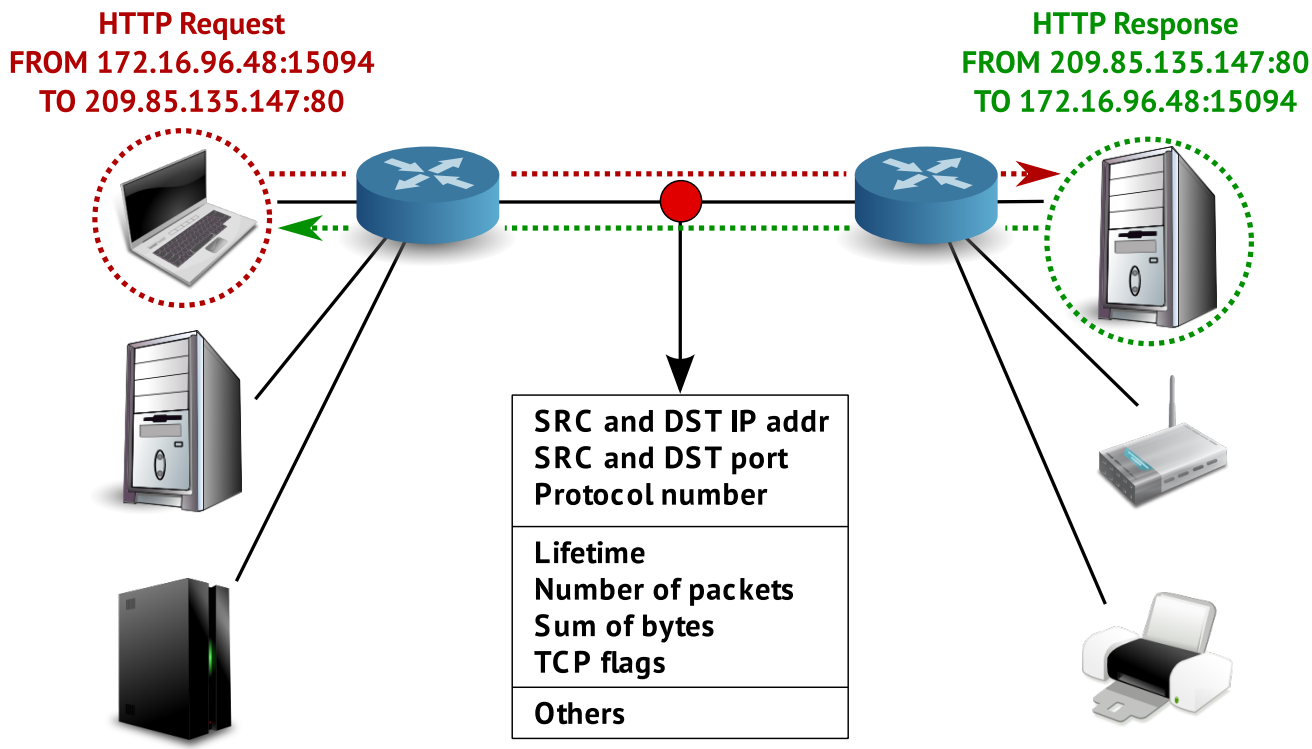
Introduction

- HTTP is the new IP and we want keep an eye on it.
- Large-scale monitoring of HTTP traffic was problematic:
  - Traditional flow-based monitoring processes only L3/L4 headers.
  - DPI is not scalable for large and high-speed networks.
- Extended flows combine the benefits of both methods.
- Can we use large-scale HTTP monitoring for security purposes?
- What types of incidents can we detect using extended flows?

Flow Monitoring

- Passive method of network monitoring.
- Suitable for large-scale and high-speed networks.
- Only the L3/L4 headers are processed.
- Aggregation of network traffic to flows.
- Network flow is a series of packets sharing 5-tuple of elements:
  - L3 protocol, source IP, destination IP, source port, destination port.

Flow Monitoring



CSIRT-MU

Extended Flow Monitoring

- Extension of traditional flow monitoring.
- Modules parse additional information from packets.
- Additional data are stored along the network flow.
- Modules are optimized to parse specific protocol/data.
- Overhead is acceptable, even for monitoring 10 Gbps links.

Research Questions

Question I.

What classes of HTTP traffic relevant to security can be observed at network level and what is their impact on attack detection?

Question II.

What is the added value of extended flow compared to traditional flow monitoring from a security point of view?

Measurement Tools and Environment

- FlowMon probes deployed in campus network of Masaryk University (/16).
- 10 Gbps links, 40,000 users, and 15,000 active IPs per day.
- NetFlow and IPFIX export protocols.
- Extension modules for parsing HTTP headers.
- Over 10 G network flows containing over 1 G HTTP requests were processed.

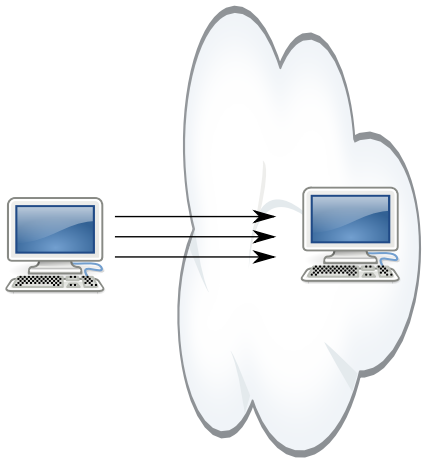
Data Elements

- Key flow elements:
  - L3Proto, srcIP, dstIP, L4Proto, srcPort, dstPort.
- Additional elements:
  - timeStart, timeEnd, packets, octets, TCPflags, ToS, srcAS, dstAS.
- HTTP elements:
  - hostname, path, userAgent, requestMethod, referrer.
  - responseCode, contentType.

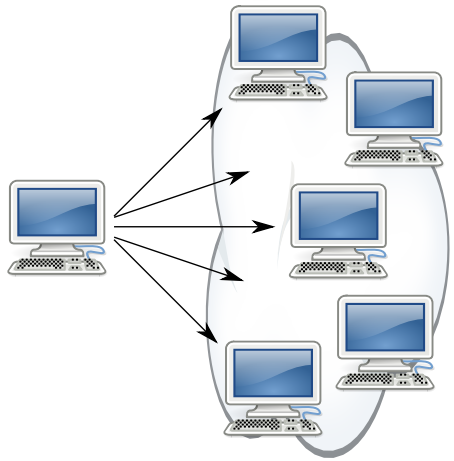
Results

Traffic of interest was found in the three classes:

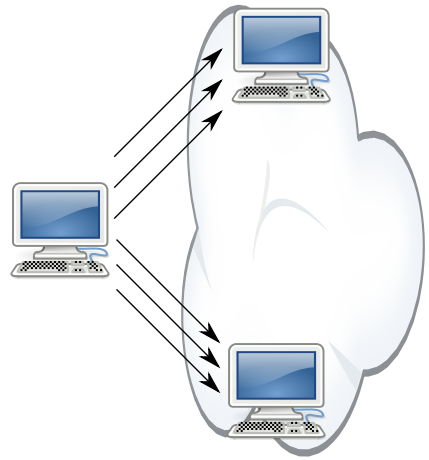
I. Repeated request on a single host.
II. Similar requests on many hosts.
III. Multiple varying requests on multiple hosts.



Class I



Class II



Class III

Class I: Repeated Requests

Guest	Host	HTTP Path	#Flows
G1	H1	/wp-**login**.php	46,031
G2	H2	/**admin**istrator/index.php	27,965
G3	H2	/**admin**istrator/index.php	27,798
G4	H3	/wp-**login**.php	25,316
G5	H4	/pub/linux/slax/Slax-7.x/7.0.8/slax-Chinese-Simplified-7.0.8-i486.iso	5,921
G6	H5	/**proxy**/lib**proxy**.pac	5,036
G7	H6	/node/	4,286
G8	H4	/pub/linux/slax/Slax-7.x/7.0.8/slax-English-US-7.0.8-i486.zip	4,170
G9	H7	/wp-**login**.php	3,632
G10	H7	/polit/wp-**login**.php	3,632

Brute-forcing and proxy servers

Two interesting subclasses were identified:

- Brute-force password attacks.
- Clients connecting to proxy servers.

Both subclasses can be recognized by repeating patterns in URLs.

Subclass	Path regular expression	Portion [%]
Proxy		**49.4**
	.*libproxy.pac	45.0
	.*sviproxy.pac	4.3
	.*proxy.php	0.1
Brute-force		**10.6**
	.*admin.*	6.7
	.*login.*	3.9
Others		**40.0**

Class II: Similar requests on many hosts

Guest	HTTP Path	#Hosts	%
G1	/myadmin/scripts/setup.php	497	100
G1	/pma/scripts/setup.php	497	100
G1	/wootwoot.at.blackhats.romanian.anti-sec:)	497	100
G1	/phpmyadmin/scripts/setup.php	495	99
G1	/phpMyAdmin/scripts/setup.php	494	99
G1	/MyAdmin/scripts/setup.php	491	99
G2	/manager/html	118	24

HTTP Scanners

- Hosts appearing in Class II.
- HTTP scanner requests the same URL from more hosts.
- Typically preceded by or accompanying TCP SYN scan.
  - Lower number of flows is needed to detect a HTTP scan.
- The adversaries are searching for popular vulnerable resources, e.g., older versions of phpMyAdmin.
- Simultaneous search for more resources is common.

Class III: Varying requests on multiple hosts

Guest	Domain Name	#Hosts
207.46.13.62	msnbot-207-46-13-62.search.msn.com	7
157.55.39.107	msnbot-157-55-39-107.search.msn.com	6
137.110.244.137	bnserver2.sdsc.edu	4
157.55.39.156	msnbot-157-55-39-6.search.msn.com	4
157.55.39.6	msnbot-157-55-39-156.search.msn.com	4
37.187.28.19	z3.sentione.com	4
137.110.244.139	integromedb-crawler.integromedb.org	3
5.135.154.106	nks02.sentione.com	3
5.135.154.98	nks03.sentione.com	3
77.75.73.32	fulltextrobot-77-75-73-32.seznam.cz	3
77.75.77.17	fulltextrobot-77-75-77-17.seznam.cz	3

Web crawlers

- Web crawlers are mostly legitimate and welcome in the network.
- Two reasons to include them in the analysis:
  - Malicious crawlers, e.g., e-mail harvesters discovering spam recipients.
  - The large number of flows they generate.
- Legitimate crawlers can be identified by reverse DNS records or well-known User-Agent in HTTP field.
- Lack of such data indicates suspicious crawler.
- All detection methods have to deal with false positive alerts.
- Identification of legitimate crawler can reduce number of FPs.

Conclusion

- Extended flows enable large-scale analysis of HTTP traffic.
- Traffic of interest was found in three classes:
  - Repeated requests - brute-force password attack or proxy server.
  - HTTP scanning.
  - Activity of web crawlers.
- Straighforward implementation of detection methods.
  - Lower thresholds are needed, e.g., for HTTP scan detection.
  - Clearer evidence of malicious intent.
- Not limited to aggregation-based methods.
  - Detection of accesses to a phishing website.
  - Communication with suspicious domains.

THANK YOU FOR YOUR ATTENTION!

muni.cz/csirt

@csirtmu

Martin Husák

husakm@ics.muni.cz

UNIVERSITAS MASARYKIANA BRUNENSIS

ÚVT

CSIRT-MU