Leaving no stone unturned - in search of HTTP malware distinctive features

Piotr Białczak
CERT Polska-NASK/Warsaw University of Technology

Toulouse, 7 December 2018

Botconf 2018

Warsaw University of Technology

Bio

Piotr Białczak
Researcher at CERT Polska/NASK
PhD student at Warsaw University of Technology
Main research areas:

- malware's network artifacts,
- sandboxing.

@bialczakp
piotr.bialczak@cert.pl

- goal: identify feature which distinguish malware and browser HTTP requests,

- goal: identify feature which distinguish malware and browser HTTP requests,
- analyzed network traffic of popular browsers and Windows malware, grouped into categories,

- goal: identify feature which distinguish malware and browser HTTP requests,
- analyzed network traffic of popular browsers and Windows malware, grouped into categories,
- features taken from other work or from research experience

Aim of the presentation

- give insight which features can be chosen to distinguish between malware and browser traffic,

Aim of the presentation

- give insight which features can be chosen to distinguish between malware and browser traffic,
- show which malware families manifest those differences,

Aim of the presentation

- give insight which features can be chosen to distinguish between malware and browser traffic,
- show which malware families manifest those differences,
- present interesting examples of HTTP anomalies in malware and browser traffic

Aim of the presentation

- give insight which features can be chosen to distinguish between malware and browser traffic,
- show which malware families manifest those differences,
- present interesting examples of HTTP anomalies in malware and browser traffic
- however, I **do not** present a detection system

Agenda

Introduction

Network traffic

Features: analyses and analyzers

Results

Summary

Introduction

HTTP request - an example

```
GET / HTTP/1.1

Host: cert.pl
User-Agent: Mozilla/5.0 (Windows NT 10.0; Win64; x64; rv:63.0)
            Gecko/20100101 Firefox/63.0
Accept: text/html,application/xhtml+xml,
        application/xml;q=0.9,*/*;q=0.8
Accept-Language: pl,en-US;q=0.7,en;q=0.3
Accept-Encoding: gzip, deflate
Connection: keep-alive
Upgrade-Insecure-Requests: 1
```

Previous work

- HTTP Header Hunter - Looking for malicious behavior into your HTTP header traffic - Rodrigo Montoro, SecTor 2011

Previous work

- HTTP Header Hunter - Looking for malicious behavior into your HTTP header traffic - Rodrigo Montoro, SecTor 2011
- HTTP header heuristics for malware detection, Tobias Lewis, SANS Institute InfoSec Reading Room

Previous work

- HTTP Header Hunter - Looking for malicious behavior into your HTTP header traffic - Rodrigo Montoro, SecTor 2011
- HTTP header heuristics for malware detection, Tobias Lewis, SANS Institute InfoSec Reading Room
- Some academic papers

Limitations of previous work

- Small data sources

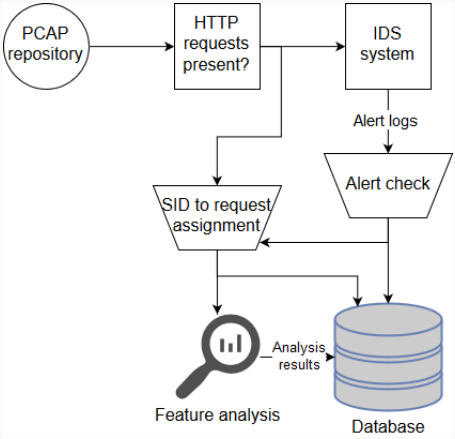
Limitations of previous work

- Small data sources
- Lack of some features

Limitations of previous work

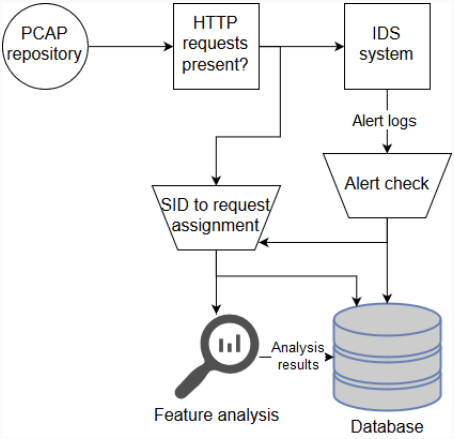
- Small data sources
- Lack of some features
- No general analysis

Analysis system overview



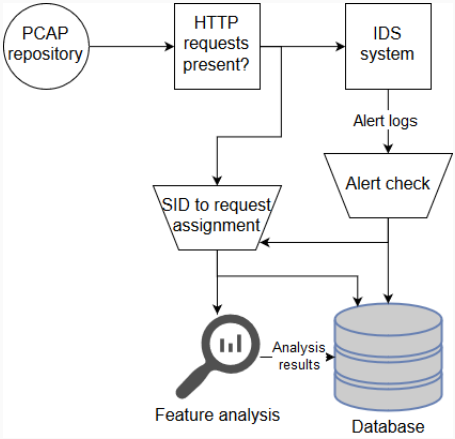
- Finding PCAPs with HTTP traffic

Analysis system overview



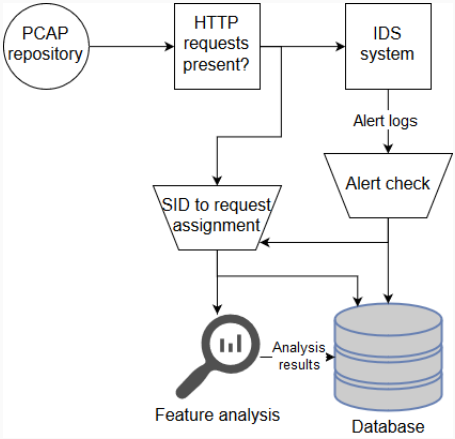
- Finding PCAPs with HTTP traffic
- Feeding them to IDS system (Snort + ET Pro)

Analysis system overview



- Finding PCAPs with HTTP traffic
- Feeding them to IDS system (Snort + ET Pro)
- Assigning SIDs to requests

Analysis system overview



- Finding PCAPs with HTTP traffic
- Feeding them to IDS system (Snort + ET Pro)
- Assigning SIDs to requests
- Performing analyzes

Network traffic

Data sources

- Malware:

Data sources

- Malware:
  - CERT Polska’a sandbox system

Data sources

- Malware:
  - CERT Polska'a sandbox system
  - Malware Capture Facility Project (http://mcfp.felk.cvut.cz/)

Data sources

- Malware:
  - CERT Polska'a sandbox system
  - Malware Capture Facility Project (http://mcfp.felk.cvut.cz/)
- Browser:

Data sources

- Malware:
  - CERT Polska'a sandbox system
  - Malware Capture Facility Project (http://mcfp.felk.cvut.cz/)
- Browser:
  - Connecting to Alexa's global top 500 websites (using Selenium)

Data sources

- Malware:
  - CERT Polska'a sandbox system
  - Malware Capture Facility Project (http://mcfp.felk.cvut.cz/)
- Browser:
  - Connecting to Alexa's global top 500 websites (using Selenium)
  - Popular desktop browsers: Microsoft Edge and Internet Explorer, Mozilla Firefox (also with Flash Player), Google Chrome

Data sources

- Malware:
  - CERT Polska'a sandbox system
  - Malware Capture Facility Project (http://mcfp.felk.cvut.cz/)
- Browser:
  - Connecting to Alexa's global top 500 websites (using Selenium)
  - Popular desktop browsers: Microsoft Edge and Internet Explorer, Mozilla Firefox (also with Flash Player), Google Chrome
  - And OSes: Windows 7, 8.1, 10

Basic information about malware PCAP files

Feature	Number
Number of pcaps in repository	36385
Number of reported IDS alerts	2559123
Number of reported IDS alerts assigned to requests	643921
Number of unique alerted IDS rules	642

Number of HTTP requests in browser data set

Browser	Number of requests
Edge - Win 10	17912
Chrome - Win 7	30621
Firefox + Flash Player - Win 7	18705
Firefox - Win 7	28178
IE - Win 7	30799
Chrome - Win 8.1	23967
Firefox - Win 8.1	18153
IE - Win 8.1	20248

Top 40 malware families

Locky	Emotet	Nymaim	H1N1
Zbot	AZORult	Necurs	Gozi
Ursnif	Loki	Graftor	Cryxos
Dreambot	Kronos	KOVTER	ColorFish
Pony	Tinba	ISFB	Banload
Nemucod	Dridex	FormBook	Adylkuzz
SmokeLoader	Upatre	Betabot	XnxxAgent
DirtJumper	Kelihos.F	Zeroaccess	Wizzcaster
Andromeda	AlphaCrypt	PadCrypt	TrickBot/Loader
Chthonic	QuantLoader	MegalodonHTTP	SpyEyes

Top 40 malware families

Locky	Emotet	Nymaim	H1N1
Zbot	AZORult	Necurs	Gozi
Ursnif	Loki	Graftor	Cryxos
Dreambot	Kronos	KOVTER	ColorFish
Pony	Tinba	ISFB	Banload
Nemucod	Dridex	FormBook	Adylkuzz
SmokeLoader	Upatre	Betabot	XnxxAgent
DirtJumper	Kelihos.F	Zeroaccess	Wizzcaster
Andromeda	AlphaCrypt	PadCrypt	TrickBot/Loader
Chthonic	QuantLoader	MegalodonHTTP	SpyEyes

- Overall 172 malware families

Top 40 malware families

Locky	Emotet	Nymaim	H1N1
Zbot	AZORult	Necurs	Gozi
Ursnif	Loki	Graftor	Cryxos
Dreambot	Kronos	KOVTER	ColorFish
Pony	Tinba	ISFB	Banload
Nemucod	Dridex	FormBook	Adylkuzz
SmokeLoader	Upatre	Betabot	XnxxAgent
DirtJumper	Kelihos.F	Zeroaccess	Wizzcaster
Andromeda	AlphaCrypt	PadCrypt	TrickBot/Loader
Chthonic	QuantLoader	MegalodonHTTP	SpyEyes

- Overall 172 malware families
- However 19% of requests of unknown malware

Features: analyses and analyzers

Reasoning behind analyses

- check for errors caused by poor quality of software development,

Reasoning behind analyses

- check for errors caused by poor quality of software development,
- search for features inherent of malicious operations, e.g. sending obfuscated data,

Reasoning behind analyses

- check for errors caused by poor quality of software development,
- search for features inherent of malicious operations, e.g. sending obfuscated data,
- identify features which reflect difference in data exchange

Header analyzes

Header features consist of analyses checking:

- occurrence frequency of headers,

Header analyzes

Header features consist of analyses checking:

- occurrence frequency of headers,
- misspellings in their name,

Header analyzes

Header features consist of analyses checking:

- occurrence frequency of headers,
- misspellings in their name,
- lack of headers,

Header analyzes

Header features consist of analyses checking:

- occurrence frequency of headers,
- misspellings in their name,
- lack of headers,
- protocol version,

Header analyzes

Header features consist of analyses checking:

- occurrence frequency of headers,
- misspellings in their name,
- lack of headers,
- protocol version,
- destination port,

Header analyzes

Header features consist of analyses checking:

- occurrence frequency of headers,
- misspellings in their name,
- lack of headers,
- protocol version,
- destination port,
- number of headers.

Header values analyzes

Header values features are analyzed in order to find unusual values. The features includes:

- value of Accept-Language, Accept-Encoding and Connection headers,

Header values analyzes

Header values features are analyzed in order to find unusual values. The features includes:

- value of Accept-Language, Accept-Encoding and Connection headers,
- value type of the Host header (IP, domain, other?)

Header values analyzes

Header values features are analyzed in order to find unusual values. The features includes:

- value of Accept-Language, Accept-Encoding and Connection headers,
- value type of the Host header (IP, domain, other?)
- unusual values in the User-Agent header,

Header values analyzes

Header values features are analyzed in order to find unusual values. The features includes:

- value of Accept-Language, Accept-Encoding and Connection headers,
- value type of the Host header (IP, domain, other?)
- unusual values in the User-Agent header,
- presence of non-ASCII characters,

Header values analyzes

Header values features are analyzed in order to find unusual values. The features includes:

- value of Accept-Language, Accept-Encoding and Connection headers,
- value type of the Host header (IP, domain, other?)
- unusual values in the User-Agent header,
- presence of non-ASCII characters,
- problems with whitespace characters and other (additional/unusual spaces, colons, other chars)

Payload analyzes

In payload we analyzed:

- length,

Payload analyzes

In payload we analyzed:

- length,
- entropy,

Payload analyzes

In payload we analyzed:

- length,
- entropy,
- presence of non-ASCII characters,

Payload analyzes

In payload we analyzed:

- length,
- entropy,
- presence of non-ASCII characters,
- presence of request pipelining.

Results

Features with no results

These features did not show any differences between malware and browser:

- misspellings in header name,

Features with no results

These features did not show any differences between malware and browser:

- misspellings in header name,
- value of Accept-Language, Accept-Encoding and Connection headers,

Features with no results

These features did not show any differences between malware and browser:

- misspellings in header name,
- value of Accept-Language, Accept-Encoding and Connection headers,
- presence of pipelining,

Features with no results

These features did not show any differences between malware and browser:

- misspellings in header name,
- value of Accept-Language, Accept-Encoding and Connection headers,
- presence of pipelining,
- payload length*

Features with limited results

- group of features which did not show significant results,

Features with limited results

- group of features which did not show significant results,
- their numbers are too low or they appear in browser traffic,

Features with limited results

- group of features which did not show significant results,
- their numbers are too low or they appear in browser traffic,
- they are shown as interesting examples

Lack of colon in header line

```
GET /space  HTTP/1.0\n
\n
```

- appeared only in browser traffic

Lack of colon in header line

```
GET /space  HTTP/1.0\n
\n
```

- appeared only in browser traffic
- two requests by IE Windows 7 and two requests by IE Windows 8.1

Lack of colon in header line

```
GET /space  HTTP/1.0\n
\n
```

- appeared only in browser traffic
- two requests by IE Windows 7 and two requests by IE Windows 8.1
- request sent to mpsnare.iesnare.com

Lack of colon in header line

```
GET /space  HTTP/1.0\n
\n
```

- appeared only in browser traffic
- two requests by IE Windows 7 and two requests by IE Windows 8.1
- request sent to mpsnare.iesnare.com
- when visited redirection to https://www.iovation.com/stopfraud/

Lack of colon in header line

```
GET /space  HTTP/1.0\n
\n
```

- appeared only in browser traffic
- two requests by IE Windows 7 and two requests by IE Windows 8.1
- request sent to mpsnare.iesnare.com
- when visited redirection to https://www.iovation.com/stopfraud/
- probably callback of device fingerprinting mechanism

Unpopular whitespace character + space before comma

- checked in search for poor quality code,

Unpopular whitespace character + space before comma

- checked in search for poor quality code,
- not present in browser traffic,

Unpopular whitespace character + space before comma

- checked in search for poor quality code,
- not present in browser traffic,
- in data set only seen in Alphacrypt ransomware

Alphacrypt

```
50 4f 53 54 20 2f 77 70 2d 63 6f 6e 74 65 6e 74   POST /wp-content
2f 70 6c 75 67 69 6e 73 2f 63 6f 6e 74 61 63 74   /plugins/contact
2d 66 6f 72 6d 2d 37 2f 69 6e 63 6c 75 64 65 73   -form-7/includes
2f 6a 73 2f 6a 71 75 65 72 79 2d 75 69 2f 74 68   /js/jquery-ui/th
65 6d 65 73 2f 73 6d 6f 6f 74 68 6e 65 73 73 2f   emes/smoothness/
69 6d 61 67 65 73 2f 62 69 6e 66 69 6c 65 2e 70   images/binfile.p
68 70 20 48 54 54 50 2f 31 2e 31 0d 0a 41 63 63   hp HTTP/1.1..Acc
65 70 74 3a 20 6a 2c 20 e8 09 4e 02 b8 8d 24 02   ept: j, è.N.¸.$.
80 01 24 02 80 01 24 02 88 01 24 02 88 01 24 02   ..$...$...$...$.
90 01 24 02 90 01 24 02 98 01 24 02 98 01 24 02   ..$...$...$...$.
28 2c 20 2c 20 2c 20 2c 20 2c 20 2c 20 2c 20 2c   (, , , , , , , ,
20 2c 20 2c 20 2c 20 2c 20 2c 20 2c 20 2c 20 2c    , , , , , , , ,
20 2c 20 2c 20 2c 20 2c 20 0d 0a 43 6f 6e 74 65    , , , , ..Conte
```

End of header line other than CRLF

```
GET /empty_flash?e=35&ud=4&up=4&qa=1827&[truncated] HTTP/1.0\n
User-Agent: Mozilla/5.0 (Windows NT 6.1; Trident/7.0; SLCC2;
            .NET CLR 2.0.50727; .NET CLR 3.5.30729; .NET CLR 3.0.30729;
            Media Center PC 6.0; .NET4.0C; rv:11.0) like Gecko\n
Referer: http://www.huffingtonpost.com/\n
Host: pixel.moatads.com\n
Connection: keep-alive\n
Accept: */*\n
\n
```

- present only in browser traffic (less than 0,1%),

End of header line other than CRLF

```
GET /empty_flash?e=35&ud=4&up=4&qa=1827&[truncated] HTTP/1.0\n
User-Agent: Mozilla/5.0 (Windows NT 6.1; Trident/7.0; SLCC2;
            .NET CLR 2.0.50727; .NET CLR 3.5.30729; .NET CLR 3.0.30729;
            Media Center PC 6.0; .NET4.0C; rv:11.0) like Gecko\n
Referer: http://www.huffingtonpost.com/\n
Host: pixel.moatads.com\n
Connection: keep-alive\n
Accept: */*\n
\n
```

- present only in browser traffic (less than 0,1%),
- again IE on Windows 7 and Windows 8.1

End of header line other than CRLF

```
GET /empty_flash?e=35&ud=4&up=4&qa=1827&[truncated] HTTP/1.0\n
User-Agent: Mozilla/5.0 (Windows NT 6.1; Trident/7.0; SLCC2;
            .NET CLR 2.0.50727; .NET CLR 3.5.30729; .NET CLR 3.0.30729;
            Media Center PC 6.0; .NET4.0C; rv:11.0) like Gecko\n
Referer: http://www.huffingtonpost.com/\n
Host: pixel.moatads.com\n
Connection: keep-alive\n
Accept: */*\n
\n
```

- present only in browser traffic (less than 0,1%),
- again IE on Windows 7 and Windows 8.1
- one example already shown

Non-ASCII character in header line

```
61 73 69 26 72 72 6f 32 3d 41 73 69 61 26 67 6d   asi&rro2=Asia&gm
30 3d 2d 31 26 6c 63 33 3d 32 37 35 31 37 34 26   0=-1&lc3=275174&
63 63 33 3d 50 4c 26 63 74 33 3d 47 64 61 c5 84   cc3=PL&ct3=Gda..
73 6b 26 63 6e 33 3d 50 6f 6c 61 6e 64 26 72 67   sk&cn3=Poland&rg
33 3d 45 55 52 26 73 74 33 3d 50 4d 26 61 6e 33   3=EUR&st3=PM&an3
3d 50 6f 6d 65 72 61 6e 69 61 26 6f 6e 33 3d 47   =Pomerania&on3=G
64 61 c5 84 73 6b 26 75 66 33 3d 45 50 47 52 26   da..sk&uf3=EPGR&
7a 70 33 3d 32 37 35 31 37 34 26 6c 73 33 3d 2d   zp3=275174&ls3=-
31 26 64 70 33 3d 6e 26 75 67 6c 61 74 33 3d 30   1&dp3=n&uglat3=0
26 75 67 6c 6f 6e 33 3d 30 26 6c 63 34 3d 32 37   &uglon3=0&lc4=27
```

Non-ASCII character in header line - browser

```
61 73 69 26 72 72 6f 32 3d 41 73 69 61 26 67 6d   asi&rro2=Asia&gm
30 3d 2d 31 26 6c 63 33 3d 32 37 35 31 37 34 26   0=-1&lc3=275174&
63 63 33 3d 50 4c 26 63 74 33 3d 47 64 61 c5 84   cc3=PL&ct3=Gda..
73 6b 26 63 6e 33 3d 50 6f 6c 61 6e 64 26 72 67   sk&cn3=Poland&rg
33 3d 45 55 52 26 73 74 33 3d 50 4d 26 61 6e 33   3=EUR&st3=PM&an3
3d 50 6f 6d 65 72 61 6e 69 61 26 6f 6e 33 3d 47   =Pomerania&on3=G
64 61 c5 84 73 6b 26 75 66 33 3d 45 50 47 52 26   da..sk&uf3=EPGR&
7a 70 33 3d 32 37 35 31 37 34 26 6c 73 33 3d 2d   zp3=275174&ls3=-
31 26 64 70 33 3d 6e 26 75 67 6c 61 74 33 3d 30   1&dp3=n&uglat3=0
26 75 67 6c 6f 6e 33 3d 30 26 6c 63 34 3d 32 37   &uglon3=0&lc4=27
```

- only seen in Windows 7 browsers: Firefox, IE, Chrome,

Non-ASCII character in header line - browser

```
61 73 69 26 72 72 6f 32 3d 41 73 69 61 26 67 6d   asi&rro2=Asia&gm
30 3d 2d 31 26 6c 63 33 3d 32 37 35 31 37 34 26   0=-1&lc3=275174&
63 63 33 3d 50 4c 26 63 74 33 3d 47 64 61 c5 84   cc3=PL&ct3=Gda..
73 6b 26 63 6e 33 3d 50 6f 6c 61 6e 64 26 72 67   sk&cn3=Poland&rg
33 3d 45 55 52 26 73 74 33 3d 50 4d 26 61 6e 33   3=EUR&st3=PM&an3
3d 50 6f 6d 65 72 61 6e 69 61 26 6f 6e 33 3d 47   =Pomerania&on3=G
64 61 c5 84 73 6b 26 75 66 33 3d 45 50 47 52 26   da..sk&uf3=EPGR&
7a 70 33 3d 32 37 35 31 37 34 26 6c 73 33 3d 2d   zp3=275174&ls3=-
31 26 64 70 33 3d 6e 26 75 67 6c 61 74 33 3d 30   1&dp3=n&uglat3=0
26 75 67 6c 6f 6e 33 3d 30 26 6c 63 34 3d 32 37   &uglon3=0&lc4=27
```

- only seen in Windows 7 browsers: Firefox, IE, Chrome,
- one request in every browser,

Non-ASCII character in header line - browser

```
61 73 69 26 72 72 6f 32 3d 41 73 69 61 26 67 6d   asi&rro2=Asia&gm
30 3d 2d 31 26 6c 63 33 3d 32 37 35 31 37 34 26   0=-1&lc3=275174&
63 63 33 3d 50 4c 26 63 74 33 3d 47 64 61 c5 84   cc3=PL&ct3=Gda..
73 6b 26 63 6e 33 3d 50 6f 6c 61 6e 64 26 72 67   sk&cn3=Poland&rg
33 3d 45 55 52 26 73 74 33 3d 50 4d 26 61 6e 33   3=EUR&st3=PM&an3
3d 50 6f 6d 65 72 61 6e 69 61 26 6f 6e 33 3d 47   =Pomerania&on3=G
64 61 c5 84 73 6b 26 75 66 33 3d 45 50 47 52 26   da..sk&uf3=EPGR&
7a 70 33 3d 32 37 35 31 37 34 26 6c 73 33 3d 2d   zp3=275174&ls3=-
31 26 64 70 33 3d 6e 26 75 67 6c 61 74 33 3d 30   1&dp3=n&uglat3=0
26 75 67 6c 6f 6e 33 3d 30 26 6c 63 34 3d 32 37   &uglon3=0&lc4=27
```

- only seen in Windows 7 browsers: Firefox, IE, Chrome,
- one request in every browser,
- it is a request for weather forecast with city name character *ń* in Polish

Non-ASCII character in header - malware

```
74 3a 20 74 65 78 74 2f 2a 2c 20 51 57 52 73 4e   t: text/*, QWRsN
32 73 72 64 6a 6c 78 55 55 64 44 59 56 70 30 61   2srdjlxUUdDYVp0a
54 42 4d 55 7a 6c 32 63 53 74 7a 59 30 4a 30 64   TBMUzl2cStzY0J0d
58 64 45 4e 44 6c 4f 4d 47 5a 6e 63 55 68 68 65   XdENDlOMGZncUhhe
47 31 42 4d 53 39 6f 53 6c 42 56 65 6a 42 6a 54   G1BMS9oSlBVejBjT
31 41 30 4d 33 4e 4b 52 57 70 46 4e 6d 31 6c 55   1A0M3NKRWpFNm1lU
45 46 58 55 46 70 75 4d 6d 6b 35 4d 6c 70 35 65   EFXUFpuMmk5Mlp5e
45 64 76 4d 46 4e 58 61 6d 6c 44 54 33 53 76 53   EdvMFNXamlDT3SvS
48 52 56 65 6b 6c 68 4d 43 39 45 63 6d 46 49 5a   HRVeklhMC9EcmFIZ
6e 45 35 4c 33 51 35 4d 47 4a 30 64 58 5a 4b 53   nE5L3Q5MGJ0dXZKS
32 67 76 62 47 78 59 51 67 3d 3d 2c 20 31 35 31   2gvbGxYQg==, 151
2e 38 30 2e 38 2e 31 2c 20 5f 5e 5b c3 e8 eb 02   .80.8.1, _^[Ãèë.
0d 0a 43 6f 6e 74 65 6e 74 2d 54 79 70 65 3a 20   ..Content-Type:
```

- an example of Graybird malware request,

Non-ASCII character in header - malware

```
74 3a 20 74 65 78 74 2f 2a 2c 20 51 57 52 73 4e   t: text/*, QWRsN
32 73 72 64 6a 6c 78 55 55 64 44 59 56 70 30 61   2srdjlxUUdDYVp0a
54 42 4d 55 7a 6c 32 63 53 74 7a 59 30 4a 30 64   TBMUzl2cStzY0J0d
58 64 45 4e 44 6c 4f 4d 47 5a 6e 63 55 68 68 65   XdENDlOMGZncUhhe
47 31 42 4d 53 39 6f 53 6c 42 56 65 6a 42 6a 54   G1BMS9oSlBVejBjT
31 41 30 4d 33 4e 4b 52 57 70 46 4e 6d 31 6c 55   1A0M3NKRWpFNm1lU
45 46 58 55 46 70 75 4d 6d 6b 35 4d 6c 70 35 65   EFXUFpuMmk5Mlp5e
45 64 76 4d 46 4e 58 61 6d 6c 44 54 33 53 76 53   EdvMFNXamlDT3SvS
48 52 56 65 6b 6c 68 4d 43 39 45 63 6d 46 49 5a   HRVeklhMC9EcmFIZ
6e 45 35 4c 33 51 35 4d 47 4a 30 64 58 5a 4b 53   nE5L3Q5MGJ0dXZKS
32 67 76 62 47 78 59 51 67 3d 3d 2c 20 31 35 31   2gvbGxYQg==, 151
2e 38 30 2e 38 2e 31 2c 20 5f 5e 5b c3 e8 eb 02   .80.8.1, _^[Ãèë.
0d 0a 43 6f 6e 74 65 6e 74 2d 54 79 70 65 3a 20   ..Content-Type:
```

- an example of Graybird malware request,
- the only other - Alphacrypt (presented before)

Destination port

- most of the traffic on port 80 in both types of applications,

Destination port

- most of the traffic on port 80 in both types of applications,
- in browser: 99,9% of traffic - 80, then 443,

Destination port

- most of the traffic on port 80 in both types of applications,
- in browser: 99,9% of traffic - 80, then 443,
- ports other than 80 present in some families: Andromeda, Dridex, Emotet, Pony, Upatre,

Destination port

- most of the traffic on port 80 in both types of applications,
- in browser: 99,9% of traffic - 80, then 443,
- ports other than 80 present in some families: Andromeda, Dridex, Emotet, Pony, Upatre,
- not clear whether a constant feature or dependent on infrastructure

Destination port

- most of the traffic on port 80 in both types of applications,
- in browser: 99,9% of traffic - 80, then 443,
- ports other than 80 present in some families: Andromeda, Dridex, Emotet, Pony, Upatre,
- not clear whether a constant feature or dependent on infrastructure

Category	Ports
Banker	443, 8080, 7080
Downloader	13404, 13405, 13526, 12267, 5450, 8080, 81
Miner	8888
Other	8080
Ransomware	443, 53717, 40219
Stealer	26123
Trojan	8080, 433

More distinctive features

- further features showed some significant difference between malware and browser traffic,

More distinctive features

- further features showed some significant difference between malware and browser traffic,
- we put more expectations into some of them (*Host* header value),

More distinctive features

- further features showed some significant difference between malware and browser traffic,
- we put more expectations into some of them (*Host* header value),
- but some other proved to be better (POST requests without *Referer* header)

Protocol version

- surprising prevalence of 1.0 protocol version requests in malware,

Protocol version

- surprising prevalence of 1.0 protocol version requests in malware,
- main categories (at least 90% of requests): spambot, clicker, ransomware, miner, adware

Protocol version

- surprising prevalence of 1.0 protocol version requests in malware,
- main categories (at least 90% of requests): spambot, clicker, ransomware, miner, adware
- none in browsers or limited numbers (less than 0,01% in IE Win7+8.1),

Protocol version

- surprising prevalence of 1.0 protocol version requests in malware,
- main categories (at least 90% of requests): spambot, clicker, ransomware, miner, adware
- none in browsers or limited numbers (less than 0,01% in IE Win7+8.1),
- 13% families sent such requests,

Protocol version

- surprising prevalence of 1.0 protocol version requests in malware,
- main categories (at least 90% of requests): spambot, clicker, ransomware, miner, adware
- none in browsers or limited numbers (less than 0,01% in IE Win7+8.1),
- 13% families sent such requests,
- these include: AZORult, Chthonic, Pony, SmokeLoader, Tinba, Neutrino

Non-ASCII characters in payload

- substantial numbers for bankers (63%), downloaders (71%) and trojans (39%),

Non-ASCII characters in payload

- substantial numbers for bankers (63%), downloaders (71%) and trojans (39%),
- single requests in browser traffic,

Non-ASCII characters in payload

- substantial numbers for bankers (63%), downloaders (71%) and trojans (39%),
- single requests in browser traffic,
- in terms of families - 15% of all,

Non-ASCII characters in payload

- substantial numbers for bankers (63%), downloaders (71%) and trojans (39%),
- single requests in browser traffic,
- in terms of families - 15% of all,
- most known: Emotet, Smokeloader, Necurs, Nymaim, Tinba, Panda, Ursnif, Locky

Entropy - what is it?

- informally: measures amount of information carried by data,

Entropy - what is it?

- informally: measures amount of information carried by data,
- higher entropy - more information per character,

Entropy - what is it?

- informally: measures amount of information carried by data,
- higher entropy - more information per character,
- more information - hint for data obfuscation,

Entropy - what is it?

- informally: measures amount of information carried by data,
- higher entropy - more information per character,
- more information - hint for data obfuscation,
- English text - typically 3,5-5,0,

Entropy - what is it?

- informally: measures amount of information carried by data,
- higher entropy - more information per character,
- more information - hint for data obfuscation,
- English text - typically 3,5-5,0,
- browser requests' payload maximum entropy value - 6,13, mean in interval 4,30-4,76,

Entropy - what is it?

- informally: measures amount of information carried by data,
- higher entropy - more information per character,
- more information - hint for data obfuscation,
- English text - typically 3,5-5,0,
- browser requests' payload maximum entropy value - 6,13, mean in interval 4,30-4,76,
- some malware categories well beyond this limit

Entropy in malware

- categories with significantly higher mean entropy than browsers: banker, trojan, clicker, downloader, spambot

Entropy in malware

- categories with significantly higher mean entropy than browsers: banker, trojan, clicker, downloader, spambot
- in terms of families - 13% of all sent request with entropy higher than 6,

Entropy in malware

- categories with significantly higher mean entropy than browsers: banker, trojan, clicker, downloader, spambot
- in terms of families - 13% of all sent request with entropy higher than 6,
- most known: AZORult, Chthonic, Dridex, Emotet, FormBook, Locky, Necurs, Nymaim, Panda

GET request with payload

- not prohibited by the RFC,

GET request with payload

- not prohibited by the RFC,
- not seen in browser traffic,

GET request with payload

- not prohibited by the RFC,
- not seen in browser traffic,
- four families sent such requests,

GET request with payload

- not prohibited by the RFC,
- not seen in browser traffic,
- four families sent such requests,
- Dreambot, Kelihos.F, Ursnif, Zeroaccess

POST without *Referer* header

- requests in browser traffic are in someway linked with this header,

POST without *Referer* header

- requests in browser traffic are in someway linked with this header,
- malware sometimes sends POST requests without previous HTTP exchange,

POST without *Referer* header

- requests in browser traffic are in someway linked with this header,
- malware sometimes sends POST requests without previous HTTP exchange,
- most of the malware categories sent POST requests without *Referer* header

POST without *Referer* header

- requests in browser traffic are in someway linked with this header,
- malware sometimes sends POST requests without previous HTTP exchange,
- most of the malware categories sent POST requests without *Referer* header
- only ransomware and backdoor have significant percentage of requests with this header (more than 20%)

POST without *Referer* header

- requests in browser traffic are in someway linked with this header,
- malware sometimes sends POST requests without previous HTTP exchange,
- most of the malware categories sent POST requests without *Referer* header
- only ransomware and backdoor have significant percentage of requests with this header (more than 20%)
- 44% of families sent such requests

Host header value

- In browsers request most of the values are domains (99,8%)

Host header value

- In browsers request most of the values are domains (99,8%)
- The same applies to malware

Host header value

- In browsers request most of the values are domains (99,8%)
- The same applies to malware
- Some categories use IP or IP + port

Host header value

- In browsers request most of the values are domains (99,8%)
- The same applies to malware
- Some categories use IP or IP + port
- These include: ransomware, miner, spambot

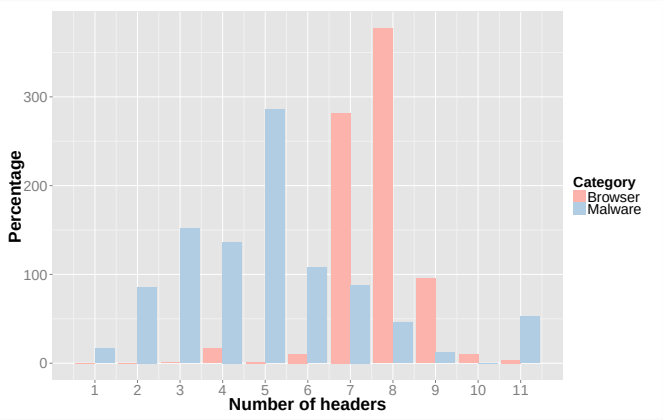
Host header value

- In browsers request most of the values are domains (99,8%)
- The same applies to malware
- Some categories use IP or IP + port
- These include: ransomware, miner, spambot
- 25% of malware families sent request with value other than domain

Host header value

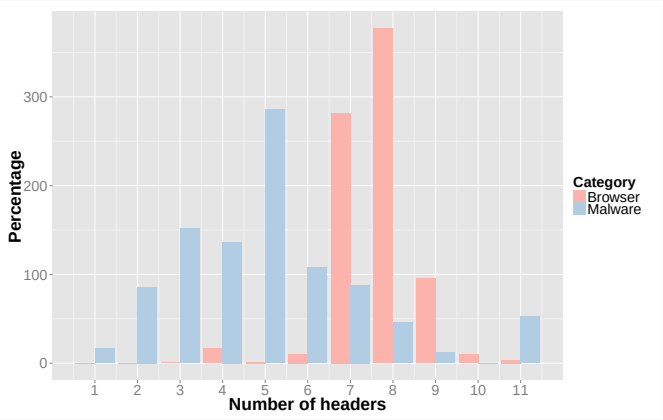
- In browsers request most of the values are domains (99,8%)
- The same applies to malware
- Some categories use IP or IP + port
- These include: ransomware, miner, spambot
- 25% of malware families sent request with value other than domain
- same remark as with destination port: the value type can depend on infrastructure

Number of headers



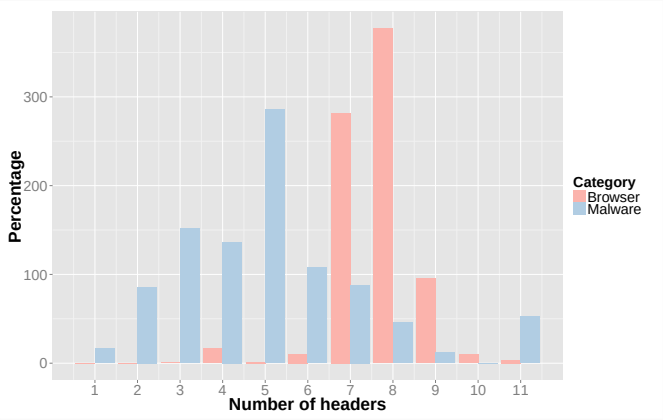
- malware tends to have less headers

Number of headers



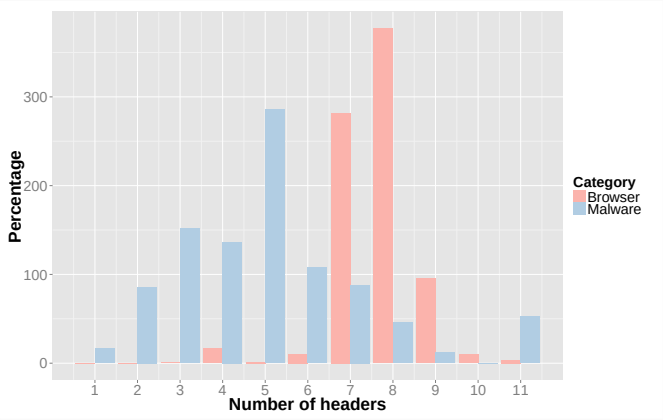
- malware tends to have less headers
- one boundary can be found at 6 headers

Number of headers



- malware tends to have less headers
- one boundary can be found at 6 headers
- noticeable number of requests with less than 3 headers (38% of families)

Number of headers



- malware tends to have less headers
- one boundary can be found at 6 headers
- noticeable number of requests with less than 3 headers (38% of families)
- these include: Dorkbot, Gozi, Formbook, Ursnif, Betabot, Smokeloader, Tinba

User-Agent

- browsers requests tend to have one standard *User-Agent* string

User-Agent

- browsers requests tend to have one standard *User-Agent* string
- IE sometimes changes the value, even to Firefox string

User-Agent

- browsers requests tend to have one standard *User-Agent* string
- IE sometimes changes the value, even to Firefox string
- malware usually carries Internet Explorer or Firefox string

User-Agent

- browsers requests tend to have one standard *User-Agent* string
- IE sometimes changes the value, even to Firefox string
- malware usually carries Internet Explorer or Firefox string
- many malware families sent requests without User-Agent header

Lack of User-Agent header - malware

- 41 families sent request without the header (24% of families)

Lack of User-Agent header - malware

- 41 families sent request without the header (24% of families)
- these are bankers, stealers, miners, ransomware, ...

Lack of User-Agent header - malware

- 41 families sent request without the header (24% of families)
- these are bankers, stealers, miners, ransomware, ...
- e.g. AZORult, Chthonic, Dreambot, Gozi, Locky, Necurs, Sage, Tinba, Ursnif, Quantloader

Lack of User-Agent header - browser

```
GET / HTTP/1.1
Host: socket.dingit.tv:8050
Upgrade: websocket
Connection: Upgrade
Sec-WebSocket-Key: evWImUfQSg18ppTsQZRo9g==
Origin: *
Sec-WebSocket-Version: 13
```

- single request sent by Chrome on Windows 7,

Lack of User-Agent header - browser

```
GET / HTTP/1.1
Host: socket.dingit.tv:8050
Upgrade: websocket
Connection: Upgrade
Sec-WebSocket-Key: evWImUfQSg18ppTsQZRo9g==
Origin: *
Sec-WebSocket-Version: 13
```

- single request sent by Chrome on Windows 7,
- other requests presented before

Examples of User-Agent strings in malware

AudioDrive	pb
AutoHotkey	post_example
AutoIt	python-requests/2.12.4
BotLoader	python-requests/2.18.4
Botnet by Danij	Python-urllib/2.7
BTWebClient/3430(40097)	Python-urllib/3.1
Uploador	Recuva
WinHttpClient	SLIMHTTP/1.1
W1pbbA((	start_page 3.50
Christmas Mystery 5.5.7	TBNotifier
Downloader 22.7	TrickLoader
EMSFRTCBVD	C:Users[user's name]AppDataRoamingv2o5g0Ie5itemp.zip

Whole presentation in one slide

Good to search for suspicious request
Number of headers smaller than 4
Lack of User-Agent string
POST without Referer header
1.0 version of protocol
Non-ASCII characters in payload
High entropy of payload
Not domain in Host header
GET request with payload

Unpopular, but depends on malware family
end of header line other than CRLF
unpopular whitespace character
space before comma
non-ASCII character in header line
destination port

Examples of rules for your sandbox system

To identify suspicious HTTP requests look for:

- GET request with payload,

Examples of rules for your sandbox system

To identify suspicious HTTP requests look for:

- GET request with payload,
- 3 or less headers and standard value of *User-Agent*,

Examples of rules for your sandbox system

To identify suspicious HTTP requests look for:

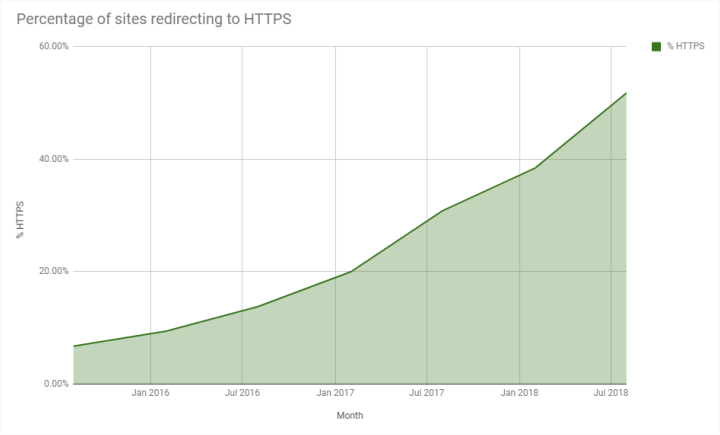
- GET request with payload,
- 3 or less headers and standard value of *User-Agent*,
- POST request without *Referer* header*,

Examples of rules for your sandbox system

To identify suspicious HTTP requests look for:

- GET request with payload,
- 3 or less headers and standard value of *User-Agent*,
- POST request without *Referer* header*,
- lack of *Accept*, *Accept-Encoding*, *Accept-Language* headers

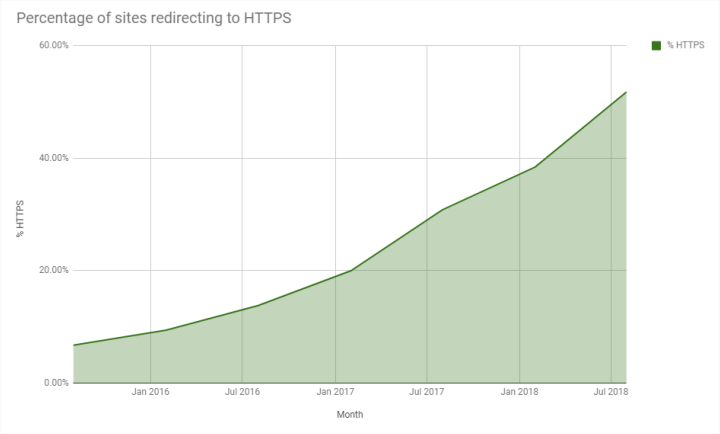
Speculations



Source: https://scotthelme.co.uk/alexa-top-1-million-analysis-august-2018/

- Increasing number of sites using HTTPS

Speculations



Source: https://scotthelme.co.uk/alexa-top-1-million-analysis-august-2018/

- Increasing number of sites using HTTPS
- Usage of HTTP as an outlier/anomaly?

Summary

Limitations

- Level of malware representation by our data sets

Limitations

- Level of malware representation by our data sets
- Problems with labeling - not all requests were labeled

Limitations

- Level of malware representation by our data sets
- Problems with labeling - not all requests were labeled
- False positives in ET rules

Summary

- Some features are relatively good to discern between malware and browser

Summary

- Some features are relatively good to discern between malware and browser
- Anomalies are hard to find and define - some browsers do produce them

Summary

- Some features are relatively good to discern between malware and browser
- Anomalies are hard to find and define - some browsers do produce them
- Presented features can be used to provide basic info about whether request is suspicious

Summary

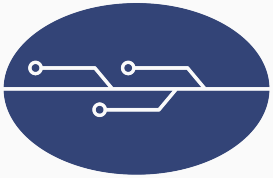
- Some features are relatively good to discern between malware and browser
- Anomalies are hard to find and define - some browsers do produce them
- Presented features can be used to provide basic info about whether request is suspicious
- If feature did not give results, it means that it isn't popular, not that it won't show anything in future or in particular malware families

Interested about details?

- We are preparing a scientific paper

Interested about details?

- We are preparing a scientific paper
- Write to me if interested



SISSDEN

Part of this research has been supported by the European Union Horizon 2020 programme under grant agreement No. 700176 (SISSDEN). The opinions expressed in this presentation are those of the authors and do not necessarily reflect the views of the European Commission.

Leaving no stone unturned - in search of HTTP malware distinctive features

Piotr Białczak
CERT Polska-NASK/Warsaw University of Technology

Toulouse, 7 December 2018

Botconf 2018

Warsaw University of Technology