HTTP Session Identification

Research project 2

Kevin de Kok
Marcus Bakker

30 June 2010

Agenda

- Introduction
- Research question
- Project scope
- Dataset
- Identification methods
- Conclusion
- Future work
- Questions?

Introduction (1)

- What is a HTTP session?

Introduction (2)

- The need to identify HTTP sessions [1]
- Not trivial to identify HTTP sessions
- HTTP is a sessionless protocol
- Request - Response

[1]
T. Kinkhorst and M. van Kleij. Busting the ghost on the web: real time detection of drive-by-infections, June 2009. URL http://www.delaat.net/~cees/sne-2008-2009/p46/report.pdf.

Research question

- How can HTTP sessions be distinguished from each other?

Project scope

- RFC 2616
- The methods to identify a HTTP session will be developed for **web 1.0 (e.g. no Ajax)**
- The HTTP session identification will be executed from a central point in the network (no host-based detection)

Dataset

- Labsite (bookmark)
  - Opened three hyperlinks
- Security.nl (bookmark)
  - Opened three hyperlinks
- 8 HTTP sessions (2 bookmarks + 6 hyperlinks)

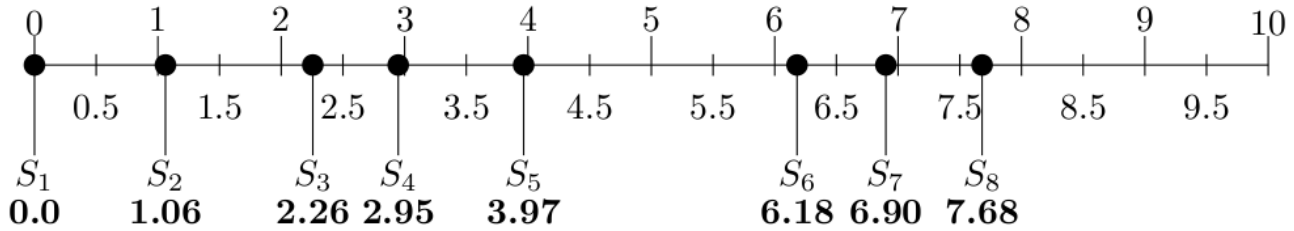


Figure 3: Start time of the HTTP session in seconds.

Identification methods

- Two categories of methods:
  - Start of a HTTP session
  - HTTP message correlation

Start of a HTTP session

- Time between successive fetches
- Hyperlink present at GET header
- No referrer

Time between successive fetches(1)

- 10 – 600ms [2]

Method 2 Time between successive fetches

$AOT = 0.6$ {The Active Off Time is set to 600ms}
for all Rs^n in Rs **do**
  if $(Rs^n.time - Rs^{n-1}.time) > AOT$ **then**
    print A new HTTP session was detected
    $S = \text{storePair}(Rs^n)$
  end if
end for

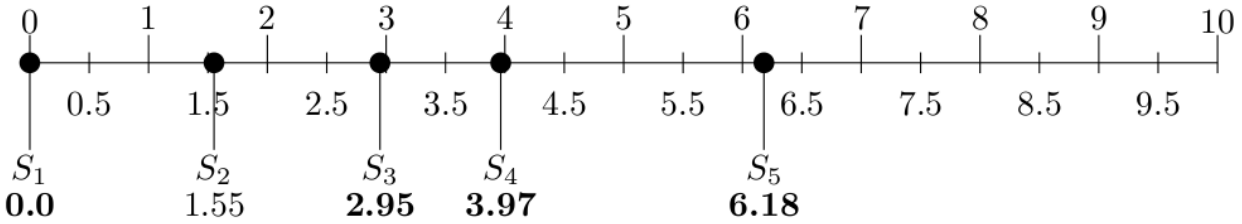
```
# ####  ..............  #### #
# #### New HTTP session: 3 #### #
 Time: 2.948454

 >>> REQUEST >>>
Packet ID:..... 86
Frame ID:..... 358
Stream ID:..... 1
Time:..... 2.948454
SRC:..... 145.100.107.4:45467
DST:..... 145.100.104.30:80
Host:..... bulbasaur.studlab.os3.nl
GET:..... /rp2/new_website/2/index.html
User-Agent:..... Opera/9.80 (X11; Linux i686; U; en-GB) Presto/2.2.15 Version
10.10
Referrer:..... http://bulbasaur.studlab.os3.nl/rp2/new_website/
--------------------------------------------------------------------------------
 <<< RESPONSE <<<
Packet ID:..... 87
Frame ID:..... 359
Stream ID:..... 1
Time:..... 2.94909
SRC:..... 145.100.104.30:80
DST:..... 145.100.107.4:45467
Response Code:.... 200 OK
Date:..... Thu, 17 Jun 2010 19:20:54 GMT
Content-Type:..... text/html
Content-Length:... 191
Body:
     <!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.0 Transitional//EN">
     <html>
     <head>
```

Proof of Concept

[2]
Y. Bhole and A.Popescu. Measurement and analysis of http traffic, December 2005.

Time between successive fetches(2)

- “Slow” browsing (mobile phone?)

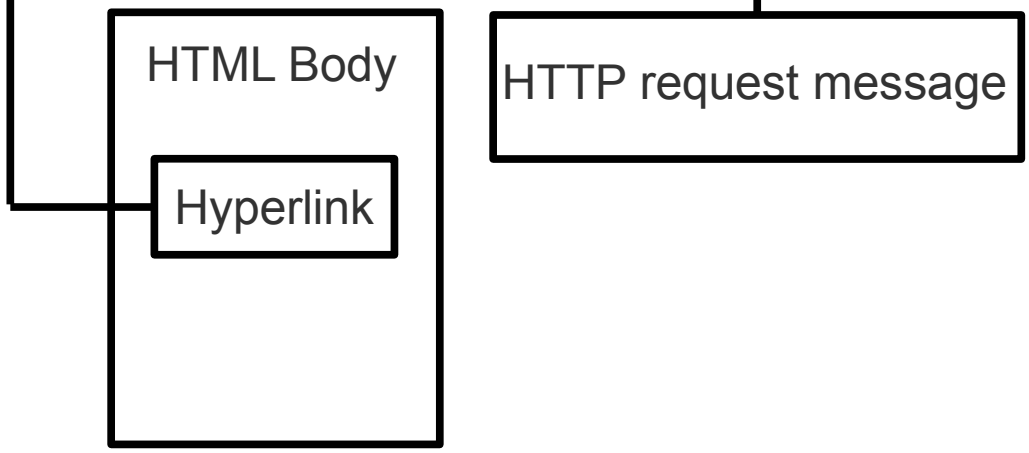


Request for	from source	false/true positive
S_1 - labsite	bookmark	true
S_2 - /favicon.ico	security.nl	false
S_3 - Hyperlink web page 2	labsite	true
S_4 - Hyperlink web page 3	labsite	true
S_5 - Hyperlink article 1	security.nl	true

Hyperlink present at GET header(1)

- Hyperlink

Hyperlink: **1/index.html**
GET header: /rp2/new_website/**1/index.html**



Method 4 Hyperlink present at GET header

```
for all Rs^n in Rs do
  for all Rq^n in Rq where Rq^n.ID > Rs^n.ID do
    if Rq^n.GET in Rs^n.hyperlinks then
      print A new HTTP session was detected
      S = storePair(Rq^n)
    end if
  end for
end for
```

Hyperlink present at GET header(2)

- 301 response message contains a hyperlink

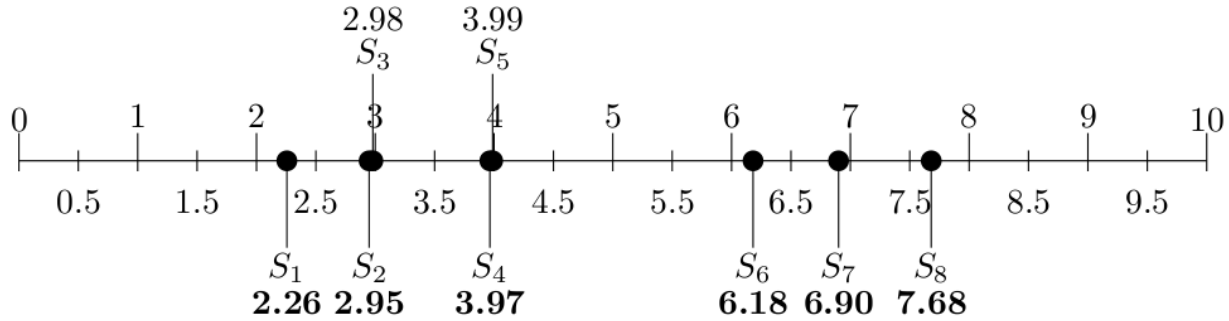


Figure 6: Start time of the HTTP session in seconds.

Request for	from source	false/true positive
S_1 - Hyperlink web page 1	labsite	true
S_2 - Hyperlink web page 2	labsite	true
S_3 - Hyperlink web page 2 (301)	labsite	false
S_4 - Hyperlink web page 3	labsite	true
S_5 - Hyperlink web page 3 (301)	labsite	false
S_6 - Hyperlink article 1	security.nl	true
S_7 - Hyperlink article 2	security.nl	true
S_8 - Hyperlink article 3	security.nl	true

No referrer(1)

- Address bar
- Bookmark

Method 3 No referrer

for all Rq^n in Rq **do**
  if $Rq^n.referrer == Null$ **then**
    print A new HTTP session was detected
    $S = \text{storePair}(Rq^n)$
  end if
end for

```
# #### New HTTP session: 2 #### #
 Time: 1.062025

 >>> REQUEST >>>
Packet ID:..... 11
Frame ID:..... 16
Stream ID:..... 3
Time:..... 1.062025
SRC:..... 145.100.107.4:46732
DST:..... 213.156.1.80:80
Host:..... www.security.nl
GET:..... /
User-Agent:..... Opera/9.80 (X11; Linux i686; U; en-GB) Presto/2.2.15 Version
10.10
--------------------------------------------------------------------------------
 <<< RESPONSE <<<
Packet ID:..... 19
Frame ID:..... 78
Stream ID:..... 3
Time:..... 1.585473
SRC:..... 213.156.1.80:80
DST:..... 145.100.107.4:46732
Response Code:.... 200 OK
Date:..... Thu, 17 Jun 2010 19:21:13 GMT
Content-Type:..... text/html
Body:
      <?xml version="1.0" encoding="UTF-8" standalone="yes"?>
      <!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN" "http://www.w3.o
g/TR/xhtml1/DTD/xhtml1-strict.dtd">
```

Proof of Concept

No referrer(2)

- Javascript removes the referrer

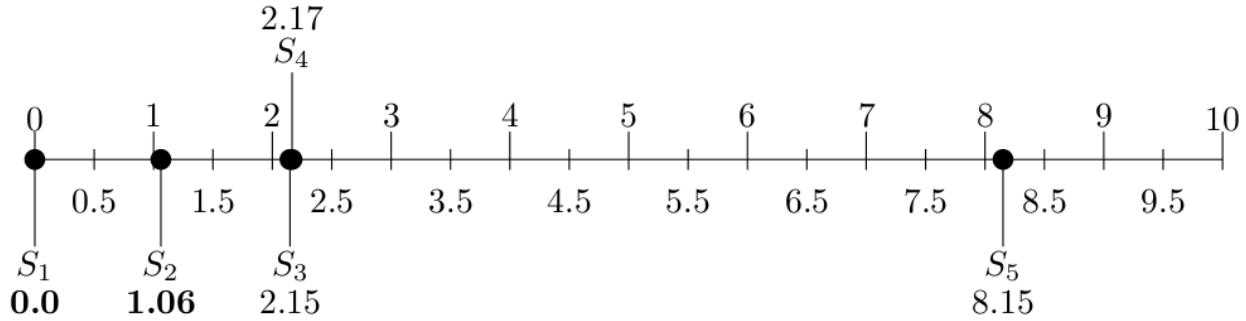


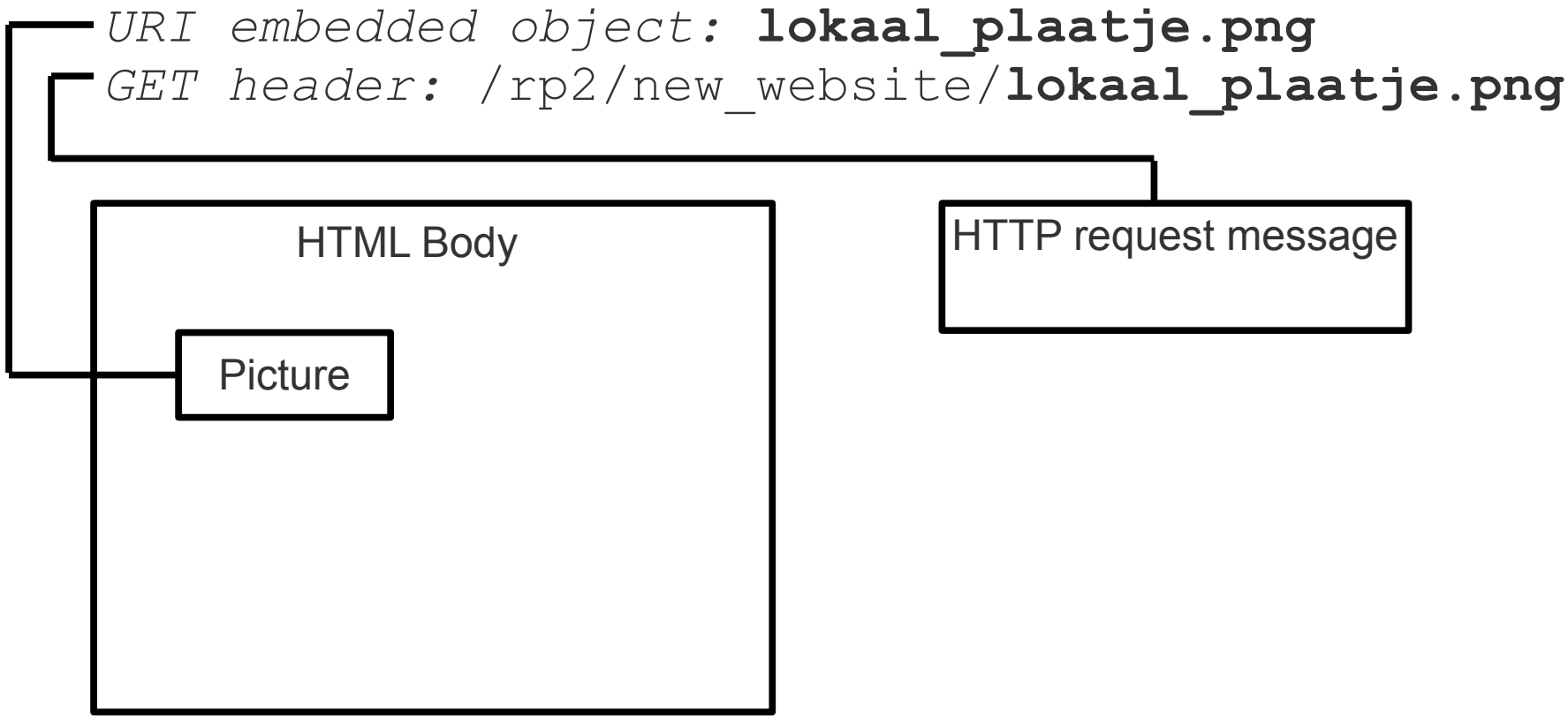
Figure 5: Sessions start time in seconds.

Request for	**from source**	**false/true positive**
S_1 - labsite	bookmark	true
S_2 - security.nl	bookmark	true
S_3 - /pagead/js/graphics.js	pagead2.googlesyndication.com	false
S_4 - /pagead/abglogo/abg-nl-100c-ffffff.png	pagead2.googlesyndication.com	false
S_5 - /pagead/sma8.js	pagead2.googlesyndication.com	false

HTTP message correlation

- HTML body HTTP GET correlation
- Link the referrers

HTML body HTTP GET correlation(1)



HTML body HTTP GET correlation(2)

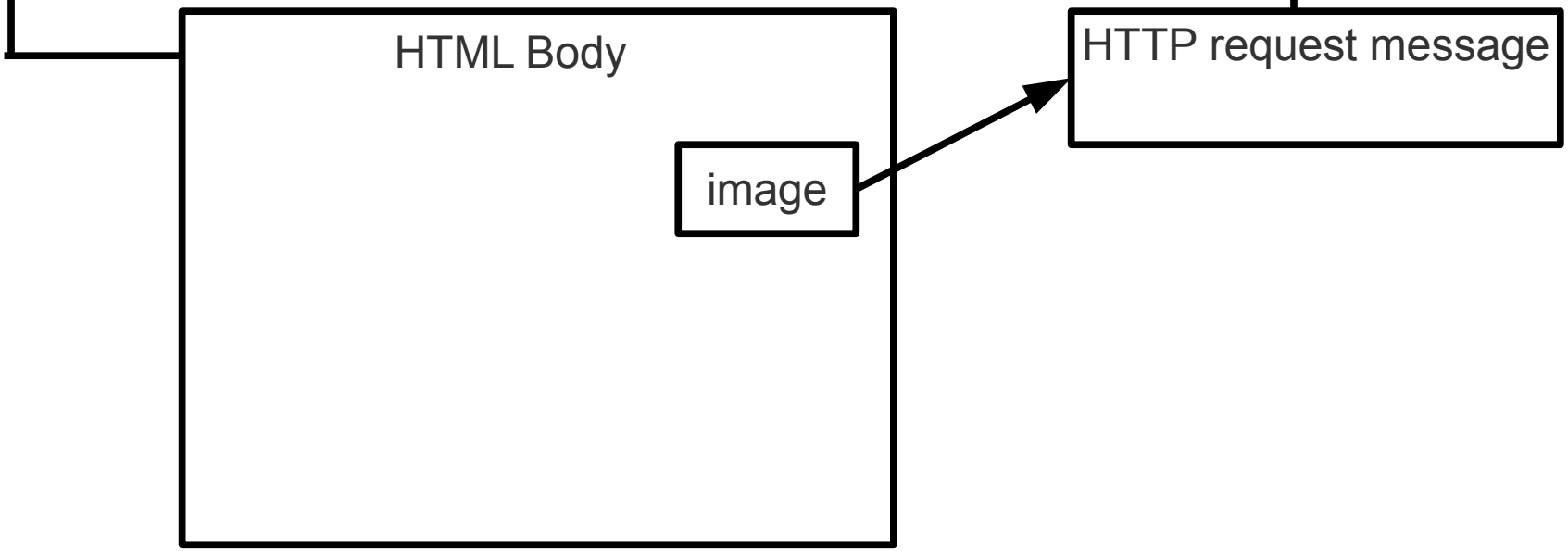
- Javascript:

```
document.write(unescape("%3Cscript src='" + gaJsHost +
"google-analytics.com/ga.js' type='text/javascript'%3E
%3C/script%3E"));
```

Link the referrers(1)

Host+GET header:
bulbasaur.studlab.os3.nl/rp2/new_website/

Referrer:
http://bulbasaur.studlab.os3.nl/rp2/new_website/



Link the referrers(2)

- Javascript can change the referrer:
  `http://pagead2.googlesyndication.com/pagead/ads?client=<VERY LONG STRING>`

Conclusion

- Start of a HTTP session
  - Time between successive fetches
  - Hyperlink present at GET header
  - No referrer
- HTTP message correlation
  - HTML body HTTP GET correlation
  - Link the referrers

Future work

- Large scale testing
- Time between successive fetches for mobile phones
- Web 2.0

Questions?