

# Reconstructing web pages from browser cache

Iwan Hoogendoorn  
&  
Edwin Schaap

University of Amsterdam

July 4, 2013



- Open Safari
- Clear Safari's cache
- Visit [www.tweakers.net](http://www.tweakers.net)

# Criminal research

- planning a crime
- committing the perfect crime
- Internet used as a resource



# Evidence by a witness

- looking at content that is against the law
- content is removed by a suspect in a later stage
- Internet used as a resource



# Forensic crime investigation



- computer forensics
- browser forensics
- web cache data forensics

---

*In what ways can one visually reconstruct websites with information retrieved from normalized browser caches that can be use for computer forensic examiners to build a case?*

---

- Raw caching data
- Reconstruction methods
- Reliability after reconstruction

# Current forensic web cache tools

- Nirsoft
- Web Cache View
- Digital Detective
- Siquest
- Foxten Software

- XIRAF
- HANSKEN
- Traces



# Popular web browsers

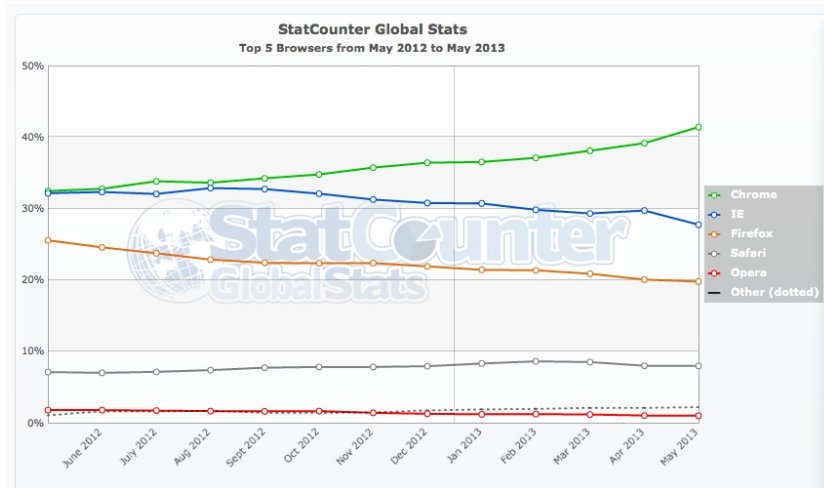


Figure 1 : Browser popularity - Worldwide

# Web cache data structure - Google Chrome

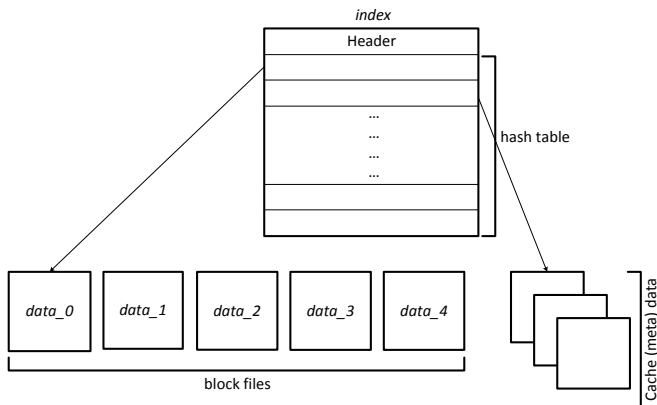


Figure 2 : Chrome web cache structure

# Web cache data structure - Mozilla Firefox

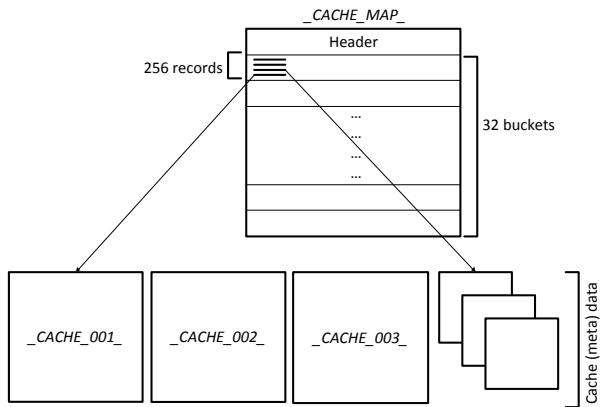


Figure 3 : Firefox web cache structure

# Web cache data structure - Apple Safari

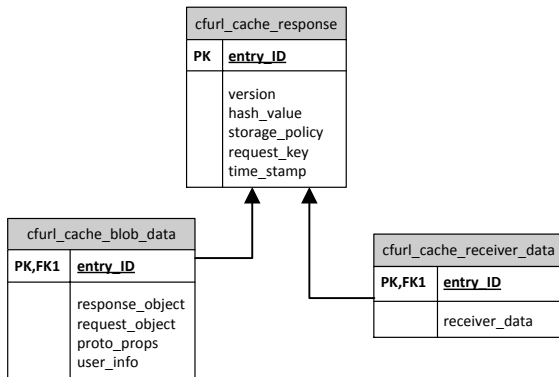


Figure 4 : Safari web cache structure

## Web cache data - before sanitization

	<b>Chrome</b>	<b>Firefox</b>	<b>Safari</b>
Unique identification	✓	✓	✓
Eviction	✓	✓	X
URL request string	✓	✓	✓
Time/Date (first request)	✓	✓	✓
Time/Date (last request)	X	✓	X
Time/Date (expire)	X	✓	X
Fetch count	X	✓	X
Client request headers	X	X	✓
Server response header	✓	✓	✓
Server response body	✓	✓	✓

Table 1 : Firefox, Chrome and Safari web cache comparison table

- Unique identification
- URL request string
- Time/Date (first request)
- Server response body

- pre-processing
- post-processing

## Pre-processing

- Advantages:
  - ① Requires no configuration of the rendering browser.
  - ② Can even run in the browser of the user enabling interaction.
- Disadvantages:
  - ① Tampering the evidence.
  - ② Hard to parse all resource identifiers, especially if JavaScript is used.
  - ③ Non-parsed resource identifiers are circumventing the application.



## Post-processing

- Advantages:
  - ① All resource identifiers are captured by the proxy.
- Disadvantages:
  - ① Requires proxy configuration of rendering browser.
  - ② SSL traffic is hard to deal with.

Proof of Concept

# Application design

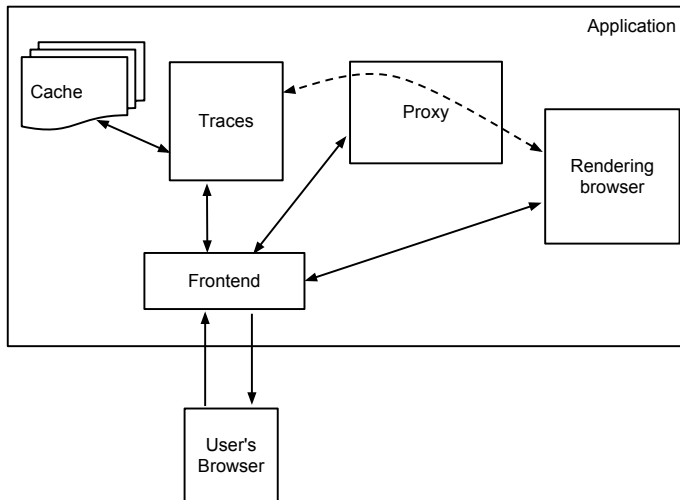
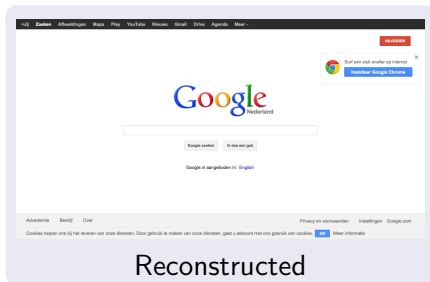
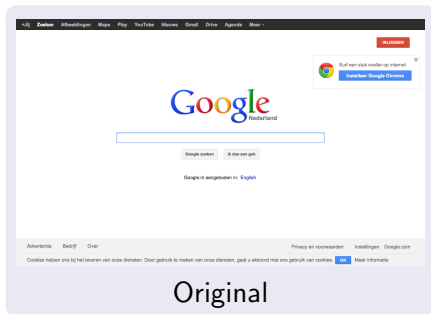


Figure 5 : Web page reconstruction application

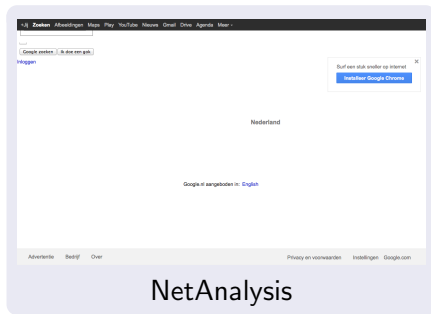
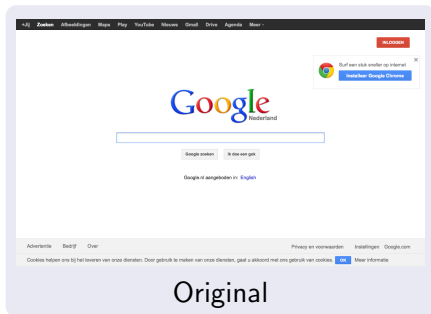
- Reconstruct web page visited at the beginning of this presentation
- Compare before and after



# Result - Simple websites I



# Result - Simple websites II



# Result - complex websites I

The original NU.nl website features a complex layout with a prominent top banner for Jumbo's BBQ Time. The navigation bar includes a search function and various category links. The main content area is divided into several sections: a large article about a fire in Rotterdam, a weather widget, a sports section, and a sidebar with a navigation menu. The design is cluttered with many elements, including multiple social media icons and a dense list of news items.

**Original**

The reconstructed NU.nl website presents a clean, modern, and user-centric design. It features a clear navigation bar, a large hero image, and a structured layout of content blocks. The design is minimalist, focusing on readability and ease of navigation. Key elements include a prominent search bar, a clear hierarchy of news items, and a clean sidebar. The overall aesthetic is professional and easy to use.

**Reconstructed**

# Result - complex websites II

**JUMBO** DE LEKKERSTE SPIESEN TE PRIJS!  
Alles voor de BBQ

**IT'S BBQ-TIME BIJ JUMBO!**

Donderdag 6 juni 2013. Het laatste nieuws het eerst op NU.nl

**Dijk langs Donau dreigt door te breken**  
Ook rivier Elbe zorgt nog voor gevaar, duizenden mensen geëvacueerd

**Volg NU.nl**  
Facebook Twitter

19°C  
Lucht vochtig 92%  
€ 1,001  
Euro 0%  
Lijkt een nieuwjaar

**Achterklap**  
Theresa Bergs vader van 2004 dood

**DE LEKKERSTE SPIESEN VOOR DE LAARSTE PRIJS!**

**NU.nl beeld**

**Economie**  
**'Gespreid betalen verkeersboete moet kunnen'**  
Ombudsmen zien veel betalingsproblemen door hoge boetes

**Sport**

Original

Middelste Middel Nieuwsport Nieuws Nieuws Nieuws Nieuws Nieuws Nieuws Nieuws

**Politiek**  
• Hongarije  
• Azerië  
• Algerije  
• Spanje  
• Nederland  
• België  
• Economie  
• Schiednis  
• Oekraïne

**Braz**  
• Sport  
• Roland Garros

**Tech**  
• Achterklap  
• Openbreuk  
• Cultuur en Media  
• Elan en serie  
• Muziek  
• Book  
• Media  
• Oerzie

**Wetenschap**  
• Gezondheid  
• Lifestyle  
• Auto  
• Nieuws  
• Economie

**Gezondheid**  
• Datalog  
• Redactie  
• Weer  
• Verkeer  
• NU.nl-apps  
• Oekraïne

**Wetenschap**

**Gezondheid**

**Lifestyle**

**Auto**

**Nieuws**

**Redactie**

**Weer**

**Verkeer**

**NU.nl-apps**

**Oekraïne**

**• Het laatste nieuws het eerst op NU.nl**

**Dijk langs Donau dreigt door te breken**

Ook rivier Elbe zorgt nog voor gevaar, duizenden mensen geëvacueerd

- Turkse agent dood bij protest
- Gewonde bij grote brand in Rotterdam
- Dieren moet acrogymnastie bezette
- Dooden, Sassenheim 'over geweldig ongeluk'
- Bewaarder: wordt gelikt over zaak Farid
- Prijs Fyra vlak voor verkoop opengeklapt
- Tientallen miljoen Ape d'Hzes nog ongebruikt
- 'Prijs Fyra vlak voor verkoop opengeklapt'

**Economie**

**'Gespreid betalen verkeersboete moet kunnen'**

Ombudsmen zien veel betalingsproblemen door hoge boetes

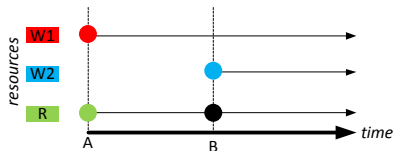
- ECB heeft nuw op 0,5 procent
- Europese Commissie wijst IMF-rapport Griekenland af
- Overheid dreigt geld te verliezen op PPS-contracten
- Europese bedrijven moeten bijpassen aflossing
- Polaire nieuwe naam Soeraya en De Sijgert
- Inflation omhoog naar 2,8 procent
- G4 en Krimprag's komen met alternatief verhuurdersheffing
- Nederlandse beïnvloed vast in Duitsland
- Nederlandse lobby in Brussel kost 60 miljoen euro

**Sport**

NetAnalysis



# Analysis - Dynamic resources



- 1 Browser S displays website W1 on time A.
- 2 Website W1 contains resource R.
- 3 Browser S displays website W2 on time B.

# Analysis - Runtime dependencies

- 1 Browser  $S$  visits website  $W$ .
- 2 Website  $W$  contains a dynamic time  $T$ .
- 3 Time  $T$  is taken from the local system time.

- Prefer post-processing
- Normalized data is sufficient
- Reliability depends on cache data

Thank you

?