Who is Fiddling with Prices?
Building and Deploying a Watchdog Service for E-commerce

sheriff_v2

Costas Iordanou Claudio Soriente
Michael Sirivianos Nikolaos Laoutaris
Walking into a brick and mortar store
everyone sees the same price, be it for fresh vegetable or a pair of jeans
Navigating into an e-commerce store

what User A sees … is not what User B gets.

Even more interestingly, you have no easy way of knowing that others see different prices.

The reason is …
An e-commerce store is like Star Trek’s holodeck

each one can have a different world painted around him
(+ he cannot see into neighboring worlds)
Enter the Price Sheriff

A first-of-its-kind transparency software that allows one to see the prices as seen by others

A first small step towards looking behind and through the matrix
What does Sheriff do?
How does Sheriff do it?
Technical challenges
Findings
$sheriff Demo
What does Sheriff do?

How does Sheriff do it?

Technical challenges

Findings
How does $\text{sheriff}$ do it?

Components

- Browser add-on
- Measurement servers
- Coordinator server
- Database server
- Proxy clients
- Peer proxy browsers
- Network of proxies

Monitoring signals

Peer browsers signals
How does $\text{sheriff}$ do it?

Serving requests

1. Request measurement server
2. Sent request
3.1 Remote Page request
3.2 Remote Page request
4. Store data
5. Show results

Browser add-on
Measurement servers
Proxy clients

database server
Peer proxy browsers
Coordinator server
Peer browsers signals
Why is $heriff interesting?

Had to solve some difficult technical challenges:

- Build a P2P proxy network
- Prevent user profile pollution (Browser and Server side)
- Protect user privacy
- Perform universal price extraction
- Automate currency detection

Gathered lots of interesting measurements:

- More than 2000 e-commerce sites,
- Including the top 400 according to Alexa
- More than 6000 products
- More than 1500 real users in 55 countries
- More than 0.7 million measurement points
What does Sheriff do?
How does Sheriff do it?

**Technical challenges**

Findings
Why hybrid network of proxies?

**Infrastructure proxy clients**
- Diverse predefined geo-locations
- Easy to setup and control
- No real users involved
- No price variation based on personal data can be observed

**Peer proxy browsers**
- Diverse real user profiles
- Price variations based on personal data
- Unpredictable availability and geo-location
- Browser side profile pollution
- Server side profile pollution
Unpredictable availability and geo-location
How we solved it ...

<table>
<thead>
<tr>
<th>Country</th>
<th>Users</th>
</tr>
</thead>
<tbody>
<tr>
<td>France</td>
<td>User1, User 6</td>
</tr>
<tr>
<td>Germany</td>
<td>User2</td>
</tr>
<tr>
<td>Italy</td>
<td>User3, User5</td>
</tr>
<tr>
<td>USA</td>
<td>User4</td>
</tr>
<tr>
<td>...</td>
<td></td>
</tr>
<tr>
<td>CountryN</td>
<td>UserN</td>
</tr>
</tbody>
</table>

Peer proxy browsers
- User1, France
- User2, Germany
- User3, Italy
- User4, USA
- User5, Italy
- User 6, France
- ... 
- UserN, CountryN

Coordinator server

Custom communication protocol over WebRTC data channel
Browser-side profile pollution
When does it happen?

Incoming remote page request
to “e-shop.com/product_A”
every time we send a remote request

How we solved it ...
If the user has never visited the domain, we delete the following

<table>
<thead>
<tr>
<th>Problem (altered state)</th>
<th>Chrome API solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Browsing history</td>
<td>chrome.history → remove the entry</td>
</tr>
<tr>
<td>Browser cached memory</td>
<td>chrome.browsingData → clean cache</td>
</tr>
<tr>
<td>Cookies - response header</td>
<td>chrome.webRequest → remove cookie</td>
</tr>
<tr>
<td>Dynamic cookies - JavaScript</td>
<td>chrome.cookies → delete inserted cookies</td>
</tr>
</tbody>
</table>

Browser extension APIs: https://developer.chrome.com/extensions/api_index
Server-side profile pollution
How does it happen?

Measurement server  Peer proxy browser

1st party server  3rd party 1  3rd party 2  3rd party 3  3rd party N

a.com/product X
b.com/product Y
c.com/product Z

= 1st party cookies  = 3rd party cookies
Server-side profile pollution
When does it happen?
in cases when the user already visited a domain

1st party server
Peer proxy browser

Measurement server

1st party server
3rd party 1
3rd party 2
3rd party 3
3rd party N

a.com

a.com/product X
b.com/product Y
c.com/product Z

1st party cookies
3rd party cookies

= 1st party cookies
= 3rd party cookies
Remote page requests to e-store domains a, b and c pollute the user profile at the 1st and 3rd party domain servers respectively.

= 1st party cookies  = 3rd party cookies
Server-side profile pollution
How we solved it ...
using “doppelgänger” profiles

Doppelgänger “an identical copy of someone”,
a ghost in the german folklore.
How do we create a doppelgänger?

Server-side profile pollution

Peer proxy browsers

- a.com: 10
- b.com: 9
- c.com: 6
- ...

Encrypted Domain vectors

Aggregator Server

- sd4a5l6kj9dlkjsa
- u9w7p4e6r7u0w
- Mn4bvm4bmlvb
- ...

Distances

Encrypted k-Means

Coordinator Server

- a.com: 8
- b.com: 7
- c.com: 5
- ...

New Centroids
Server-side profile pollution
How do we create a doppelgänger?

Peer proxy browsers

Aggregator Server

Encrypted Domain vectors

sd4a5l6kj9dlkjsa
u9w7p4e6r7u0w
Mn4bvm4bmlvb

... 

Coordinator Server

Encrypted k-Means

a.com: 8
b.com: 7
c.com: 5

... 

Crawling

Coordinator Server

Doppelgangers Server

domain: a.com
userID: 985631

domain: b.com
userID: 654852

domain: c.com
userID: 851236

... 

Use

Store
Server-side profile pollution
How do we create a doppelgänger?

By replacing the original cookies of a user with those of a doppelganger...

remote page requests will pollute the doppelganger profile.
What does Sheriff do?
How does Sheriff do it?
Technical challenges

Findings
Price variations

Three types of results:

Different geo-location

Same geo-location

Temporal monitoring
Findings Summary

1. Price variation across countries
   - 76 domains out of 1994
   - price variation up to 600%

2. Price variation within the same country
   - 7 out of 76 domains (3 repeatable)
   - price variation up to 7%

3. No price discrimination based on personal data detected yet
Prices vary depending on the country.

Systematic crawling dataset

# Requests

Normalised Price Difference

30%
Prices also vary within the same country.
Who is Fiddling with Prices?

Building and Deploying a Watchdog Service for E-commerce

Costas Iordanou
Universidad Carlos III de Madrid, Telefonica Research
kostas.iordanou@telefonica.com

Michael Sirivianos
Cyprus University of Technology
michael.sirivianos@cut.ac.cy

Claudio Soriente
Telefonica Research
claudio.soriente@telefonica.com

Nikolaos Laoutaris
Data Transparency Lab
nikos@datatransparencylab.org

ABSTRACT

We present the design, implementation, validation, and deployment of the Price $heriff, a highly distributed system for detecting various types of online price discrimination in e-commerce. The Price $heriff uses a peer-to-peer architecture, sandboxing, and secure multiparty computation to allow users to tunnel price check requests through the browsers of other peers without tainting their local or server-side browsing history and state. Having operated the Price Sheriff for several months with approximately one thousand real users, we identify several instances of cross-border price discrimination based on the country of origin. Even within national borders, we identify several retailers that return different prices for the same product to different users. We examine whether the observed differences are due to personal-data-induced discrimination or A/B testing, and conclude that it is the latter.

commonplace among e-commerce sites. These studies have mostly established that the location of a customer, and in particular the country of origin, inferred via his IP address and language settings, often affects the observed price in ways that cannot be explained in terms of currency, taxation, duty, or shipping costs.

In a few cases, researchers have even managed to reverse engineer, or at least hypothesize, about the suspected causal relationship between location and price and have shown, for example, that prices appear to be adjusted using simple multiplicative factors depending on the country of the customer [18]. Despite this initial progress in unveiling cross-border online PD, little is known about other aspects of dynamic pricing. For example, despite anecdotal evidence, there’s little work in measuring dynamic pricing within national borders. Do customers within the same country see different prices for the same product by the same vendor? If they do, can this be...
Main takeaways

1. $heriff is a first-of-its-kind transparency software

2. We communicate the challenges involved in the development of such system

3. Our architecture and implementation choices can help others build similar services
$\textit{sheriff}_v^2

http://sheriff-v2.dynu.net

Available for:

\begin{itemize}
  \item \text{chrome}
  \item \text{Firefox}
\end{itemize}

email: costas.iordanou@telefonica.com
website: http://www.tid.es/research/researchers/costas-iordanou