I Know Where You've Been: Geo-Inference Attacks via the Browser Cache

Yaoqi JIA
Department of Computer Science
National University of Singapore

black hat ASIA 2015

NUS National University of Singapore

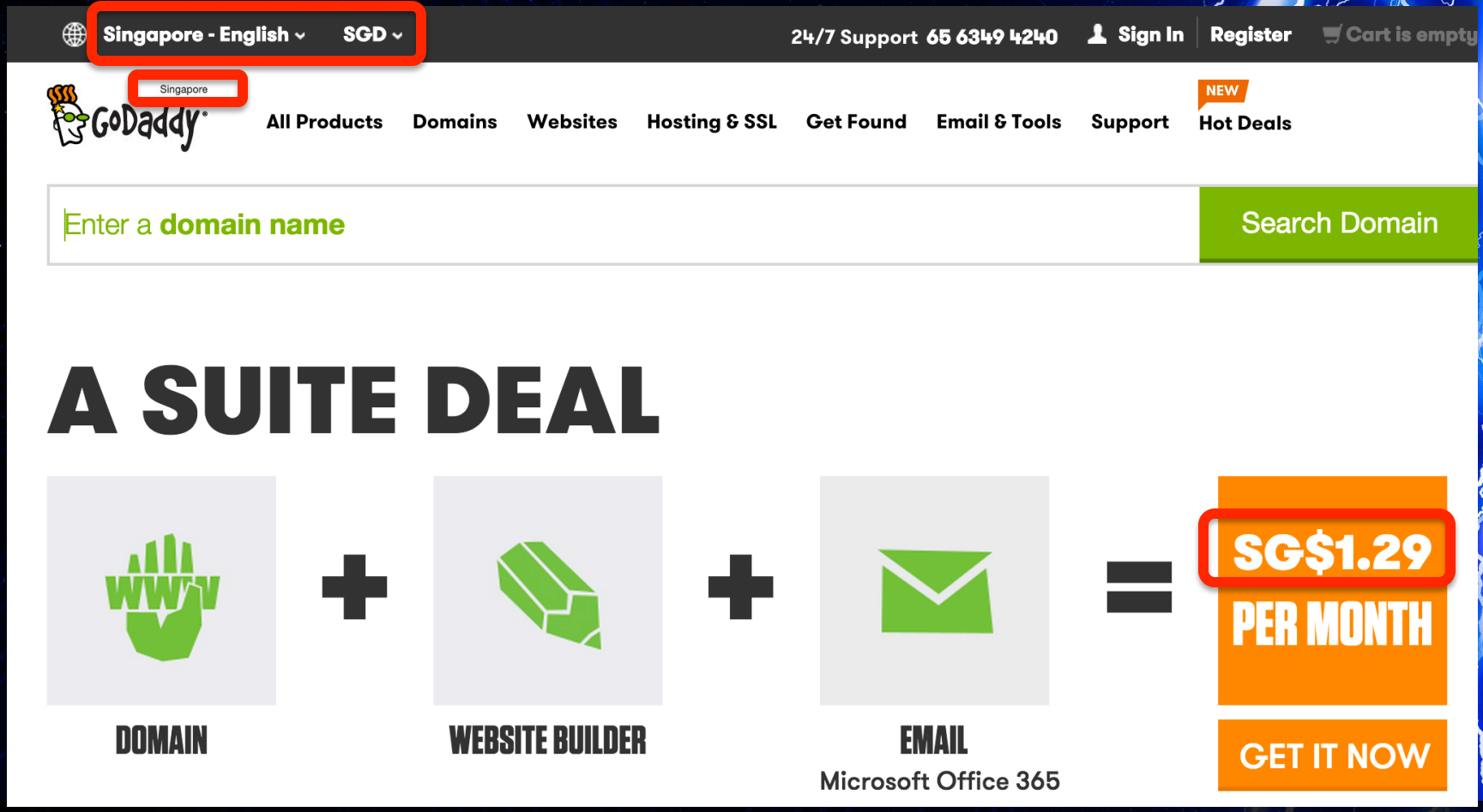
Do You Care About your Geo-location?

Video : How to Infer Your Geo-location without Your Consent

Our Agenda

- Background of geo-locations in browsers, browser cache, and timing channels
- Geo-inference attacks via the browser cache
- Prevalence of geo-inference attacks
- Pros & cons of potential solutions
- Demo Video for attacks in TorBrowser
- Q & A

Geo-location in Browsers



Geo-location in Browsers

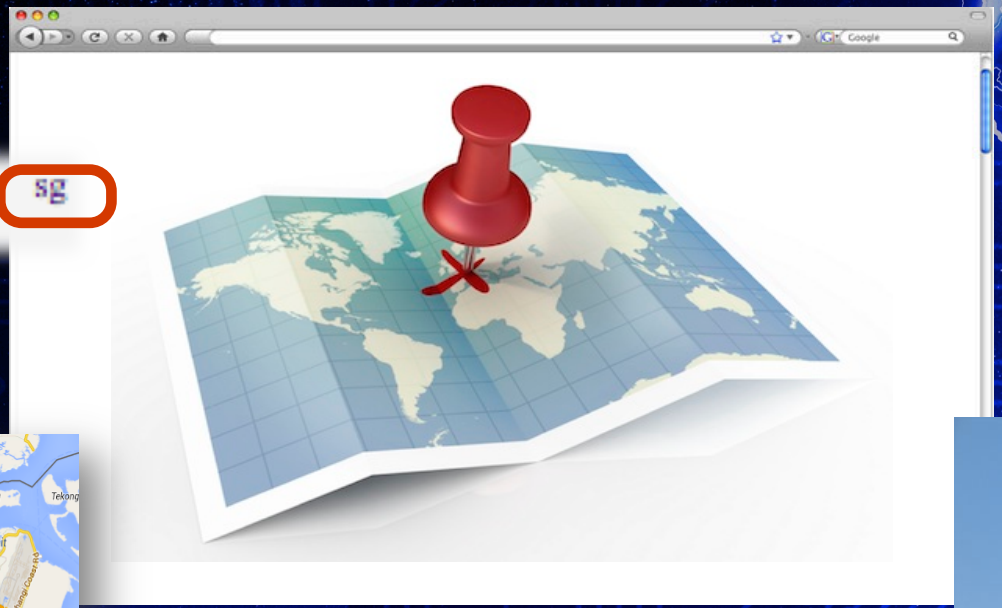
craigslist open source
craigslist blog
event
farm+garden
skill'd trade
sm biz ads
cds/dvd/vhs
cell phones
sporting
tickets
technical support
transport

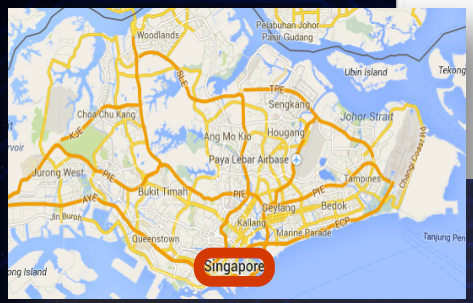
Geo-location in Browsers: Benefits & Threats

Benefits

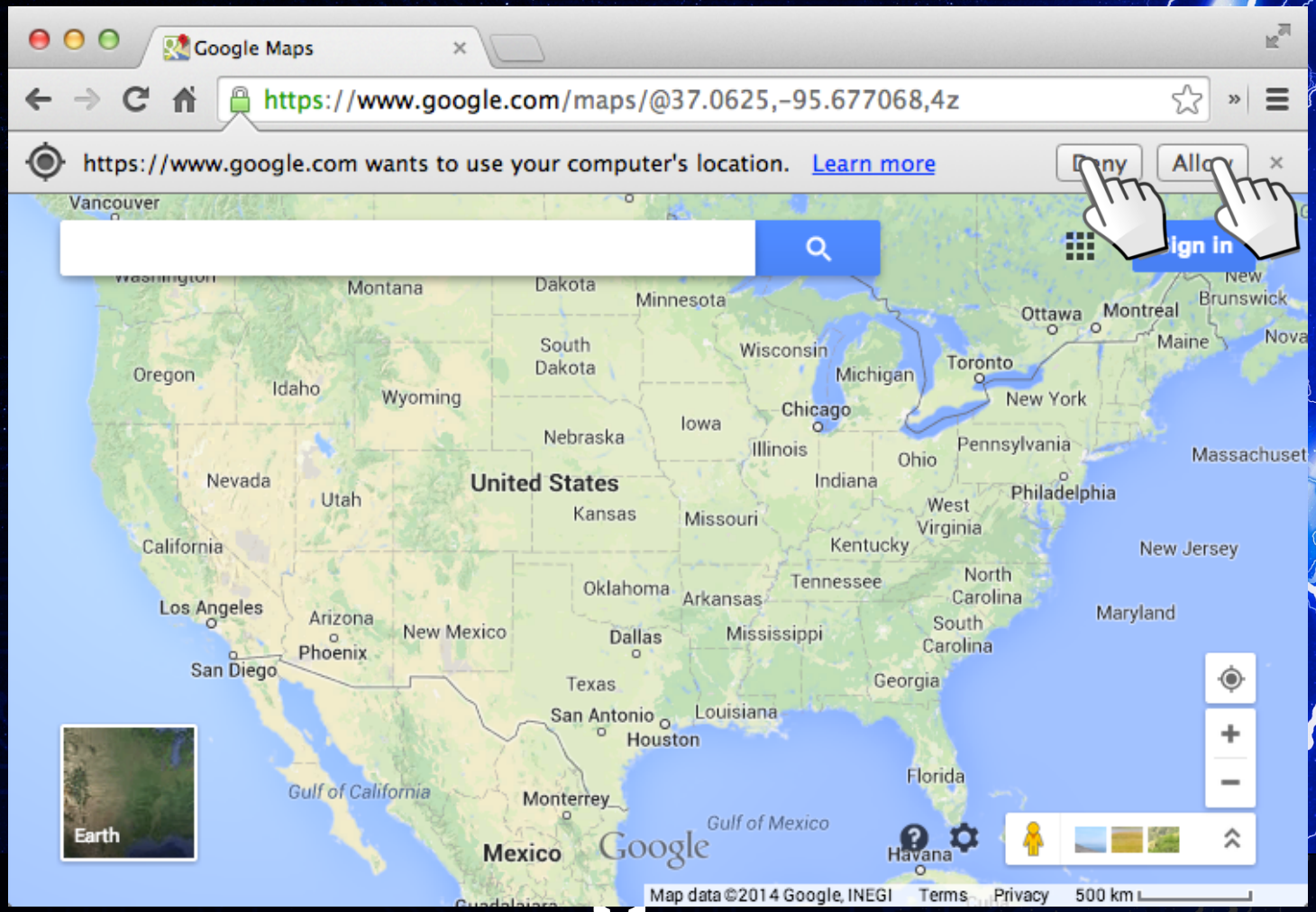
Threats

craigslist sg

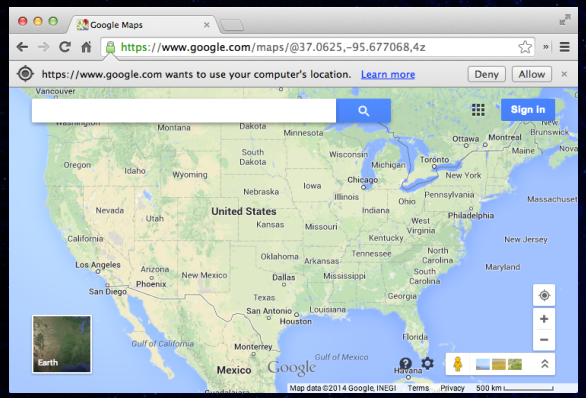


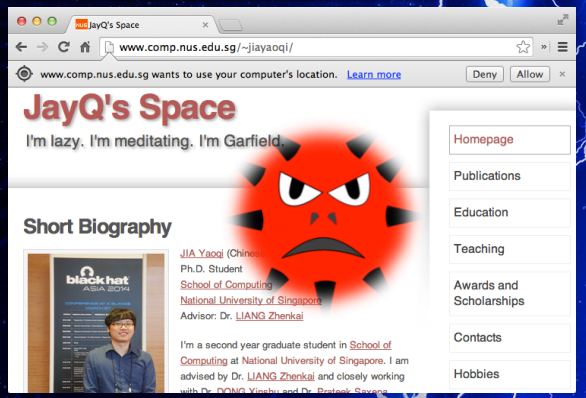


May I Access Your Geo-location?



Sources of Users' Geo-locations

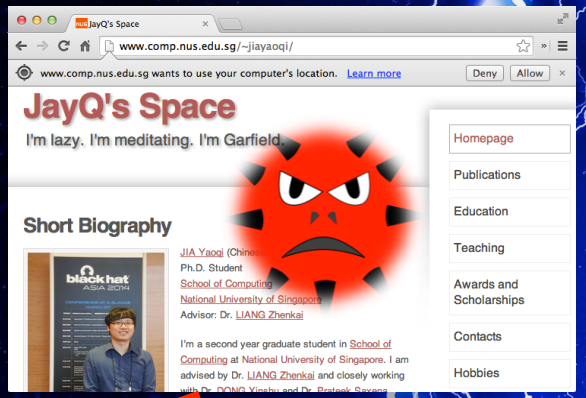


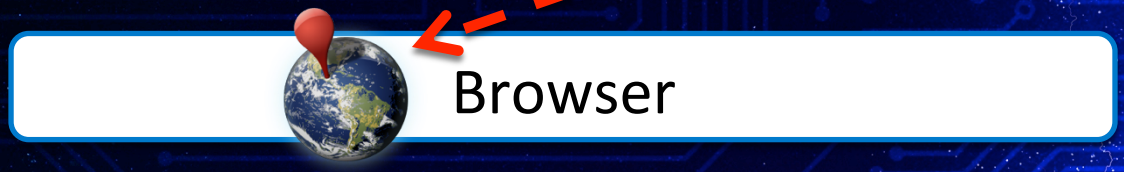


Browser

Not reliable

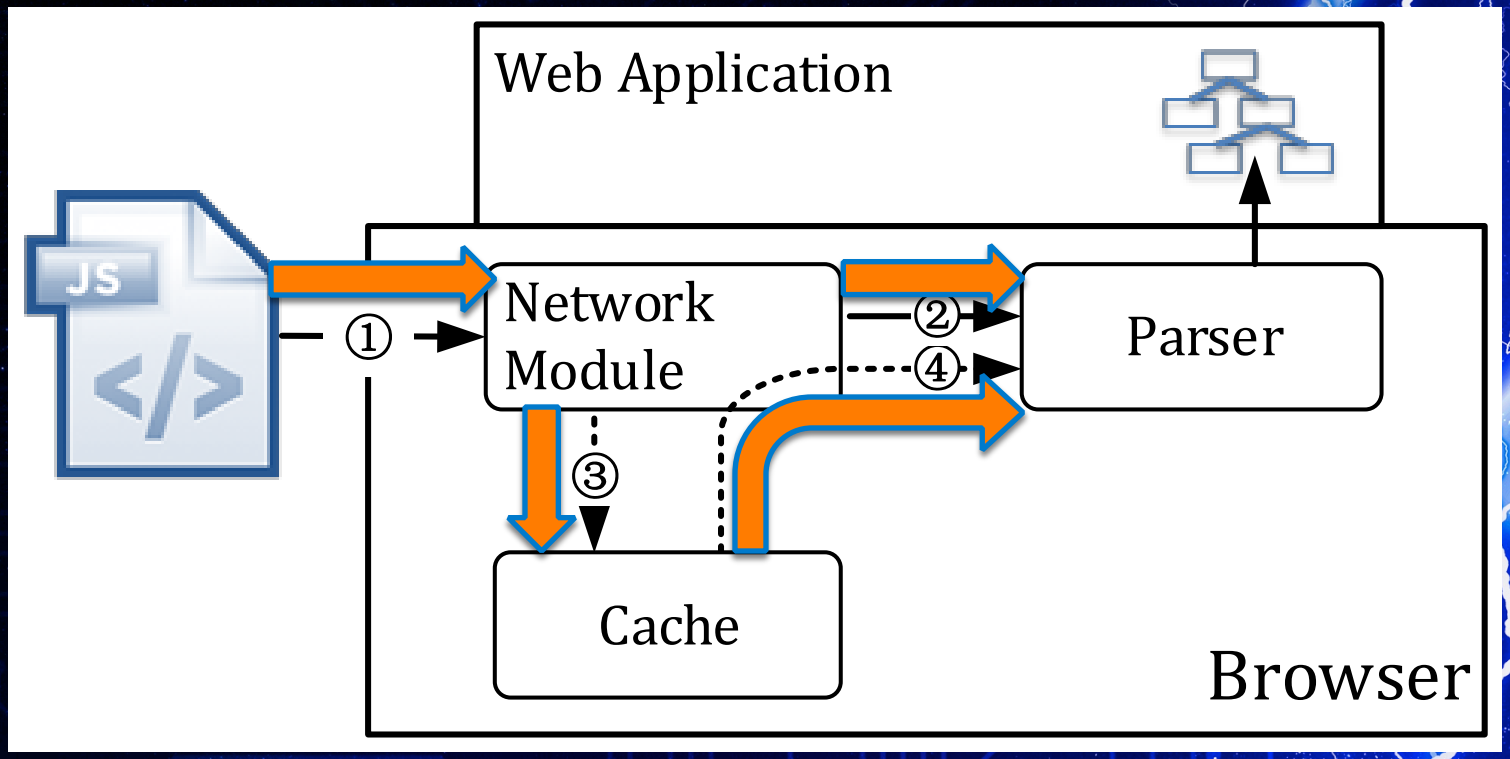
Problem Statement





Can the attacker infer the user's geo-location from his browser?

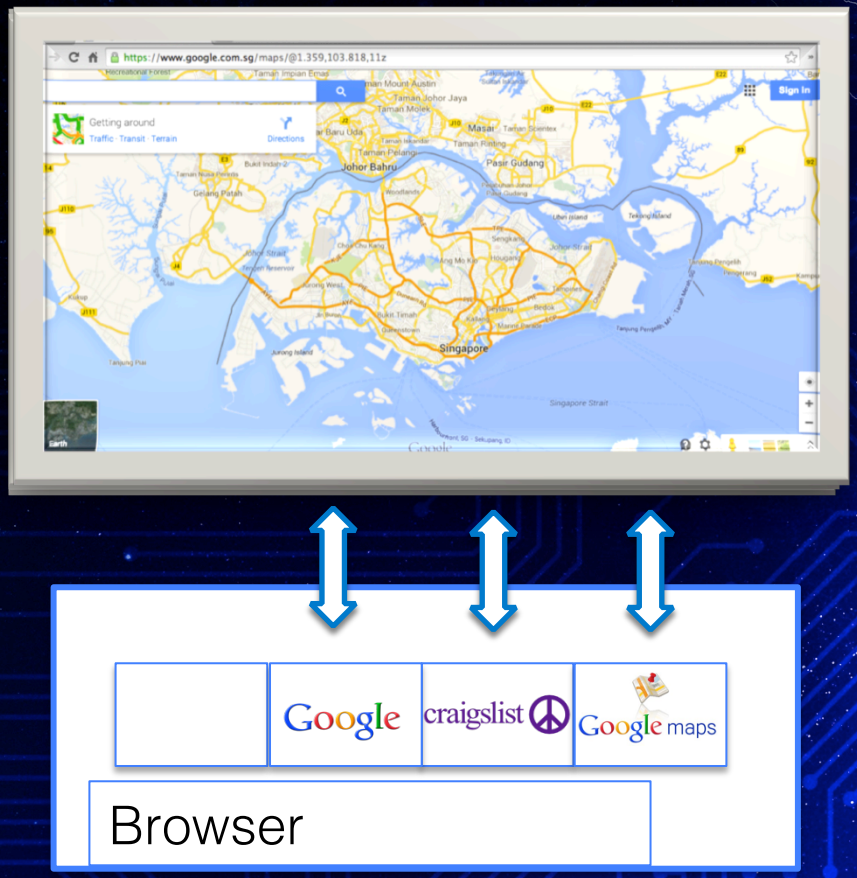
Background: Browser Cache



Directives in Response Headers to Control Cache

- **Static resources:**
  - Expires, Cache-Control: max-age, Last-Modified
- **Dynamic and sensitive resources:**
  - Cache-Control: no-cache, no store; Pragma: no-cache; Expires: 0

Browser Cache Stores Static Resources



Browser stores site-related states

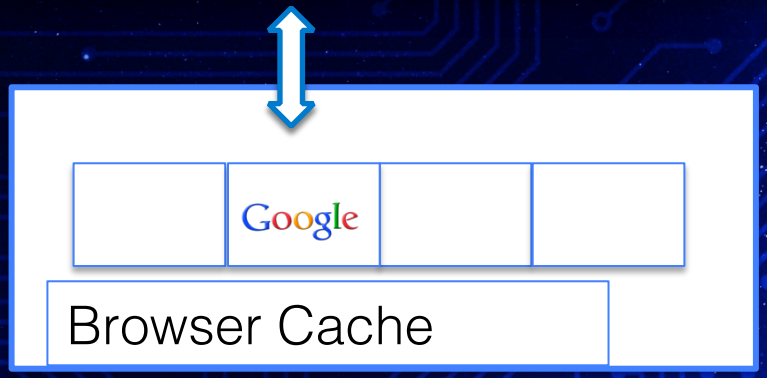
Benefits of Browser Cache

1st: 1360ms

2nd: 320ms

3rd: 350ms

Save Time!

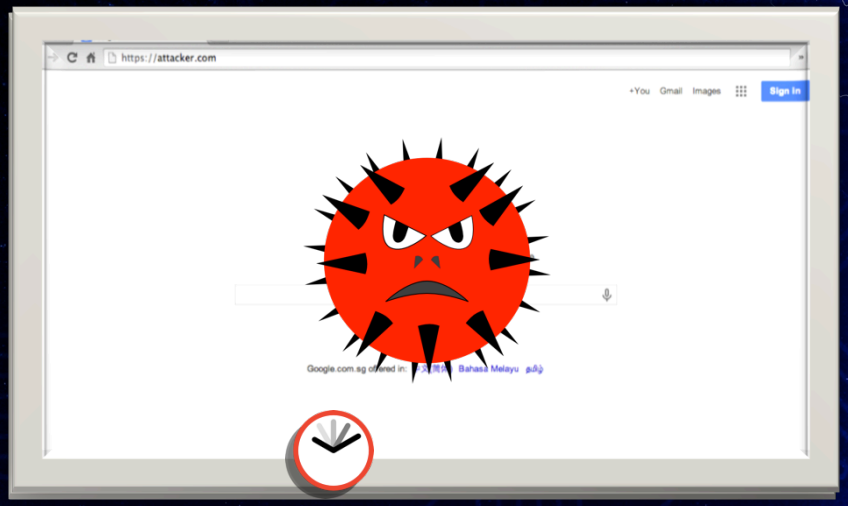


Timing Channels via the Browser Cache

1st: 1360ms

2nd: 320ms

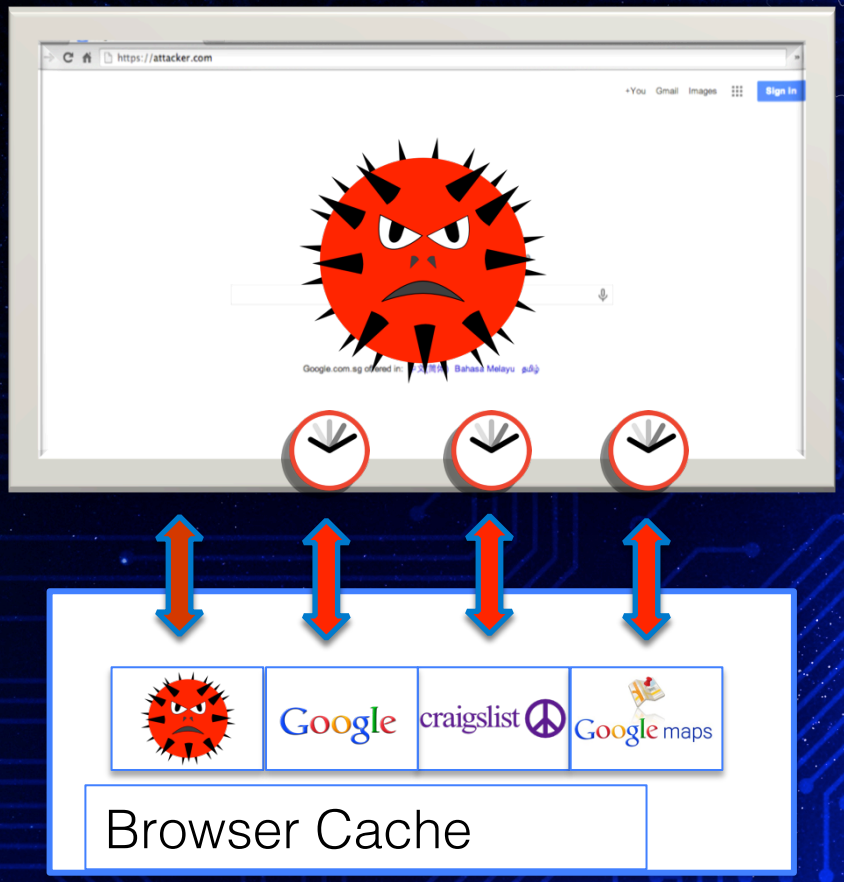
3rd: 350ms



Geo-Inference Attacks via the Browser Cache

Browser cache is shared across all sites

Infer users' geo-locations!



Our Attacks:
Infer a User's Geo-location without the Manual Input, Accessing GPS Sensors or IP Addresses

What are the Techniques to Determine the Cache Status of Targeted Resources?

Attack Vector (I) : Measuring Image Load Time

Before Loading

img.onload Fires

```
var image = document.createElement(`img');

image.setAttribute(`startTime', (new
Date().getTime()));

image.onload = function()

{

    var endTime = new Date().getTime();

    var loadTime = endTime -
parseInt(this.getAttribute(`startTime'));

    ......

}
```

attacker.com

Attack Vector (II) : Measuring Page Load Time

Before Loading

iframe.onload Fires

```
var page = document.createElement(`iframe');

page.setAttribute(`startTime', (new
Date()).getTime());

page.onload = function ()

{

    var endTime = (new Date()).getTime();

    var loadTime = ( endTime -
parseInt(this.getAttribute(`startTime')));

    ......

}
```

attacker.com

Attack Vector (III) :Measure the Load Time of XMLHttpRequests

onloadstart Fires

onloadend Fires

```
var starTime, endTime, loadTime;
var xmlhttp = new XMLHttpRequest();
xmlhttp.onloadstart = function(){
    startTime = (new Date()).getTime();
}
xmlhttp.onloadend = function(){
    endTime = (new Date()).getTime();
    loadTime = endTime - startTime;
    ......}
```

attacker.com

Attack Vector (IV) : Use <img>'s complete Property

```
function cached(url)

{

    var image = document.createElement(`img');

    image.src = url;

    return image.complete || image.width+image.height >
0;

}
```

attacker.com

Examples: What Can We Achieve?

- User's country?
- User's city?
- User's streets or neighborhood?

How to Infer a User's Country? (I)

- Google has 191 regional sites.
- One site represents one country or region.

google.com.sg/images/srpr/logo11w.png

How to Infer a User's Country? (II)

Google Canada

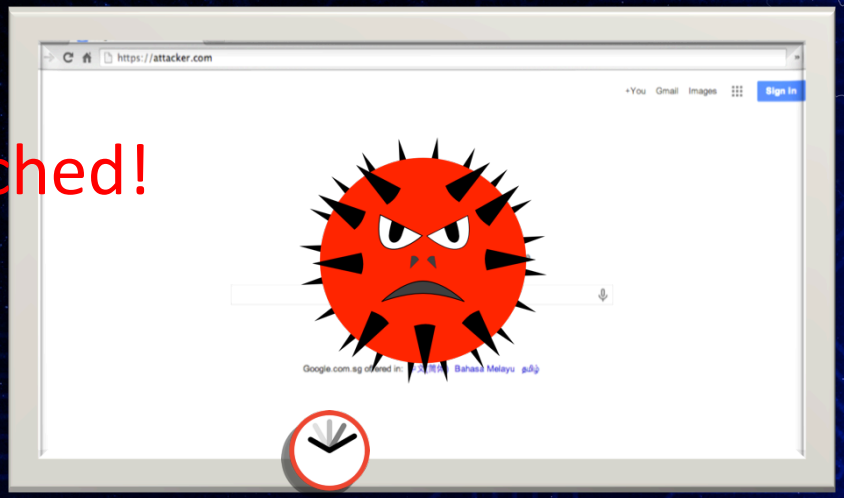
Google Singapore

Google Japan

Google UK

Google Deutschland

Cached!



How to Infer a User's City? (I)

- Craigslist provides local classifieds advertisements and forums for jobs, housing, etc.
- Craigslist has 712 city-specific sites.
- Users buy or sell second-hand stuff in their Craigslist's city-specific sites.

How to Infer a User's City? (II)

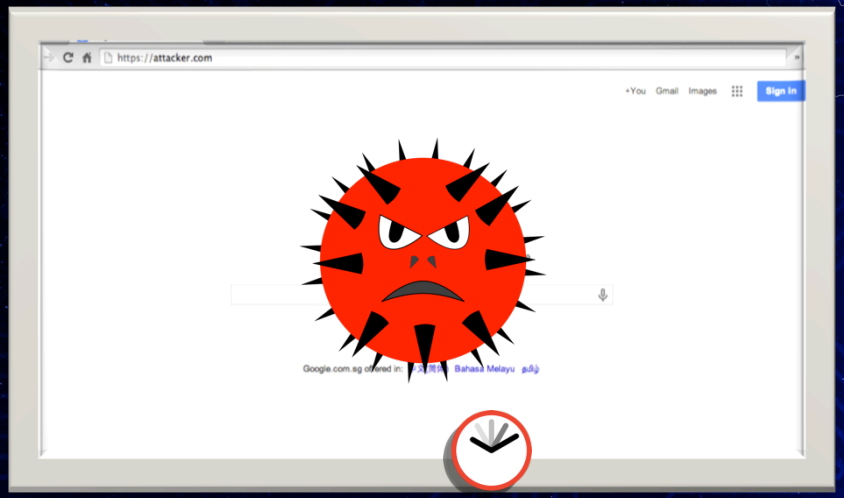
chicago.craigslist.org

sfbay.craigslist.org

newyork.craigslist.org

singapore.craigslist.com.sg

tokyo.craigslist.jp



Cached!

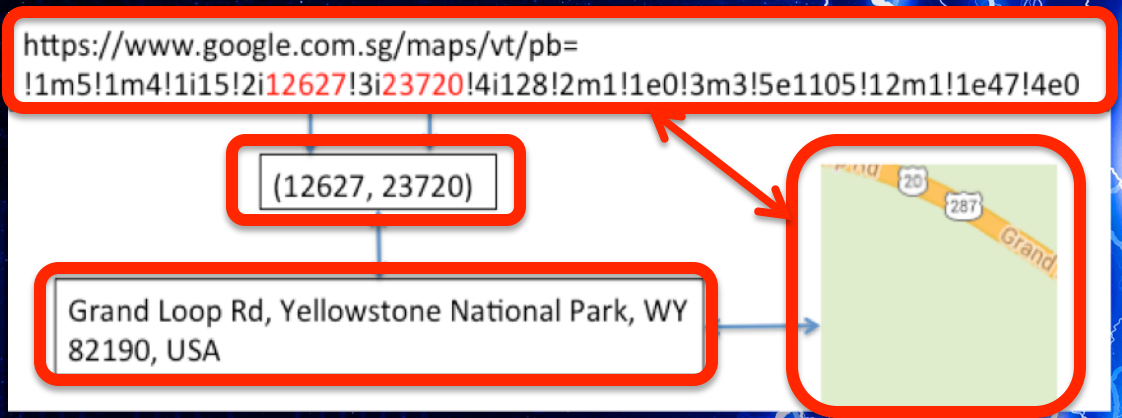
How to Infer a User's Neighborhood?(I)

Predictable URLs

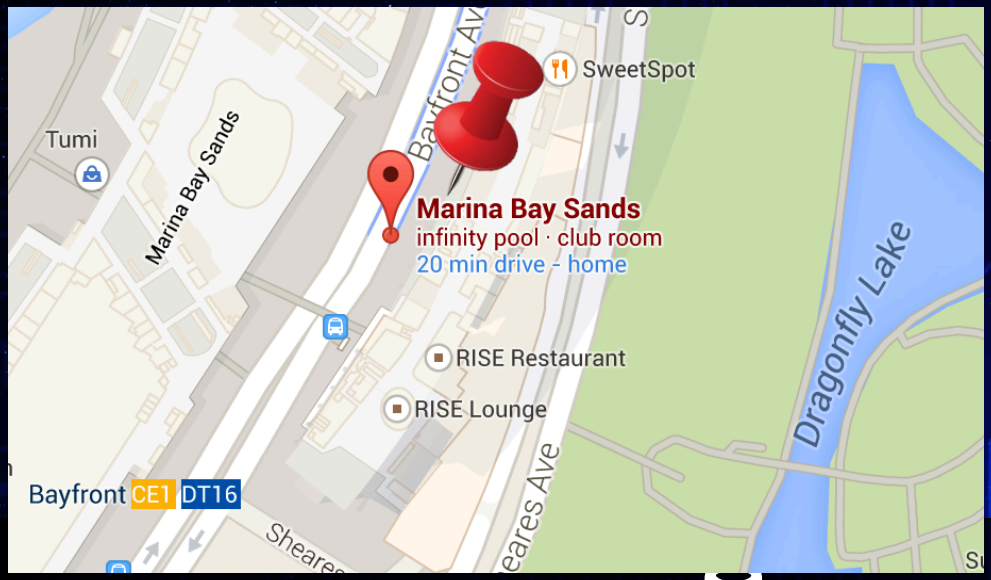
https://www.google.com.sg/maps/vt/pb=
!1m5!1m4!1i15!2i12627!3i23720!4i128!2m1!1e0!3m3!5e1105!12m1!1e47!4e0

(12627, 23720)

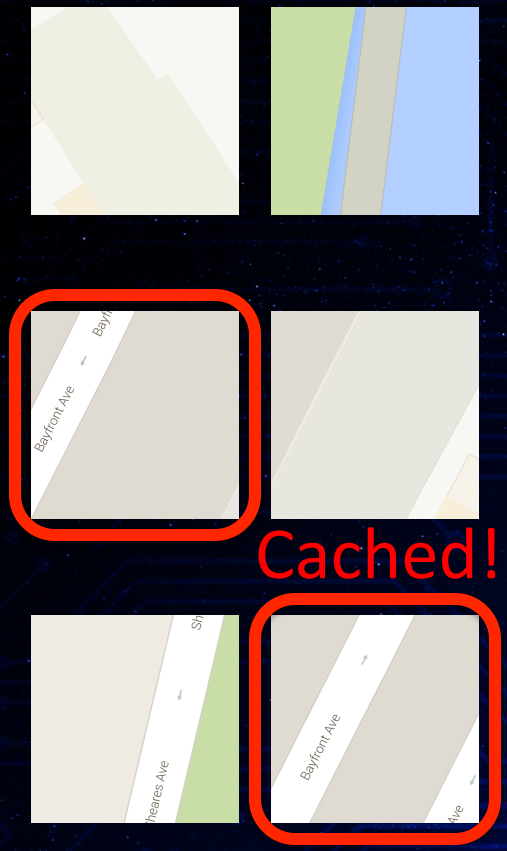
Grand Loop Rd, Yellowstone National Park, WY 82190, USA

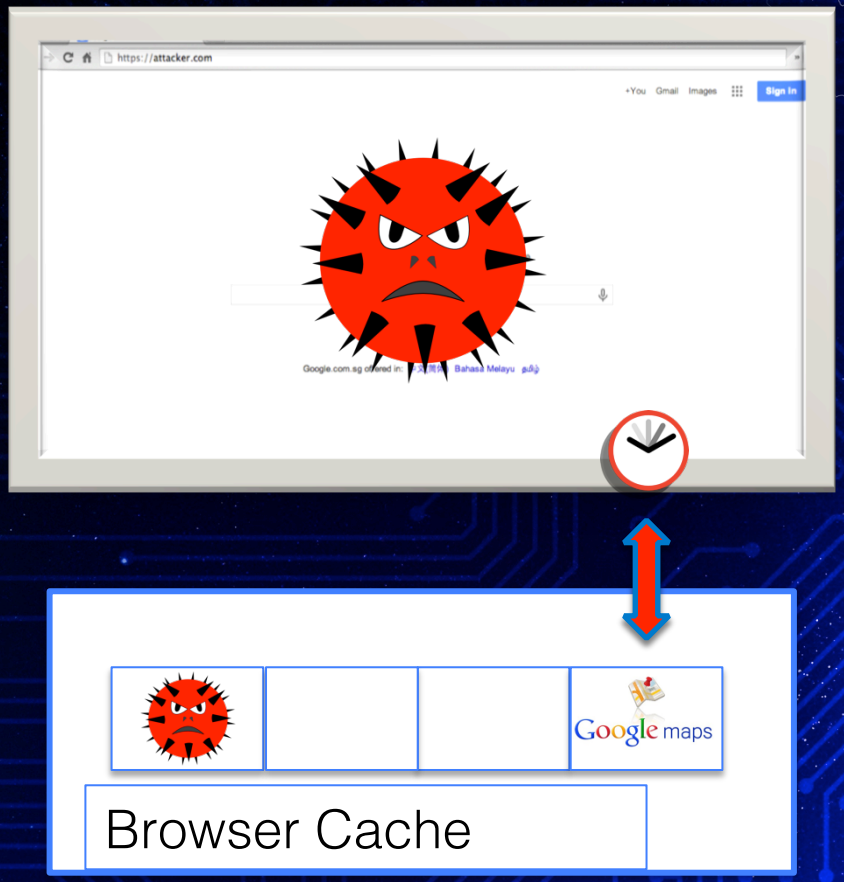


Map Tiles



How to Infer a User's Neighborhood? (II)





Evaluation

Questions to be answered:

- (Prevalence) How many websites and browsers can be utilized to conduct attacks?
- (Reliability) How big is the time difference between the loading time of resources without cache and that with cache?

Evaluation Setup

- Websites: 191 Google's sites, 100 Craigslist's sites, and 55 top Alexa sites.
- Maps: Google Maps, and other 10 map service sites.
- Browsers: Five mainstream browsers and TorBrowser
- Locations: US, UK, Australia, Singapore, and Japan.

How Many Websites and Browsers can be Utilized to Conduct Attacks?

Alexa Top Websites with Location-Related Resources

62% of 55 top Alexa global sites

singapore.craigslist.com.sg

sg.yahoo.com

www.ebay.com.sg

Map Websites with Location-Related Resources

Google maps

YAHOO! Maps

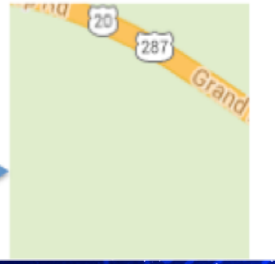
bing™ maps

All of 11 map service sites

https://www.google.com.sg/maps/vt/pb=
!1m5!1m4!1i15!2i12627!3i23720!4i128!2m1!1e0!3m3!5e1105!12m1!1e47!4e0

(12627, 23720)

Grand Loop Rd, Yellowstone National Park, WY 82190, USA



Susceptible Browsers & Platforms

Mainstream Browsers

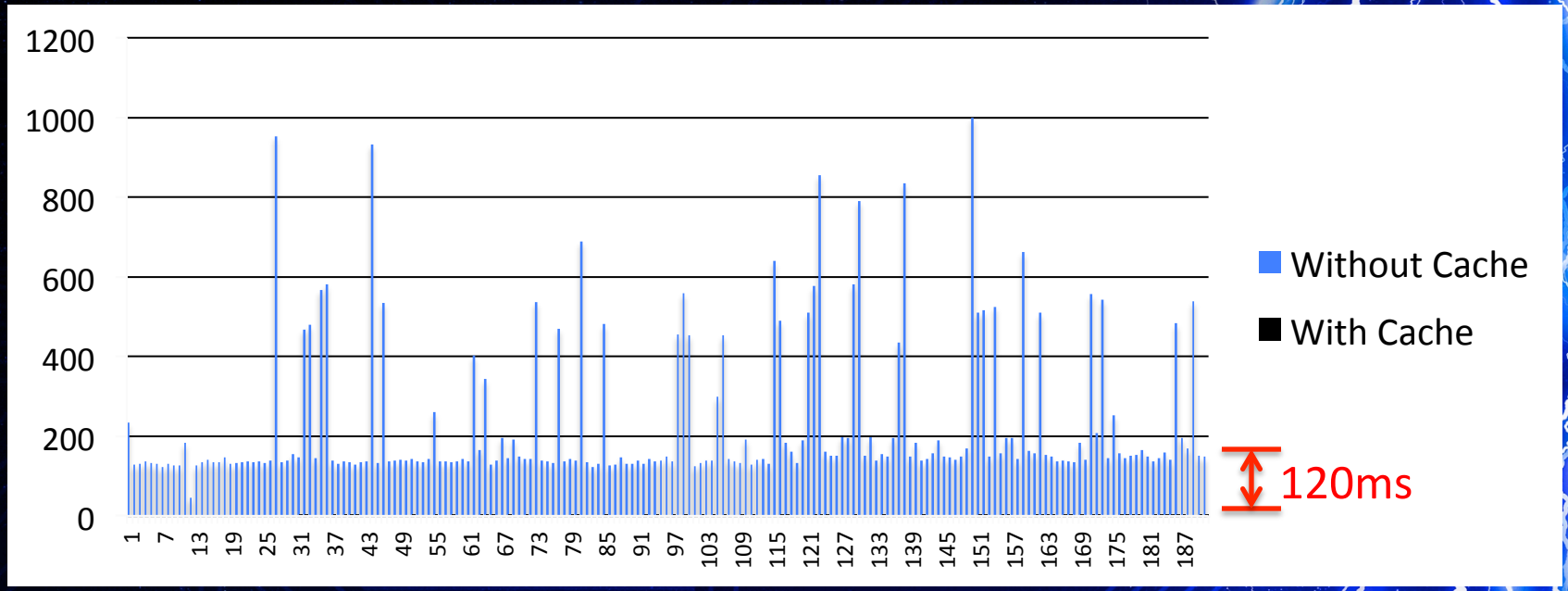
Partial

Desktop Platforms

Mobile Platforms

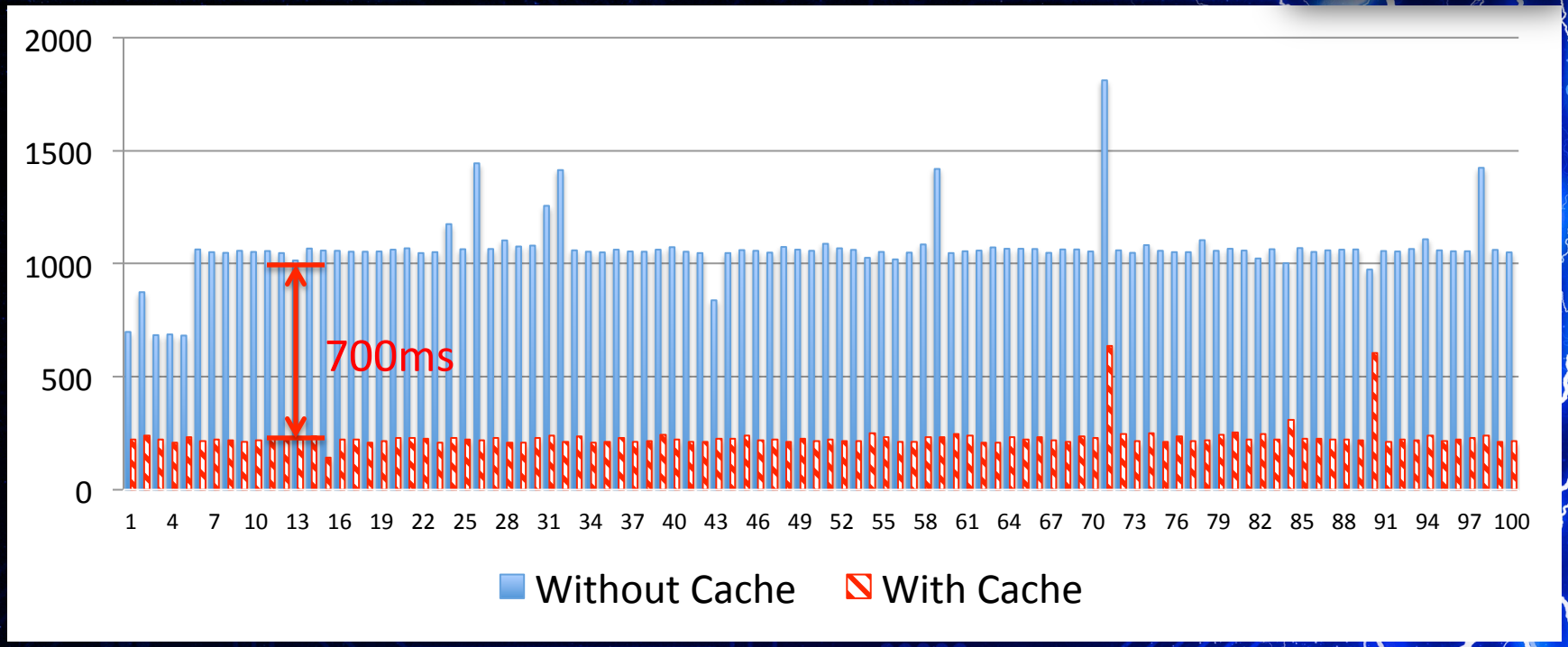
How Significant is the Time Difference between the Loading Time of Resources without Cache and that with Cache?

Loading Time： Without Cache v.s. With Cache I



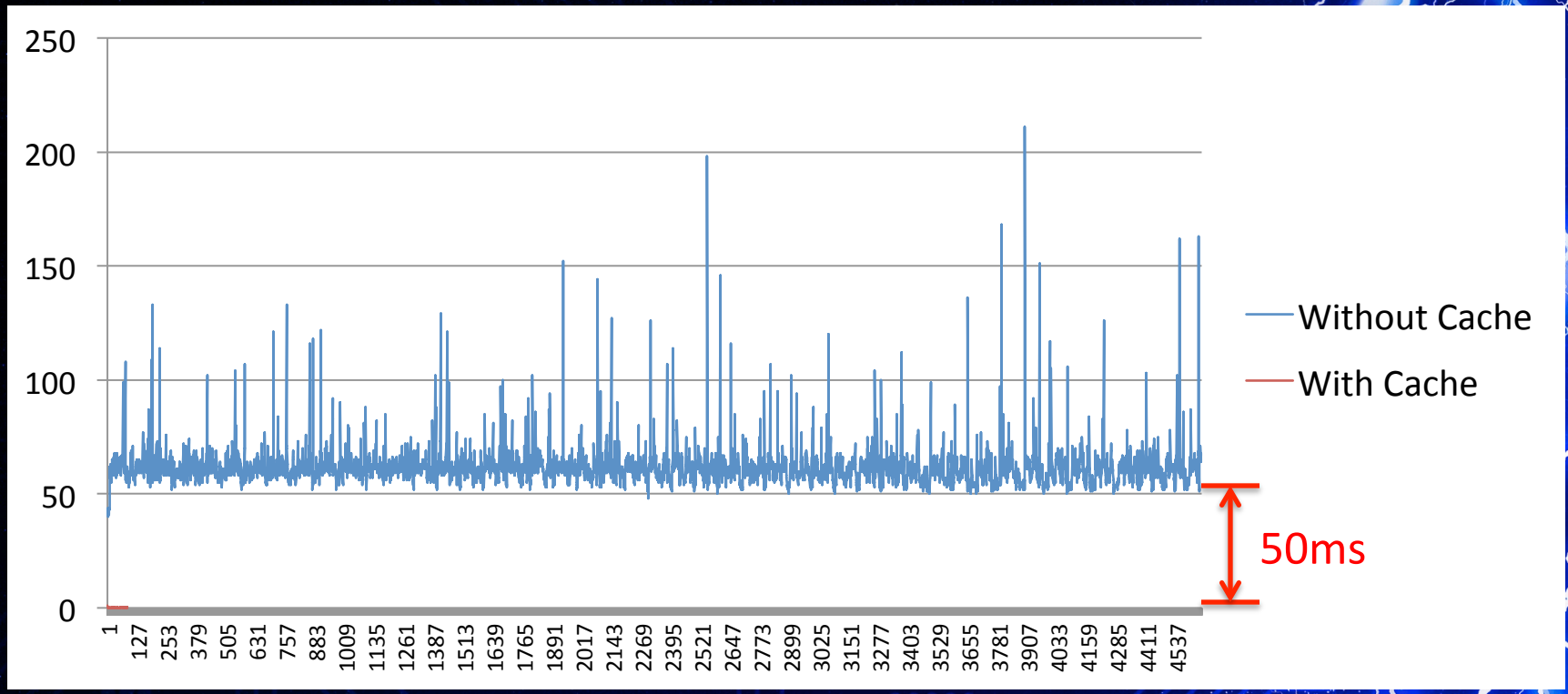
Difference in image load time (in millisecond): Without Cache (> 129 ms) v.s. With Cache (0 ~ 1 ms), for 191 Google's regional domains in Chrome on Mac OS X

Loading Time： Without Cache v.s. With Cache II



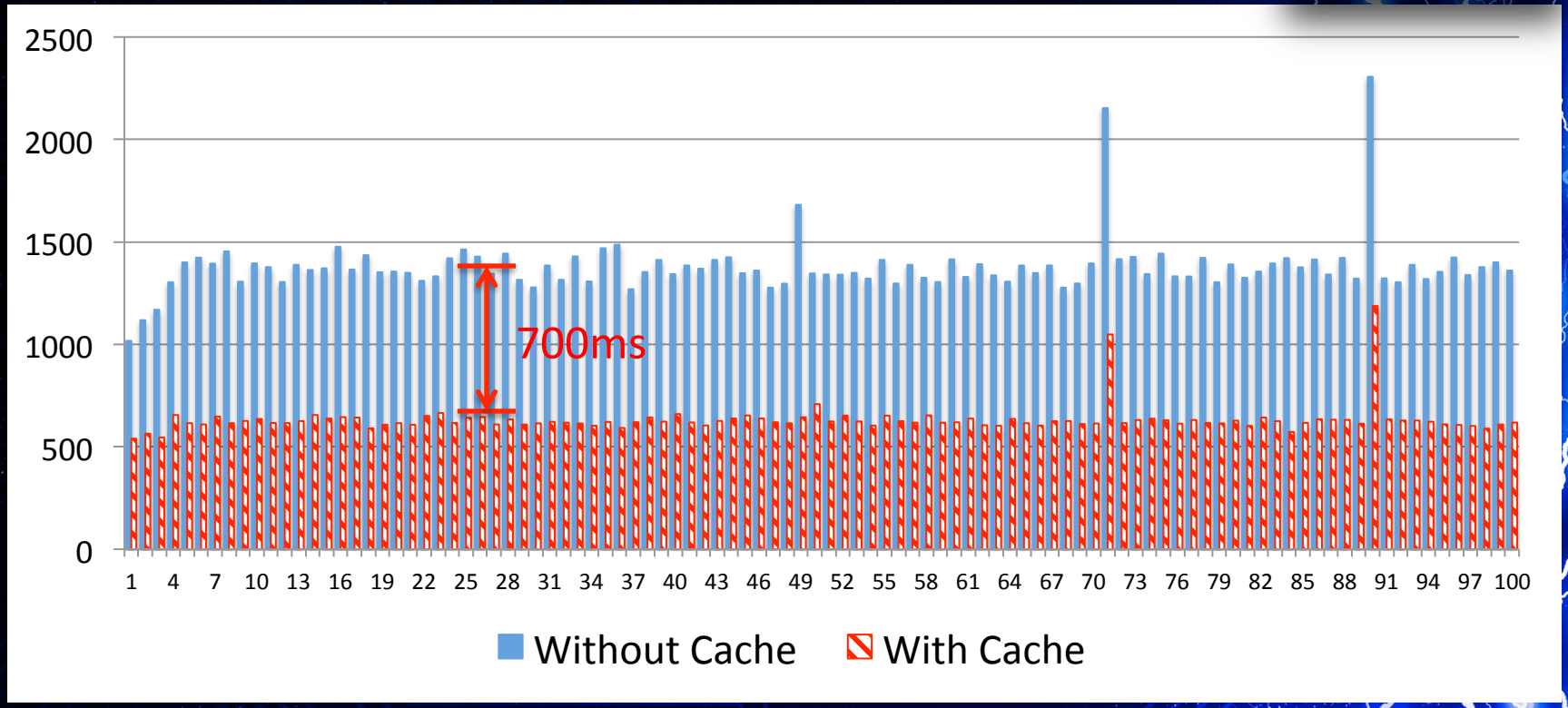
The significant difference between the page load time (in millisecond) of 100 Craigslist sites without cache (> 1000 ms) and with cache (≈ 220 ms) indicates geo-inference attacks with Craigslist

Loading Time : Without Cache v.s. With Cache III



Difference in page load time (in millisecond): Without Cache (> 50 ms) v.s. With Cache (0 ~ 1 ms), for 4,646 map tiles of New York City from Google Maps in Chrome on Mac OS X.

Loading Time (Android)



The page load time of 100 Craigslist sites on Android.

How to Protect Users from Geo-inference Attacks

black hat ASIA 2015

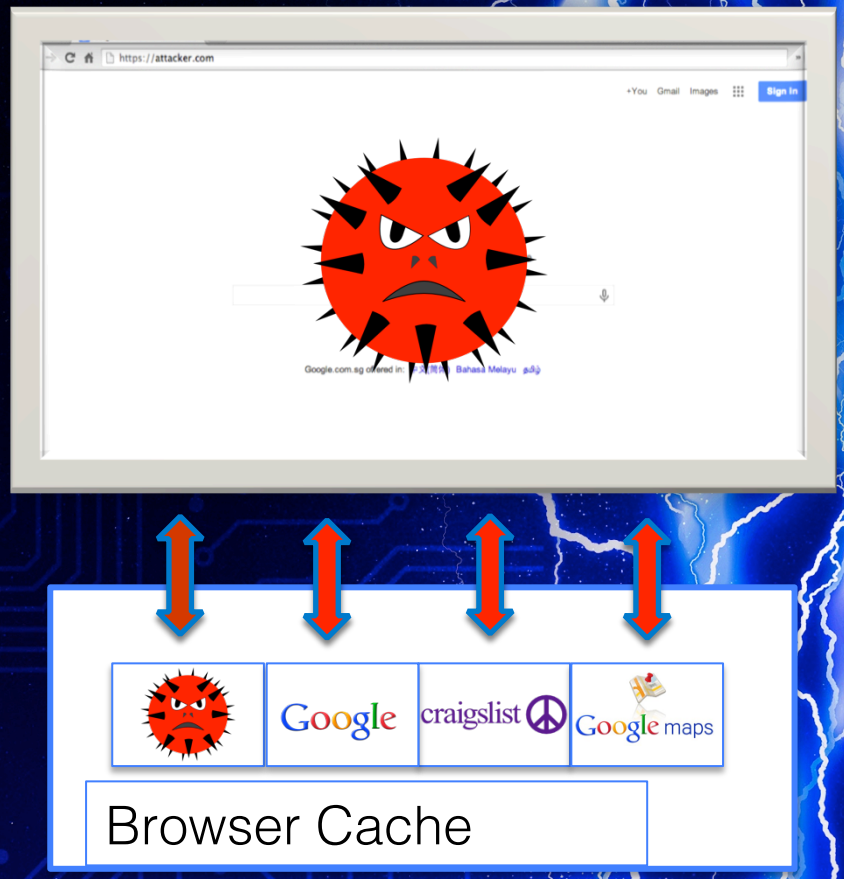
Discussion of Defense Solutions

- Private Browsing Mode
- Randomizing timing measurements
- TorBrowser and Segregating browser cache

Private Browsing Mode is not the Cure

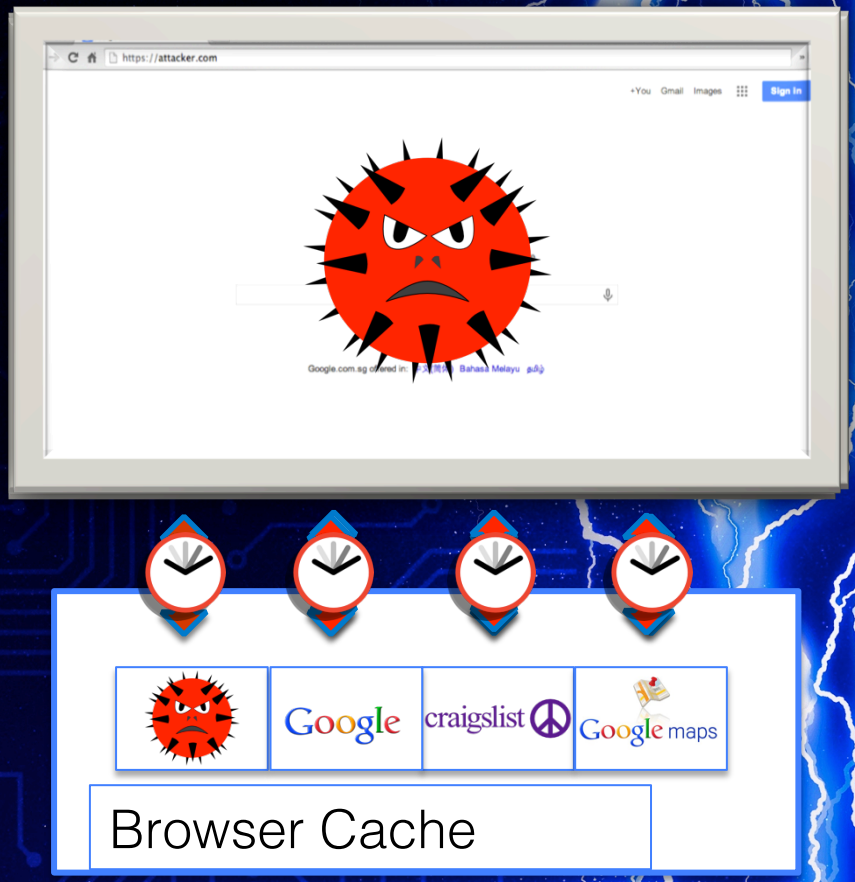
Private Browsing Mode

- Clear browser cache after closing the window.
- Disable disk cache, not the in-memory cache.
- It cannot prevent one site from inferring the user's geo-location from other sites.



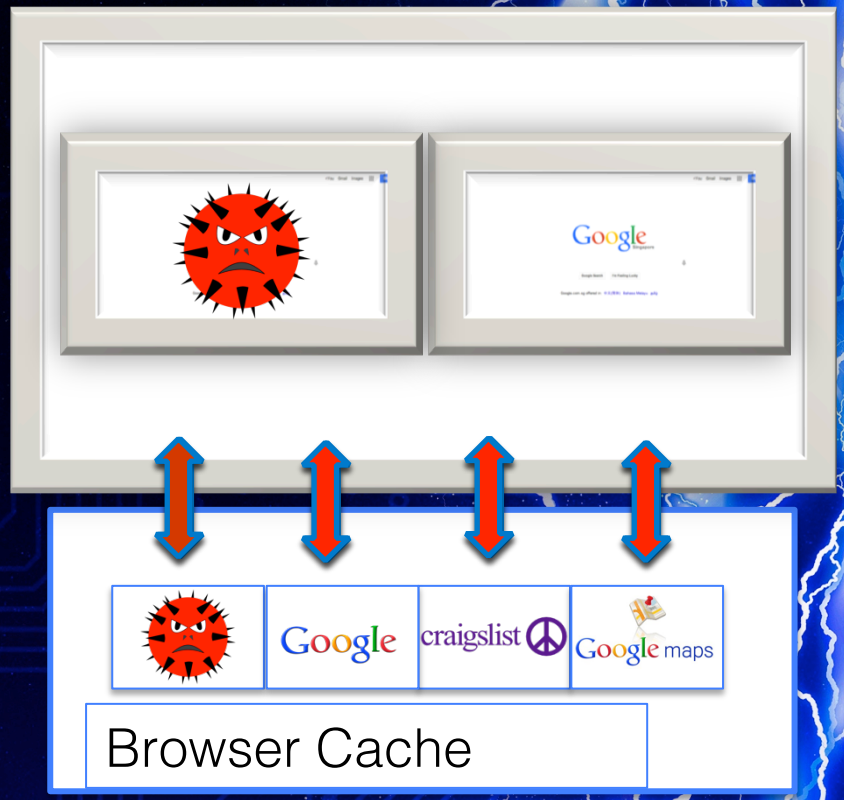
Randomizing Timing Measurements

- Add noise into timing measurement mechanisms.
- Affect web applications' functionalities
- Intricate engineering effort.



TorBrowser is not Perfect

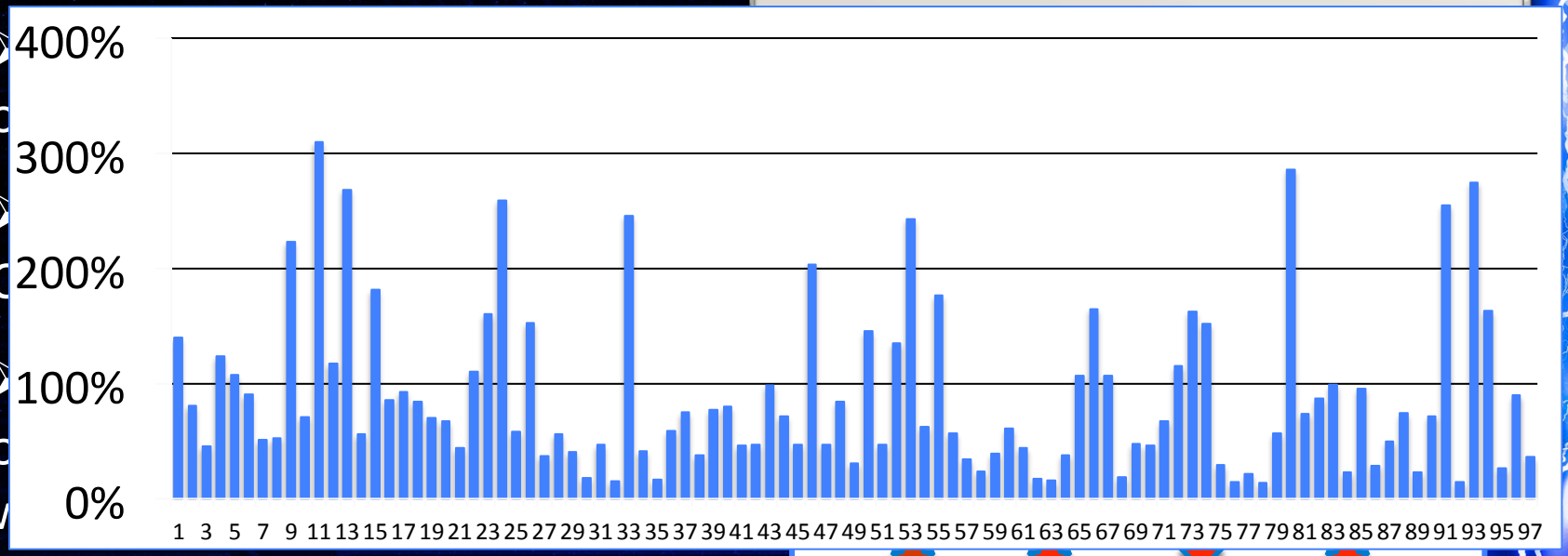
- Adds an additional "id=string" property to label every cache entry with the top-level window's domain.
- Insufficient for mashup websites, all the embedded sites in frames share the same top-level window's domain, i.e., the mashup's domain.



Demo Video

Video: Geo-inference Attacks in TorBrowser

Segregating Browser Cache



To Cache or Not To Cache?

- No cache for location-sensitive resources (0.7% to 20.7% overhead).
  - Cache-Control: no-cache for HTTP response header
- Pre-fetch redundant location-sensitive resources.
- Open challenge to design an efficient and secure caching mechanism in browsers.

Take-away

- Timing channels are still open on mainstream browsers.
- Knowing the power and prevalence of geo-inference attack (inferring country, city, neighbourhood) and be cautious about it.
- Disable cache? No JavaScript?
- Never give additional permissions to unfamiliar sites or open it for a long time.
- Clear cache before and after visiting a site with your private information, e.g., online banking site.

Yaoqi JIA
E-mail: jiayaoqi@comp.nus.edu.sg

References

D. Akhawe, A. Barth, P. E. Lam, J. Mitchell, and D. Song, "Towards a formal foundation of web security," in *Computer Security Foundations Symposium (CSF), 2010 23rd IEEE*, 2010.

A. Bortz and D. Boneh, "Exposing private information by timing web applications," in *Proceedings of the 16th international conference on World Wide Web*, 2007.

G. Wondracek, T. Holz, E. Kirda, and C. Kruegel, "A practical attack to de-anonymize social network users," in *Security and Privacy (SP), 2010 IEEE Symposium on*, 2010.

Z. Weinberg, E. Y. Chen, P. R. Jayaraman, and C. Jackson, "I still know what you visited last summer: Leaking browsing history via user interaction and side channel attacks," in *Security and Privacy (SP), 2011 IEEE Symposium on*, 2011.

M. Jakobsson and S. Stamm, "Invasive browser sniffing and countermeasures," in *Proceedings of the 15th international conference on World Wide Web*, 2006.

G. Aggarwal, E. Bursztein, C. Jackson, and D. Boneh, "An analysis of private browsing modes in modern browsers," in *Proceedings of the 19th USENIX Conference on Security*, ser. USENIX Security'10, 2010.