Imperial College London

CO331 – Network and Web Security

21. Web user tracking

Dr Sergio Maffeis
Department of Computing
Course web page: http://www.doc.ic.ac.uk/~maffeis/331

Web tracking

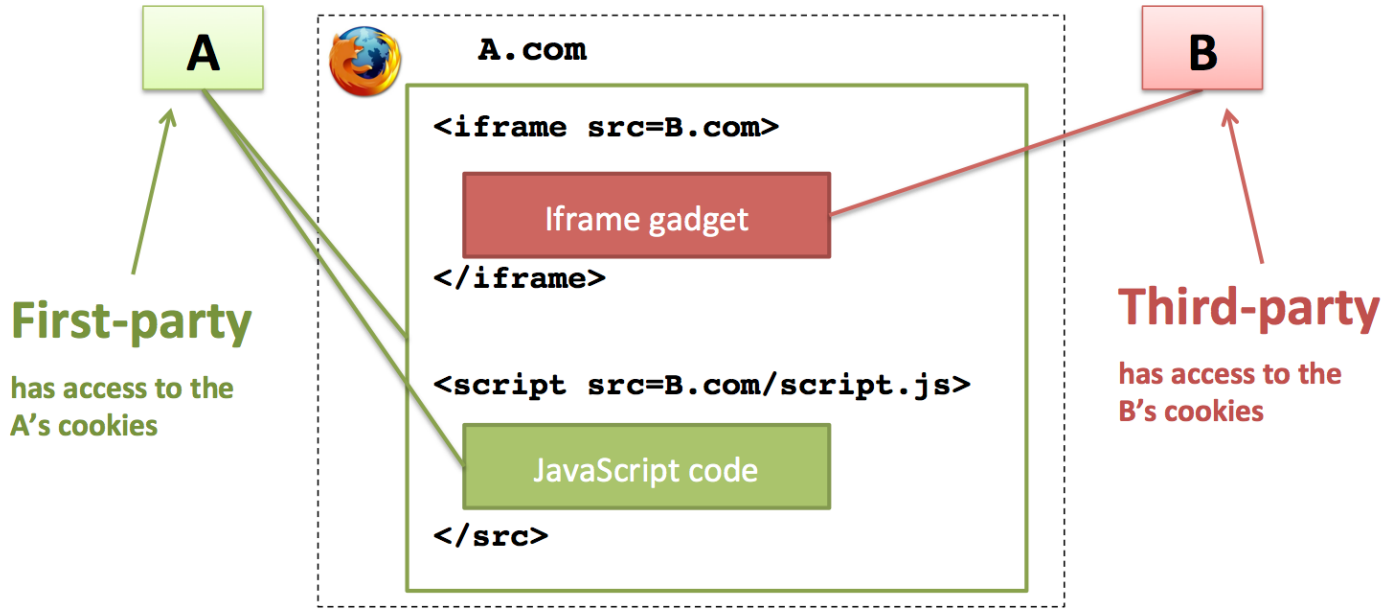
Imperial College London

- Examples (spot the controversial cases…)
  - User-specific website settings maintain a consistent browsing experience across a sequence of related HTTP requests and responses
  - Secure session recognize requests coming from the same user, that has already been authenticated, and provide privileged access
  - Browsing history, personal preferences and demographic data are harvested by marketers to profile users and provide “relevant advertising”
  - User presence online on different devices is correlated by governments to identify individuals
- Tracking is a complex and pervasive issue
  - 1st party trackers
    - iframes and scripts on origin of website visited by the user
  - 3rd party trackers
    - Cross-domain iframes and their resources, included by visited websites
  - There are *legitimate* and *illegitimate* usages
    - Do not necessarily coincide with *desirable* and *undesirable* usages
  - Can happen across devices

Browser support for tracking

Imperial College London

- **Trackers need to store information in the browser about the user**
- Cookies: again 1st/3rd party distinction



Browser support for tracking

Imperial College London

- **Trackers need to store information in the browser about the user**
- Supercookies
  - HTML5 storage: Local/sessionStorage, Web SQL, IndexedDB …
  - Plugin storage: Flash Local Shared Objects
- Cache
  - Tell browser to cache a script that assigns state to a variable
  - Response header
    - `Cache-Control: private, Max-Age=31536000`
  - Resource name can be change to reset state `state.js?v=1.x`
- ETag header
  - Originally intended for cache validation
  - Can be used like a simplified cookie
  - Response header sets ETAG value
    - `ETag: "A23C42BF890DFE"`
  - Subsequent request header reflects the value
    - `If-none-match: "A23C42BF890DFE"`
  - Other request/response header pairs similar to Etag
    - `If-Modified-Since/Last-Modified`

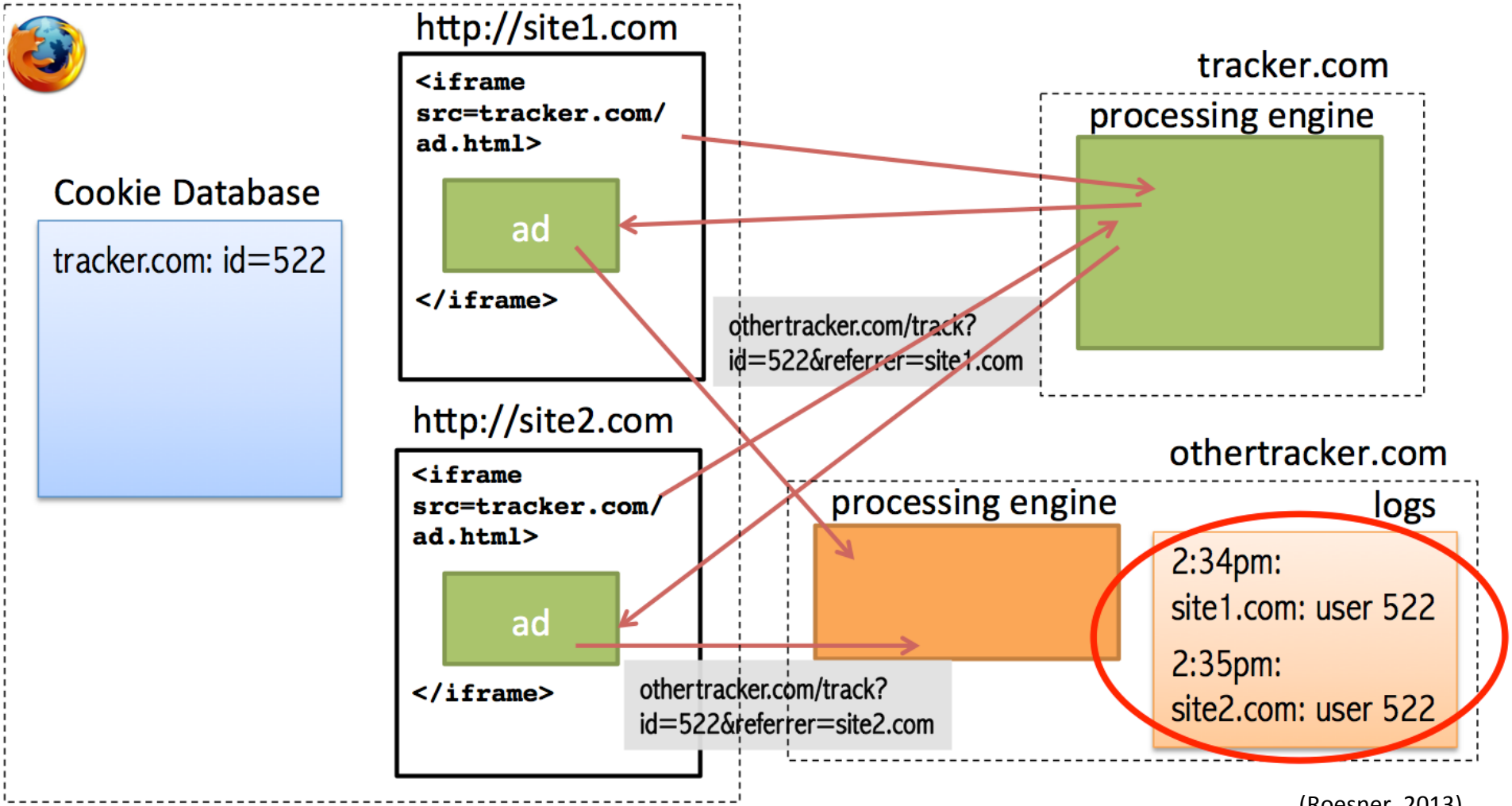
Browser support for tracking

Imperial College London

- **Trackers need ways to send user information back to the server**
- HTTP request and responses
  - Explicit communication
    - User clicks on a link
    - Loading of page resources (iframes, images, etc.)
    - AJAX and JavaScript-triggered page loading or navigation events
  - Implicit communication
    - W3C Beacons
      - "*asynchronous and non-blocking delivery of data that minimizes resource contention with other time-critical operations*"
      - `navigator.sendBeacon('/collector', data);`
    - In Chrome, opening a new tab sends a `new_tab` request to Google
    - Search bar may send in the background one request per character you type
    - *Pre-rendering*
      - Browser loads resources linked on current page in case you later click
- Other Plugin communications
  - Flash, Java, Active X controls can use sockets

Example: cross-site tracking

Imperial College London



(Roesner, 2013)

Browser countermeasures

Imperial College London

- Do Not Track header
  - W3C Tracking Protection Working Group's brainchild
  - Request header: `DNT : 1`
  - Mostly interpreted as *do not target the users based on collected data*
    - Data is still collected
- Private browsing/Incognito mode
  - Prevents caching, history, cookies, preferences
  - A bit of a drastic solution
  - A lot can still be achieved using JavaScript, side-channels, ETag header, etc.
    - "*Going incognito doesn't hide your browsing from your employer, your internet service provider, or the websites you visit.*"
- Block 3rd party cookies in browser settings
  - 3rd party trackers cannot set cookies
  - In some browsers, 3rd party cookies are still sent if they already exists
  - A 3rd party iframe can open a popup which is now a 1st-party cookie setter

Zombie cookies

Imperial College London

- Deleted your cookies?
- Tracking data saved in other headers, cache or supercookies can be used to resuscitate them!
  - *Cookie respawning,* aka ***Zombie cookies***
  - In fact, who needs cookies if you have JavaScript + localStorage?
- Cleared also cache and local/sessionStorage?
  - Respawning via Flash cookies (LSOs)
  - Thanks to Flash, zombie cookies can migrate across browsers!
    - (LSOs can be shared by various Flash plugin instances)
- Key role of **fingerprinting** in tracking
  - Respawn tracking data associated to a known fingerprint even if browser and Flash data was reset

KISSmetrics and Hulu got sued for that trick in 2011

More countermeasures

Imperial College London

- Disable plugins (eradicate Flash)
- Disable JavaScript (stop browsing?)
- Use the TOR Browser
- Use anti-fingerprinting countermeasures
- Install anti-tracking extensions: Ghostery, AdBlock+, Privacy Badger, ShareMeNot…

Research on tracking

- Anti-tracking extensions use blacklists to stop requests to tracking websites
  - How to automatically populate such blacklists?
  - Ongoing research on machine learning techniques to identify trackers
- Trackers leave a trail of information visible from the browser
  - Ongoing research on data analytic techniques to spot tracking patterns
  - Monitorito browser extension (distinguished student project)
    - Monitor and visualise network events in real time
    - Identify trackers using graph analytics