Control What You Include!

Server-Side Protection against Third Party Web Tracking

Dolière Francis Somé, Nataliia Bielova, Tamara Rezk

Inria Rennes - 10th May, 2017

thanks to ePrivacy directive 2009

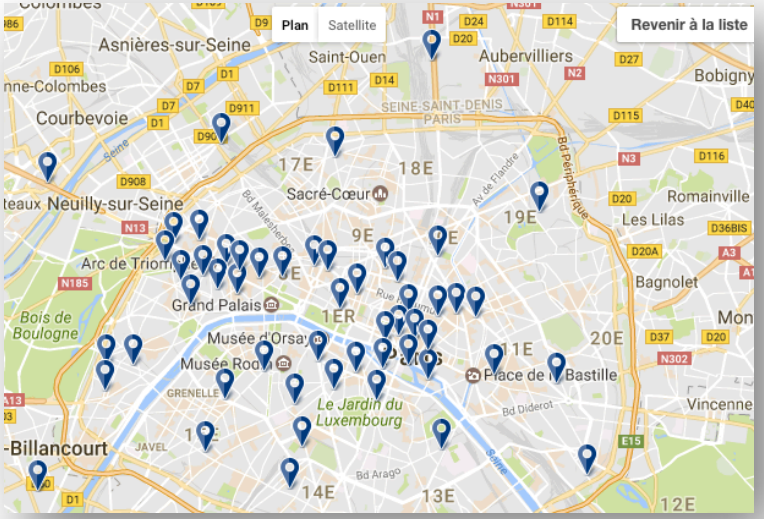
Third party in websites & tracking

- Up to 34 distinct third parties on a single website
- 90% of content is tracking users

- Unintentional tracking

Why web developers include so many third party contents?

Functionality ☺ Privacy ☹



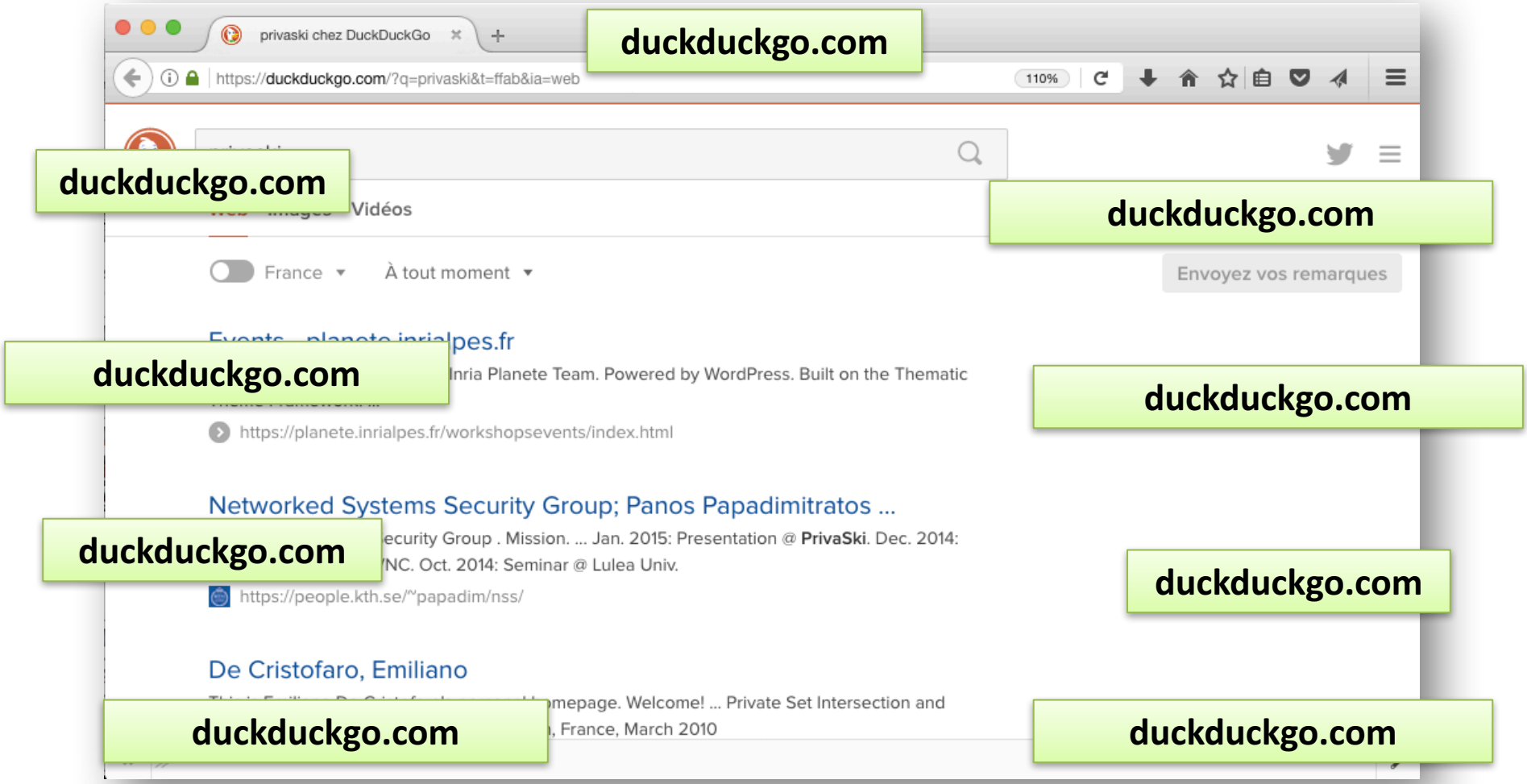
jQuery
write less, do more.

Mitigating tracking

- Many client side solutions for users to protect themselves
  - Browser extensions
- (Self) regulations and standards

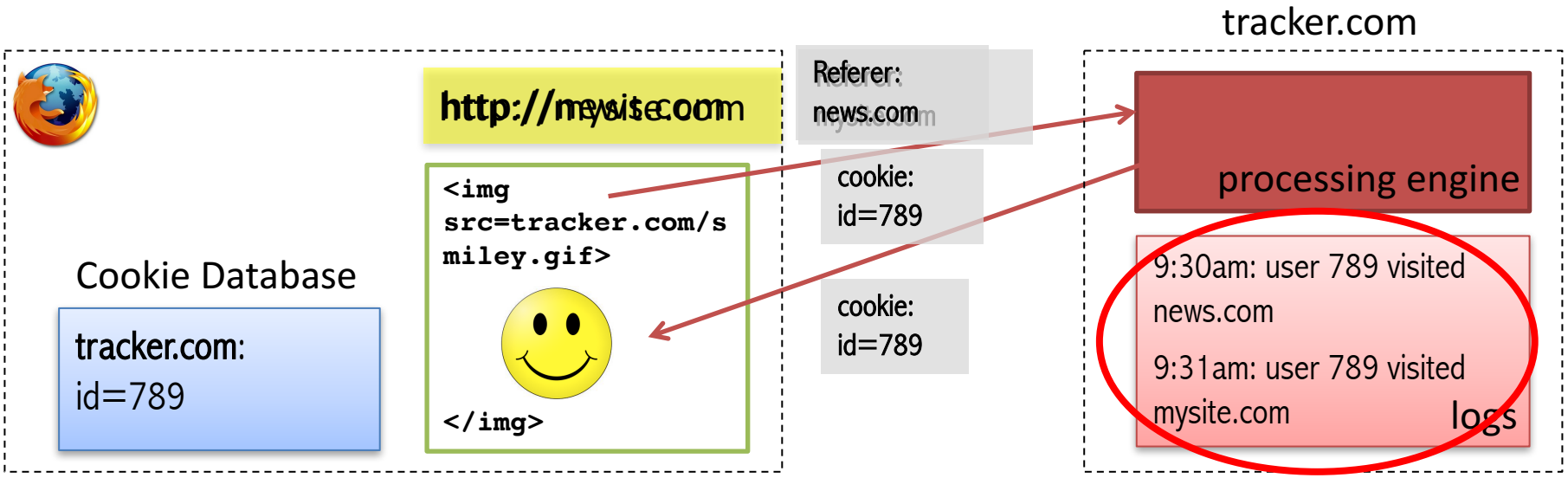
- No third parties ?
  - Tremendeous functionalities.
  - New ways to embed third parties !
- ePrivacy update [1]: website owners are liable if third parties track their users

Privacy ☺


privaski chez DuckDuckGo
https://duckduckgo.com/?q=privaski&t=ffab&ia=web
110%
Vidéos
France
À tout moment
Envoyez vos remarques
https://planete.inrialpes.fr/workshopsevents/index.html
Networked Systems Security Group; Panos Papadimitratos ...
https://people.kth.se/~papadim/nss/
De Cristofaro, Emiliano
duckduckgo.com
duckduckgo.com
duckduckgo.com
duckduckgo.com
duckduckgo.com
duckduckgo.com
duckduckgo.com

How can developers include third party content and guarantee privacy?

Third party tracking via cookies

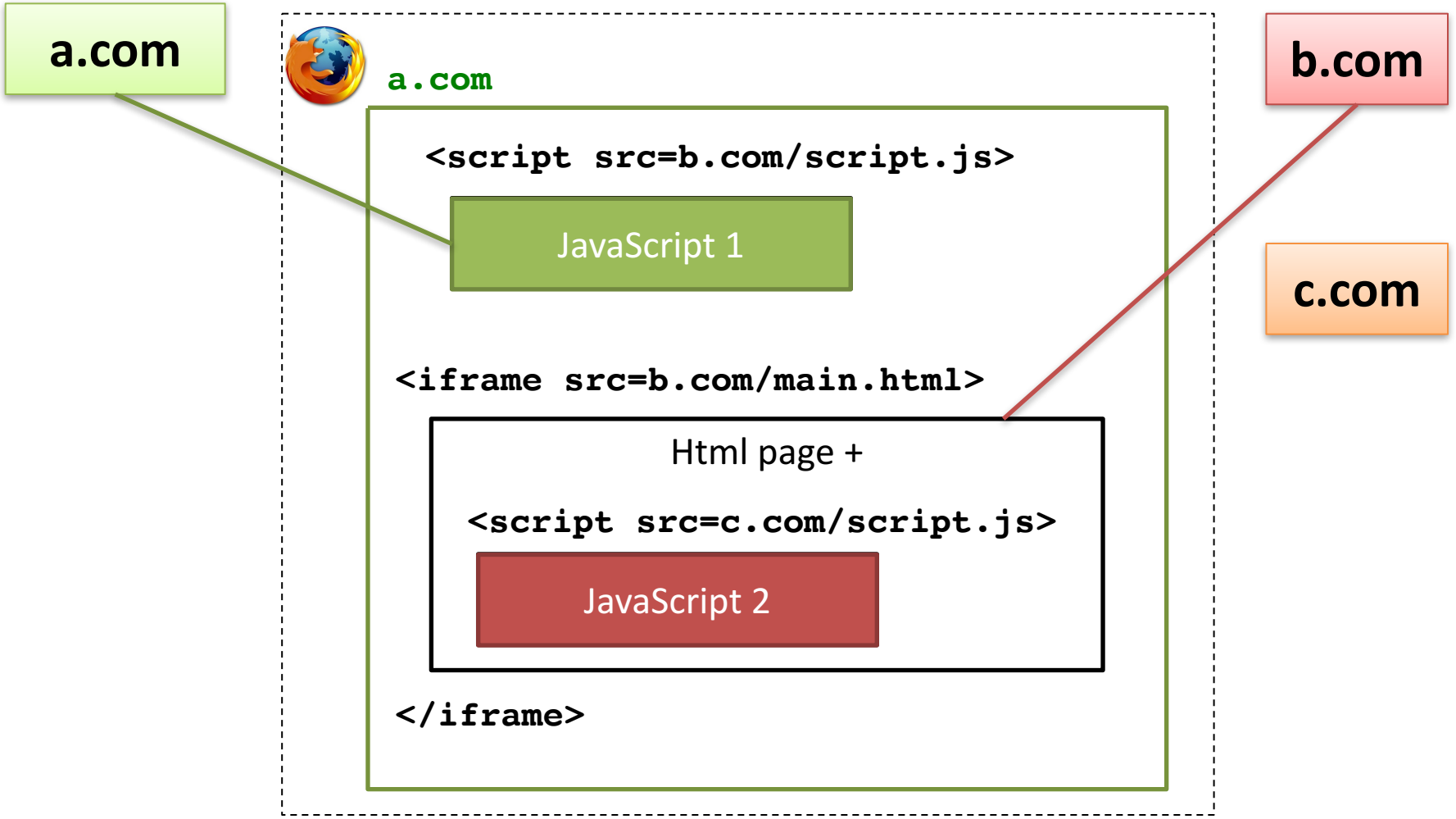


Mechanisms Required By Trackers

- Ability to create/store user identity in the browser and communicate it back to tracker
  - HTTP cookies, browser cache, browser Storages
  - device fingerprinting

- Ability to communicate website visited back to the tracker
  - HTTP Referer header
  - `APIs: window.location, document.URL, document.referrer`

In what context each content is running?



Which third party content is controllable?

	HTML Tags	Third Party Content
controllable **in-context**	<link>	Stylesheets
	<img>	Images
	<audio>	Audios
	<video>	Videos
	<form>	Forms
	<script>	Scripts
not controllable **cross-context**	<(i)frame>, <frameset>, <a>	Web pages
	<object>, <embed>, <applet>	Plugins and Web pages

Table 1. Third party content and execution context.

Tracking capabilities

	User Recognition		Website Identification	
	Passive	Active	Passive	Active

in-context	HTTP cookies
	Cache-Control
	Etag
	Last-Modified
		-	Referer
			Origin
				`document.URL`
				`document.location`
				`window.location`

cross-context		Flash LSOs
		`document.cookie`
		`window.localStorage`
		`window.indexedDB`
			Referer	`document.referrer`

Fig. 2. Stateful tracking mechanisms

Privacy-preserving web architecture

Goal

- Remove tracking from functional third-party content

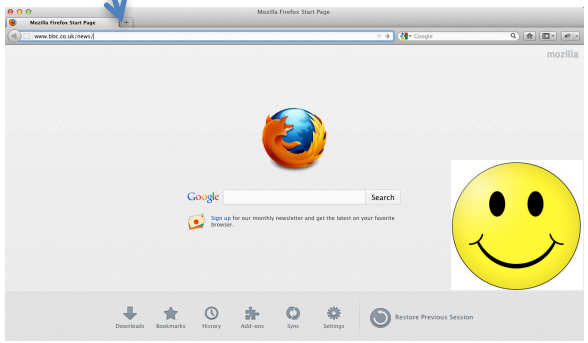
Idea

- Rewrite static third-party content
- Redirect dynamic third-party content
- Restrict communication between third-parties within the application

Current web architecture

http://mysite.com

Web browser



Referer:
mysite.com

cookie:
id=789

tracker.com

mysite.com

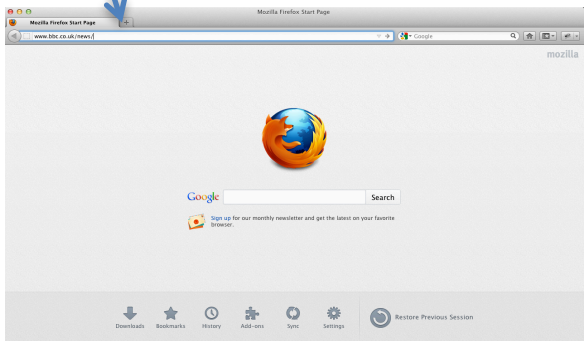
Our architecture

tracker.com

http://mysite.com

Redirect third parties to middle.com
Intercept dynamically created in-context content
Add CSP (to avoid bypassing)

Web browser



rewrite.com

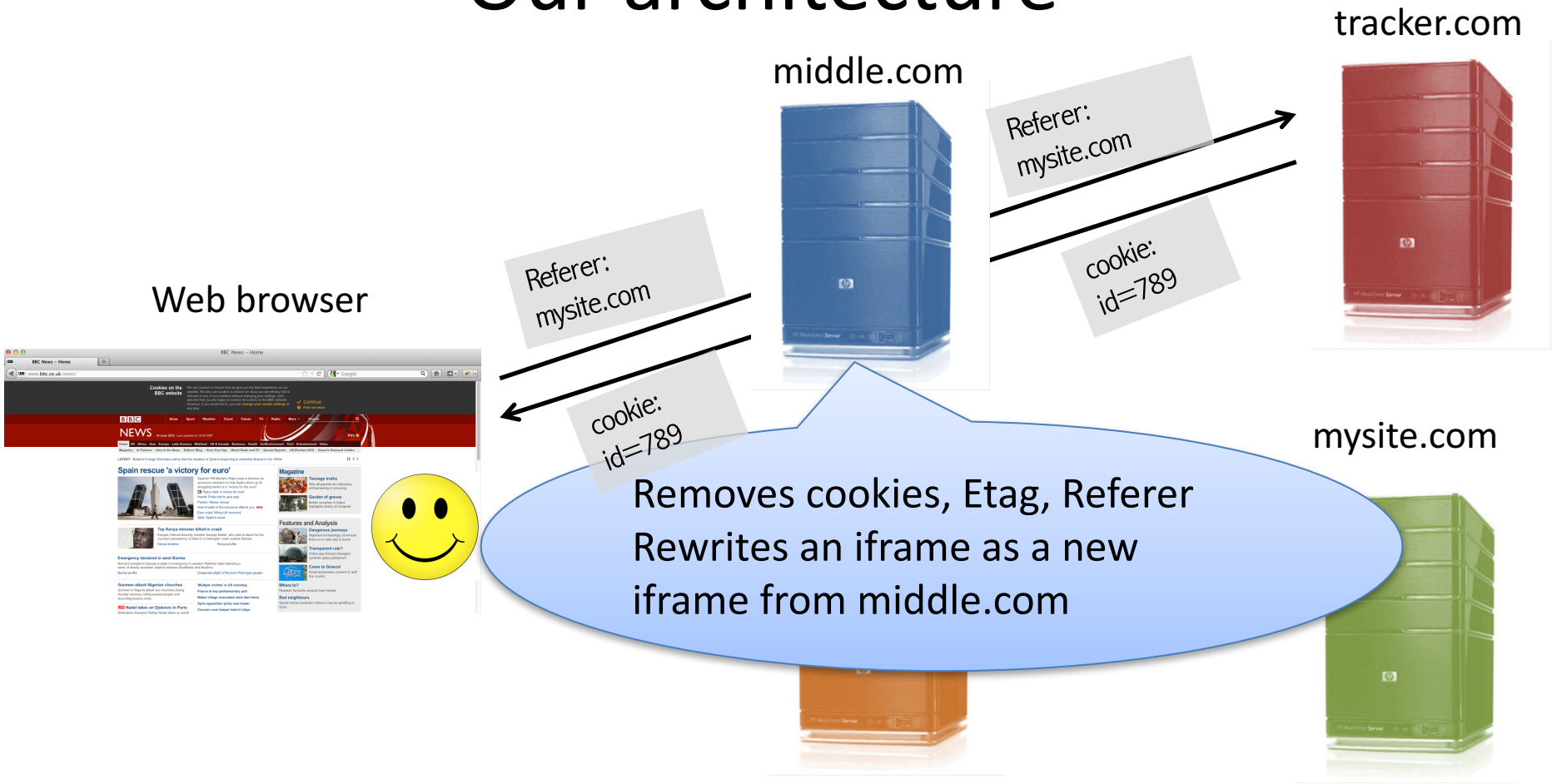
mysite.com

```
http://tracker.com/smiley.gif →
http://middle.com/?src=http://tracker.com/smiley.gif
```

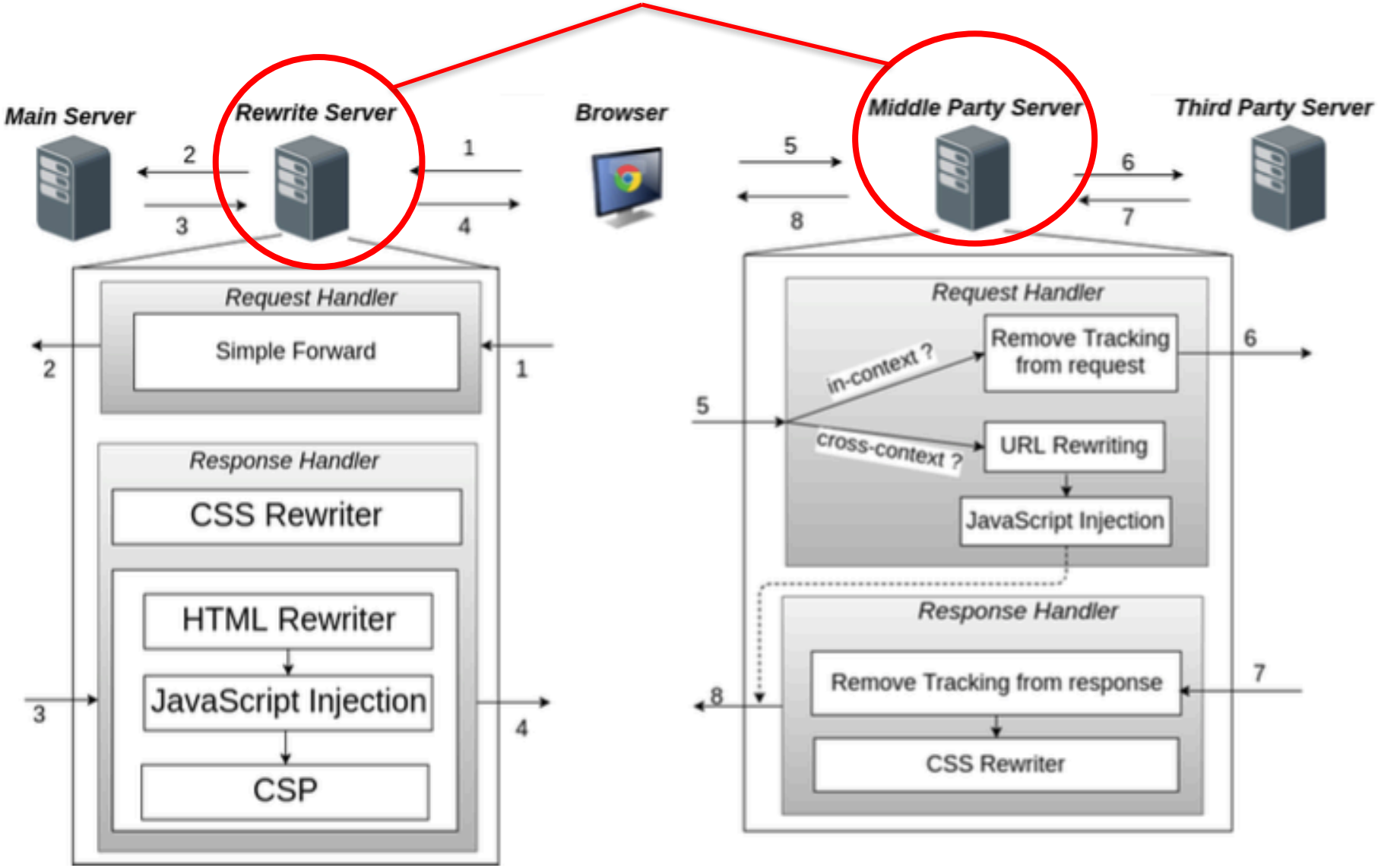
Content-Security-Policy: default-src 'self' 'middle.com';

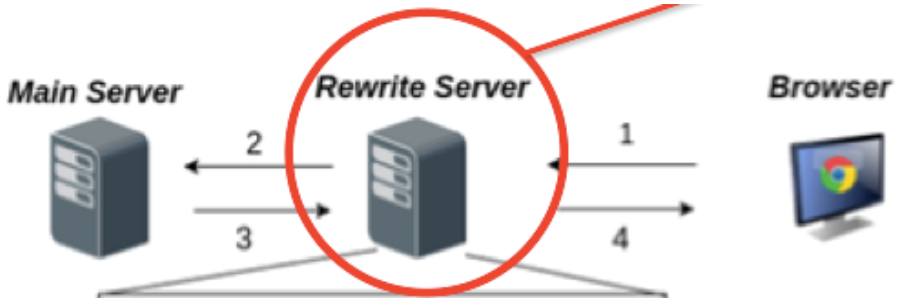
object-src 'self';

Our architecture



Controlled by the website owner

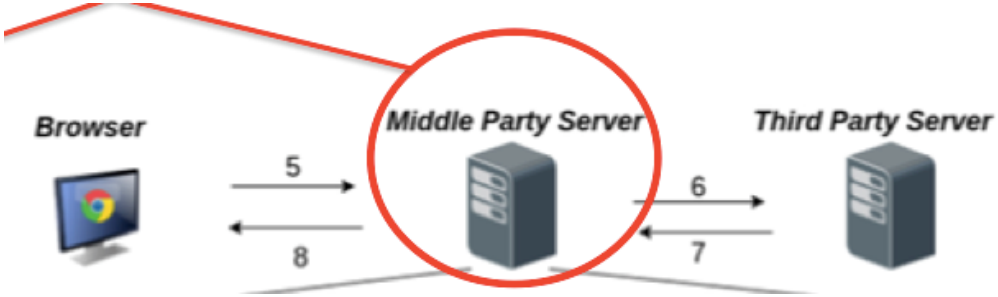




- HTML Rewriter

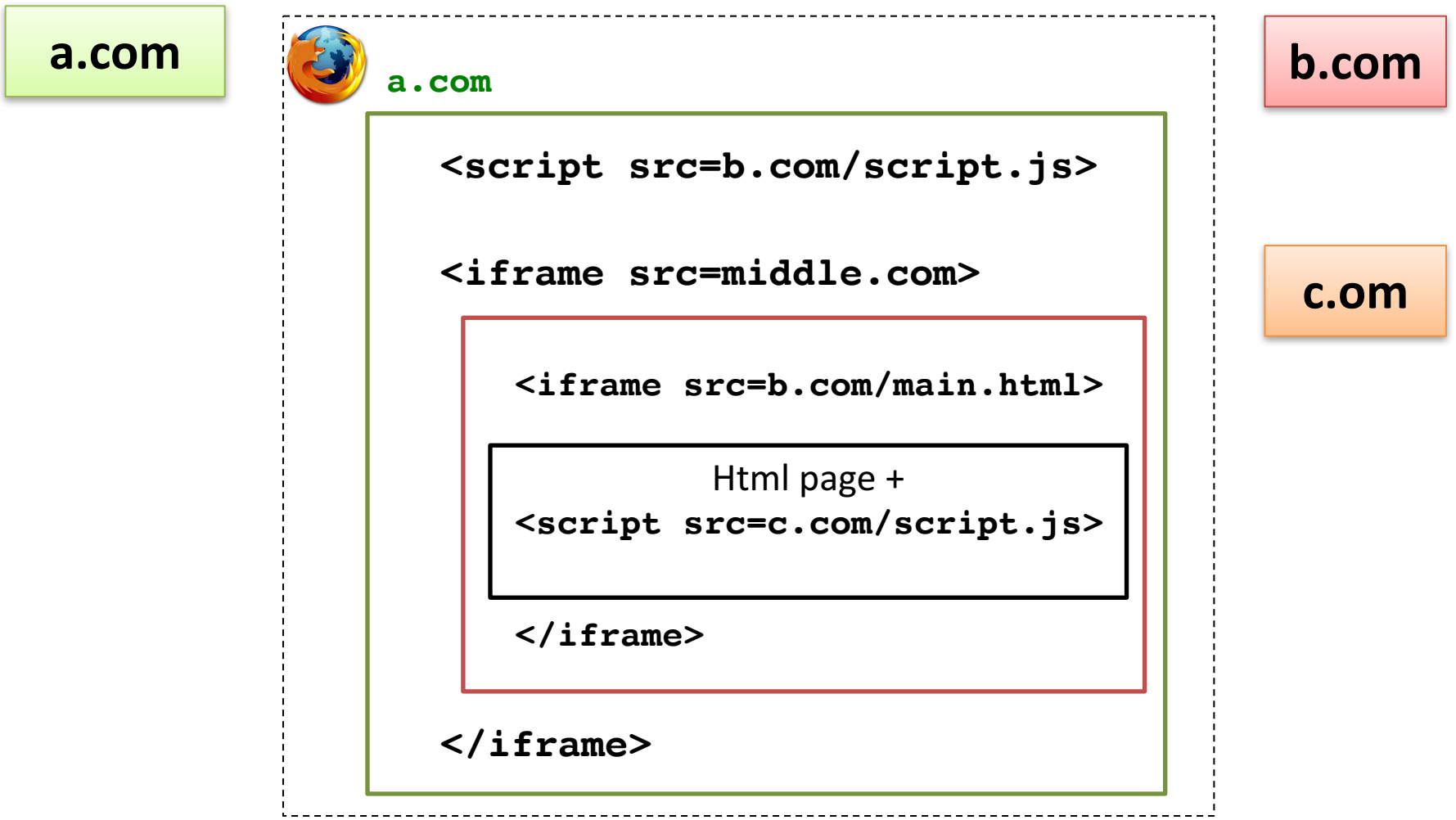
```
http://third.com/script.js
→
http://middle.com/?src=http://third.com/script.js
```

- JavaScript injected to intercept dynamically created content
- Content Security Policy (CSP) injected to avoid bypassing



- Removes tracking HTTP headers
  - Cookies, Etag, Referer,…
- Rewrites cross-context content
  - Prevents browser from sending Referer header
  - Disables document.referer
- Disables cross-context communication by placing third party content in isolated iframes

Handling Cross-Context contents



Case study & conclusions

- All websites work properly
  - Demo website with youtube videos, google maps, and various contents from third parties
  - News: www.bbc.com
  - Movies: www.imdb.com
  - Shopping: http://vertbaudet.fr

- **Our architecture**
  - for website developers
  - allows to embed third party contents
  - while preserving users privacy

Future Work

- Performance evaluation
- Compatibility on various browsers
  - Tested on Firefox, Google Chrome, Chromium, Safari, Opera
- Test implementation with real developpers to improve
  - Developper own contents such as CDNs that require HTTP Referrer, cookies

Thanks !