Control What You Include!

Server-Side Protection against Third Party Web Tracking

Dolière Francis Somé, Nataliia Bielova, Tamara Rezk

Privaski 2017

thanks to ePrivacy directive 2009

Third party content on websites

Today

- up to 34 distinct third parties on a single website
- 90% of content is tracking users
- Users protect themselves with browser extensions
  - Ghostery, Disconnect, etc.

Tomorrow

- ePrivacy update [1]: website owners are liable if third parties track their users

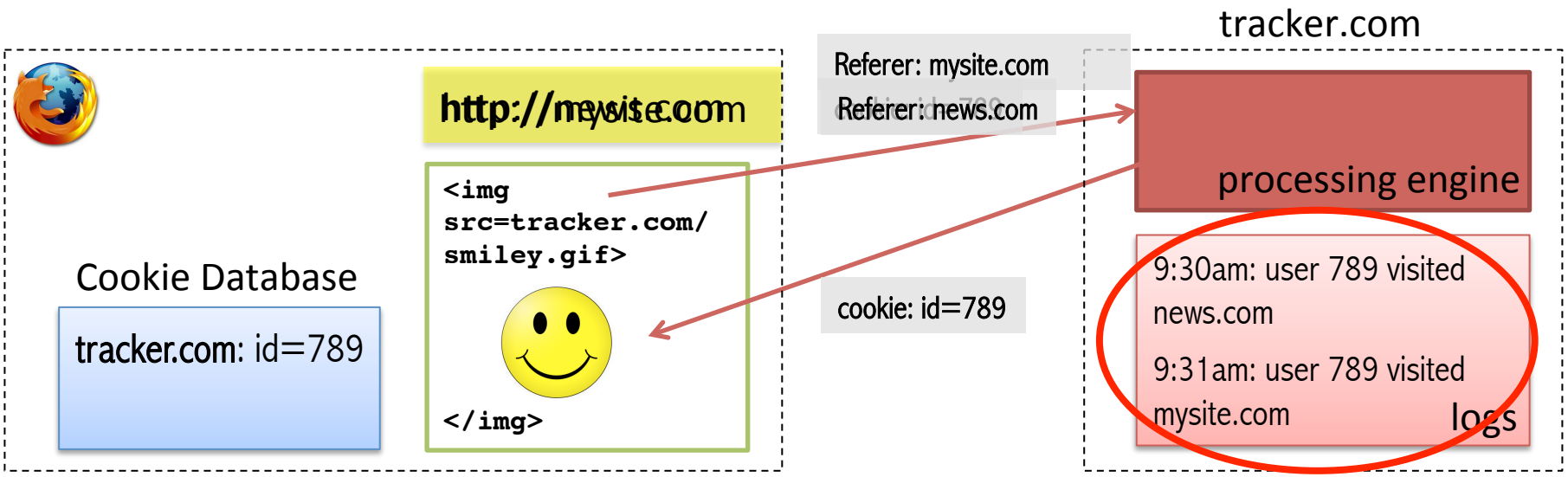
  => Website owners want to control third party content they include

Mechanisms Required By Trackers

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

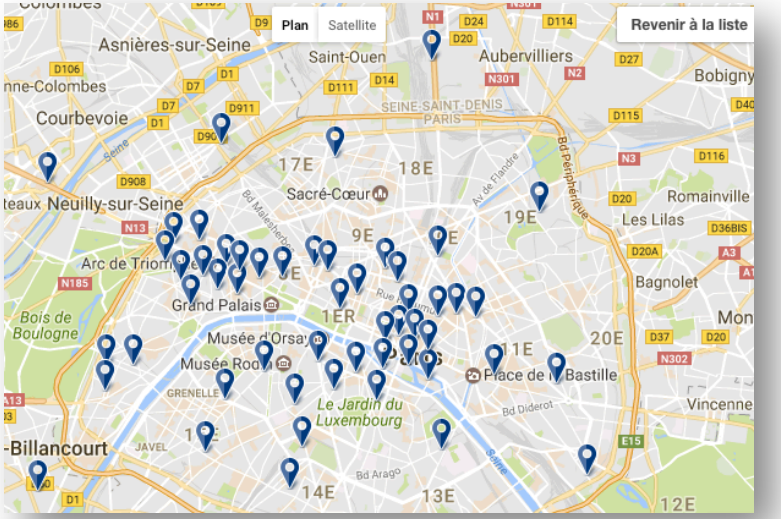
- Ability to communicate website visited back to the tracker
  - HTTP Referer header
  - `document.referrer`

Third party tracking via cookies



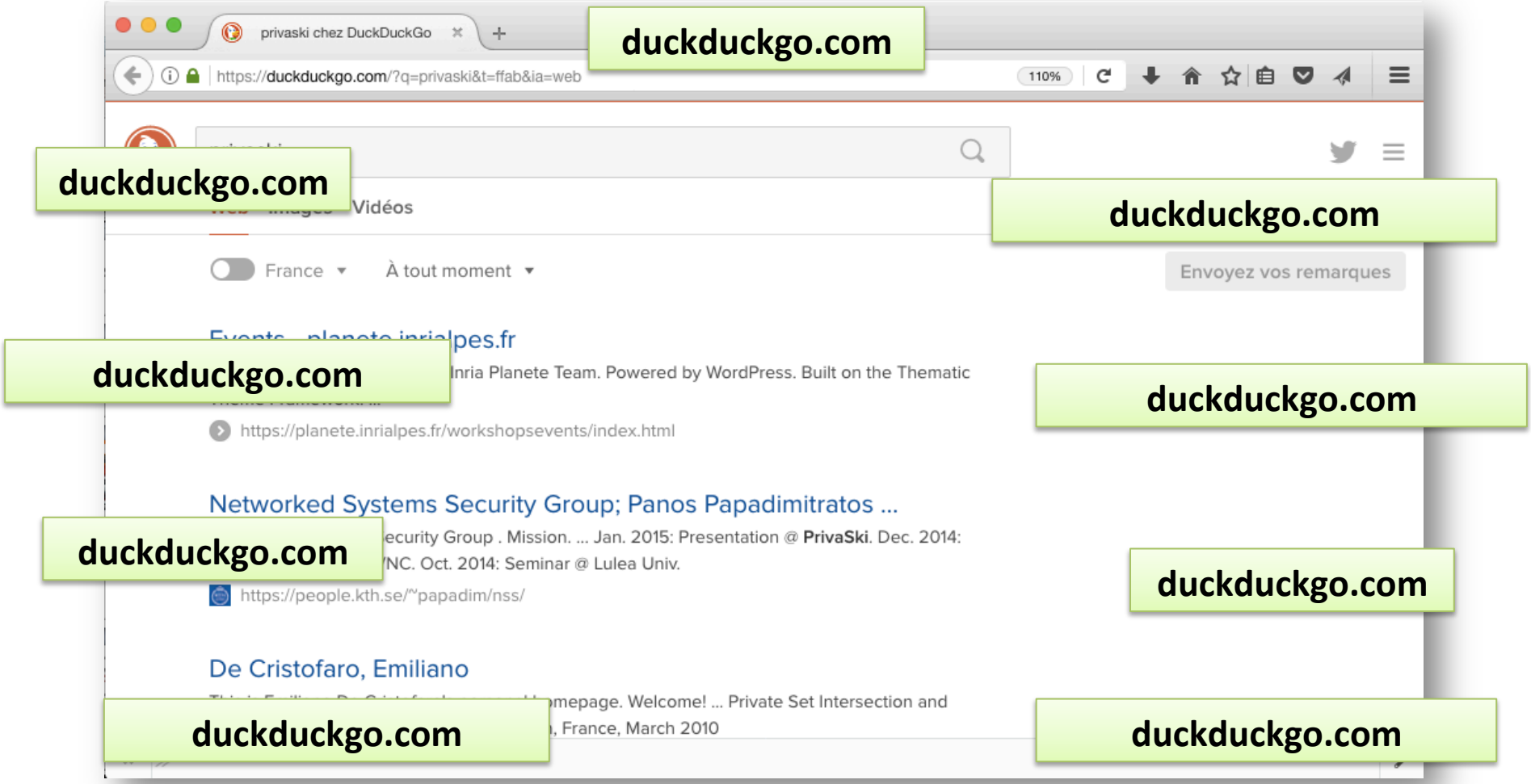
Why web developers include so many third party contents?

Functionality ☺ Privacy ☹



jQuery
write less, do more.

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
duckduckgo.com

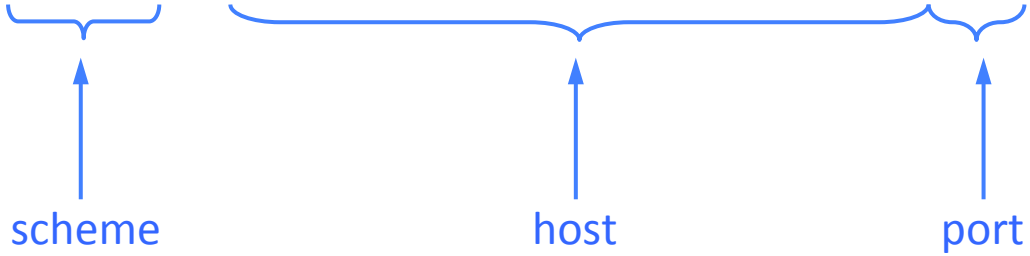
How can developers include third party content and guarantee privacy?

Same Origin Policy (SOP)

- SOP implemented in all web browsers:
  - **"Scripts can only access properties associated with documents from the same origin"**

- Origin = [scheme, host, port]

http://www.example.com:81/dir/page.html



In what origin each script is running?

a.com

a.com

```
<script src=b.com/script.js>
```

JavaScript 1

```
<iframe src=b.com/main.html>
```

Html page +

```
<script src=c.com/script.js>
```

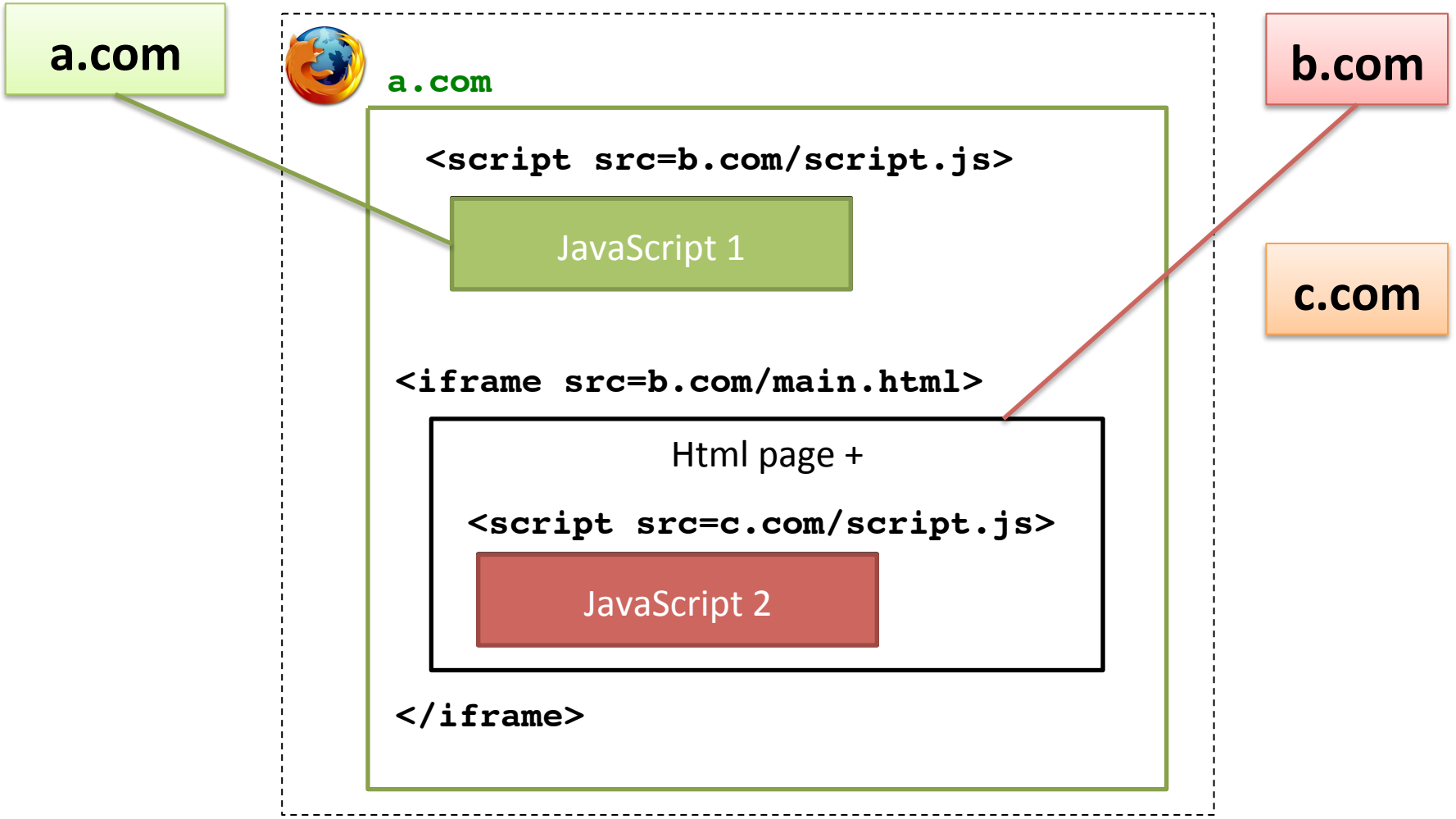
JavaScript 2

```
</iframe>
```

b.com

c.com

In what origin each script 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.

Privacy-preserving web architecture

Goal

- Remove tracking from functional third-party content

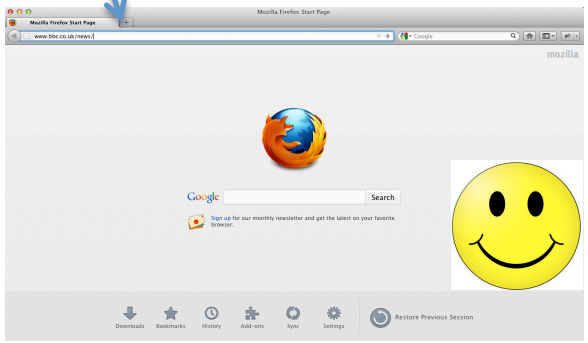
Idea

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

Current architecture

http://mysite.com

Web browser



Referer: mysite.com

cookie: id=789

tracker.com

mysite.com

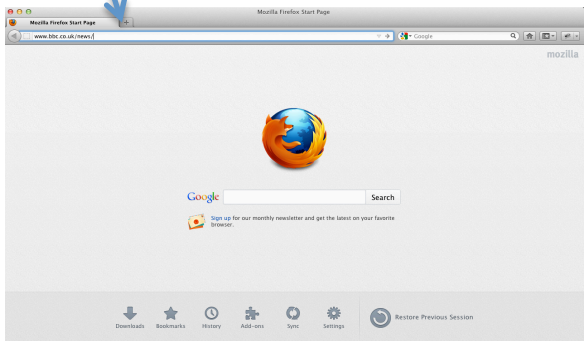
Our architecture

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

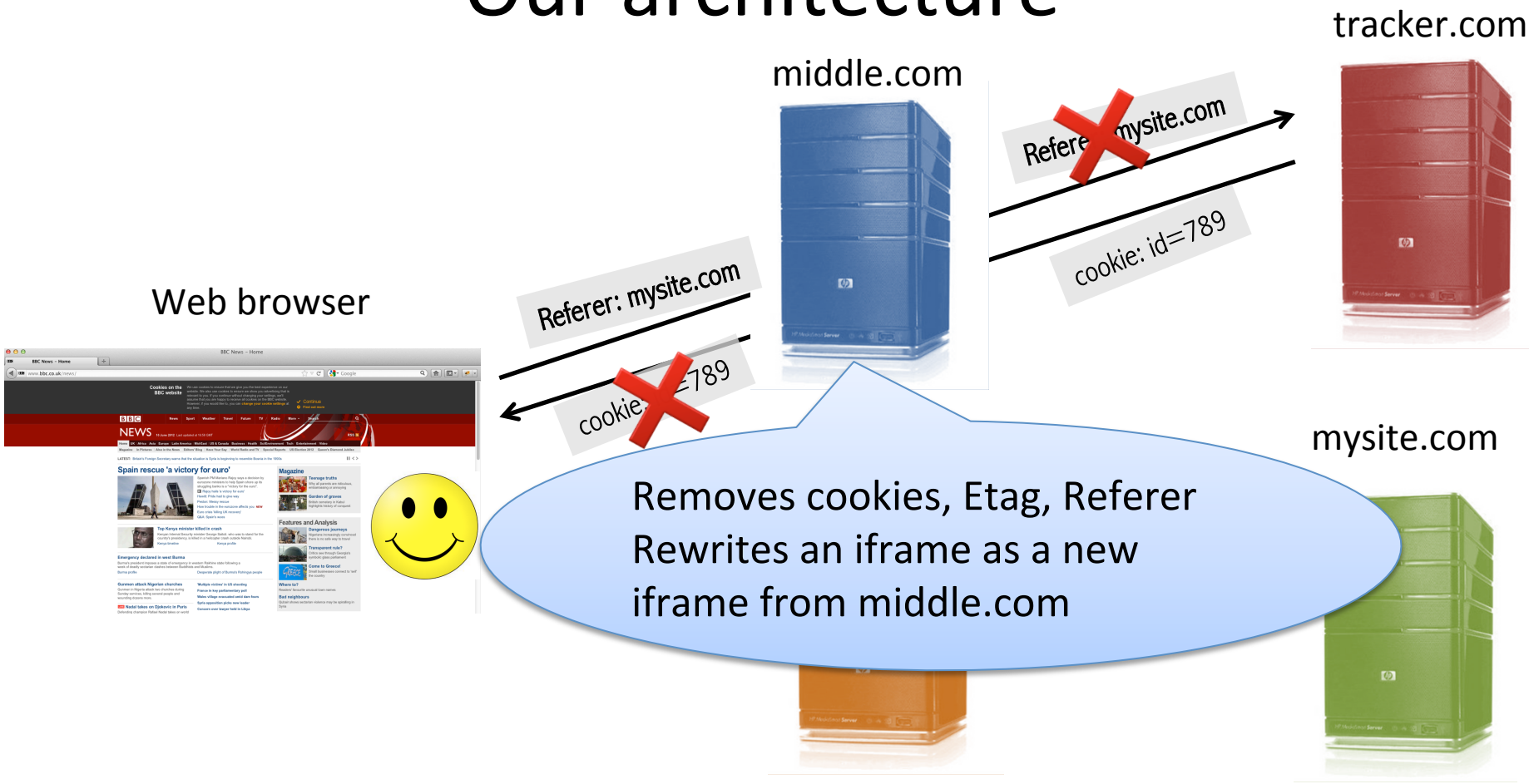
mysite.com

```
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



Case study & conclusions

- All websites work properly
  - Demo website with youtube videos
  - 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