Towards Privacy-Friendly Online Advertising

Julien Freudiger, Nevena Vratonjic, and Jean-Pierre Hubaux

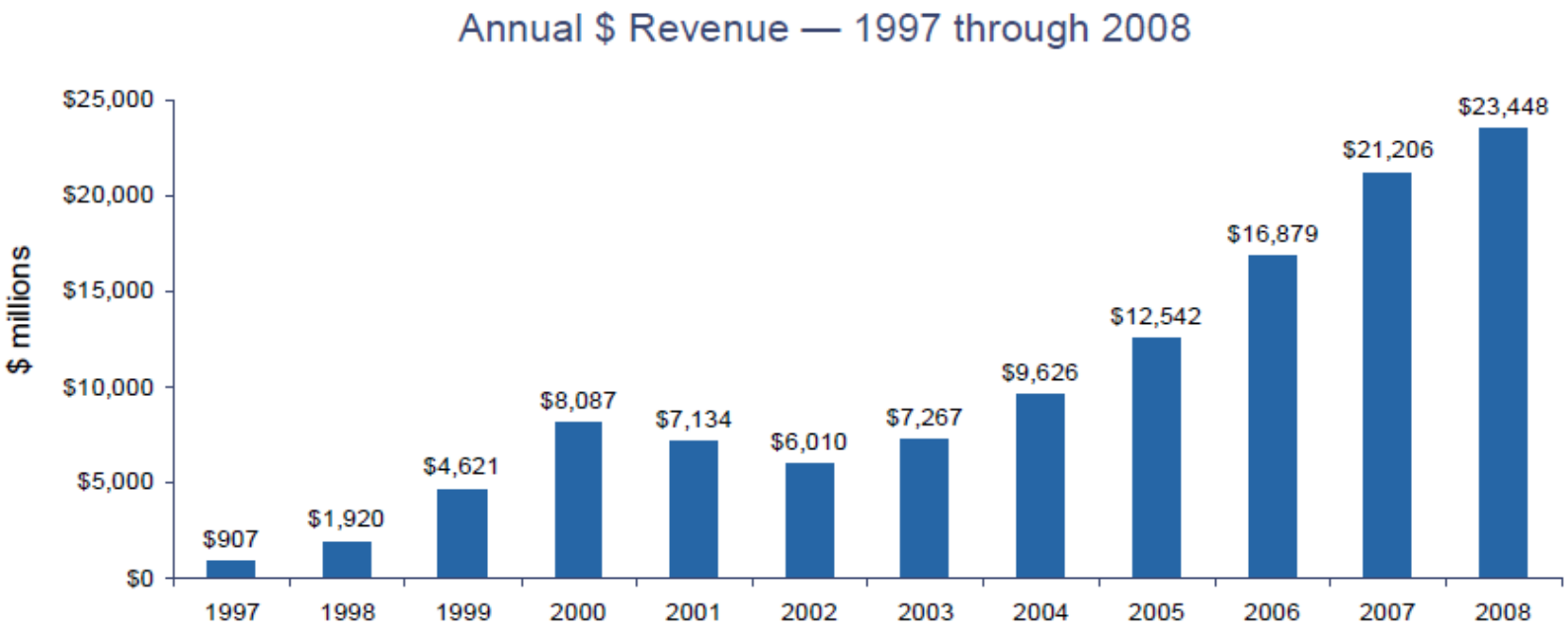
May 2009, W2SP

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Internet Economy

Online advertising is at center of Internet economy

- Immediate and personalized
- Enables **Behavioral** targeting



Source: Interactive Advertising Bureau Internet Advertising revenue report, 2008

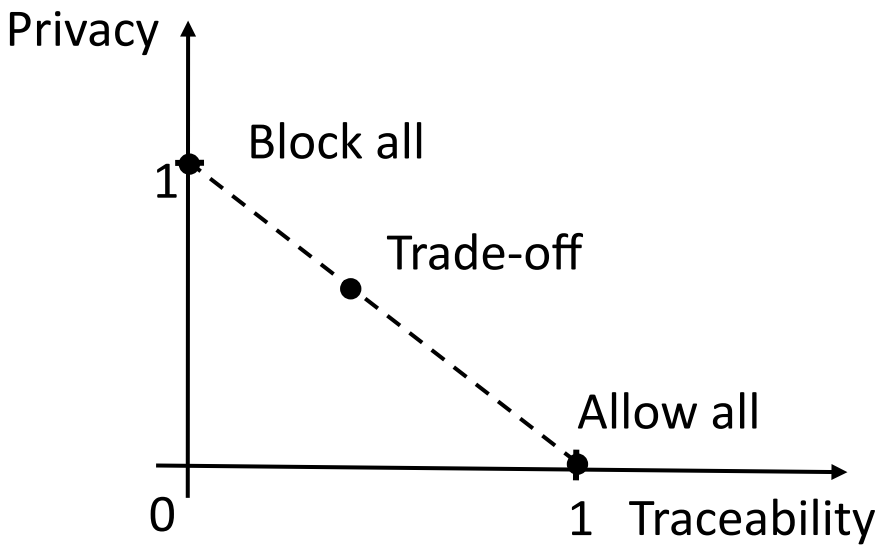
Benefits

- For users
  - Relevance of ads
  - Sponsored services
- For websites
  - Generate profit from ads
  - New business models

Privacy Concerns

- Track user activities online
  - Interests (visited websites, search terms)
  - Conversations (email)
  - Friends (social networks)
- Privacy footprint (Krishnamurthy and Wills)
  - 72% of web servers share at least one advertiser
  - 3 third-party domains contacted on average per accessed web site

Privacy/traceability Trade-off



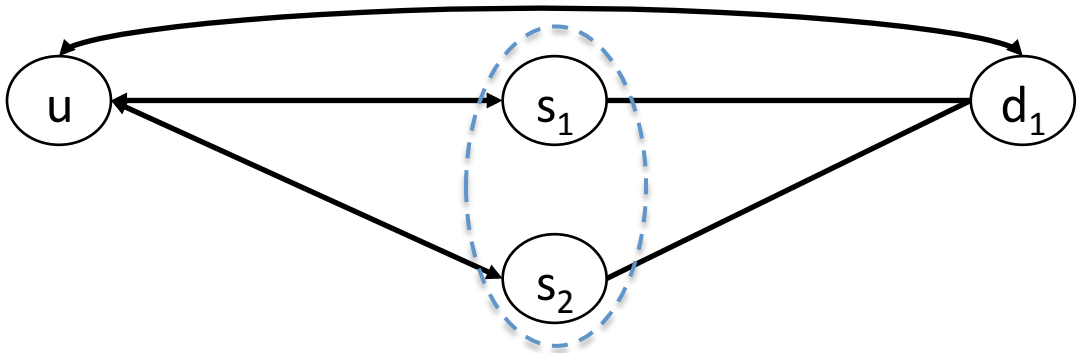
Provide a way to control amount of information shared

Outline

1. Online Advertising
   - Privacy Implications
   - Existing Solutions
2. Proposed Solution
   - Privacy friendly Cookie management
   - User centric
3. Preliminary Evaluation
   - Firefox Extension

Online Advertising

Users **U** | Visible servers **S** | Hidden servers **D**



u -> s_1:	www.nytimes.com	u -> s_2:	www.google.com
s_1-> u :	index.html	s_1-> u:	index.html
u -> d_1:	ads.com, **TP-cookie**	u -> d_1:	ads.com, **TP-cookie**
d_1-> u:	ads	d_1-> u:	ads

B. Krishnamurthy and C. E. Wills. Generating a privacy footprint on the Internet. IMC 2006

Traceability

- TP-Cookies enable
  - **Spatial** tracking: Track over different domains
  - **Temporal** tracking: Identify subsequent visits
- Referrer reveals visited website

- Example of data collected by advertisers:
  - 10h00: www.nytimes.com, cookie
  - 10h02: www.nytimes.com, cookie
  - 11h00: www.facebook.com/friends, cookie

Existing Solutions

- All or nothing
  - Block requests to advertisers
  - Block TP-cookies
  - Allow all

- **Same origin policy**
  - "Only the server that sets a cookie can access it"
  - Prevents loss of data confidentiality or integrity
  - But too permissive with respect to online tracking

Outline

1. Online Advertising
   - Privacy Implications
   - Existing Solutions
2. Proposed Solution
   - Privacy friendly Cookie management
   - User centric
3. Preliminary Evaluation
   - Firefox Extension

Proposed Solution

- Trade-off privacy and traceability
  - Limit spatial and temporal tracking
  - User-centric solution

- Define policies for use of cookies
  - User privacy preferences
  - User advertisement preferences
  - Visited web site

Key Idea

- Maintain a collection of cookies in parallel
  - Sent cookie depends on the visited web site and advertiser

Domain	Cookie
ads.com	**c1**

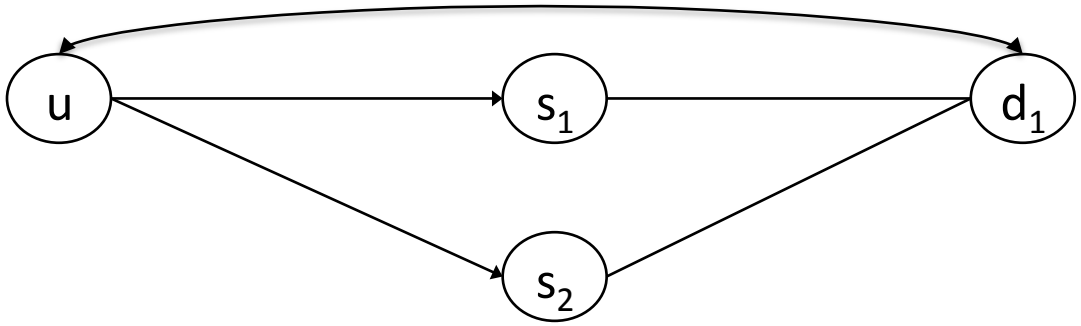
Domain	Website	Cookie
ads.com	nytimes.com	**c1**
ads.com	google.com	**c2**

Key Technique

- To obtain a new cookie
  - Do not send existing cookies in HTML header
  - Server assigns a new cookie
- Privacy-Friendly cookie management
  - Alternate among cookies in collection

Approach 1

Limit use of TP-cookies per domain
Use for a limited number of times



u-> $\mathbf{d_1}$: ads.com, www.nytimes.com, c1

u-> $\mathbf{d_1}$: ads.com, www.nytimes.com/technology, c1

u-> $\mathbf{d_1}$: ads.com, www.google.com , c2

because nytimes != google

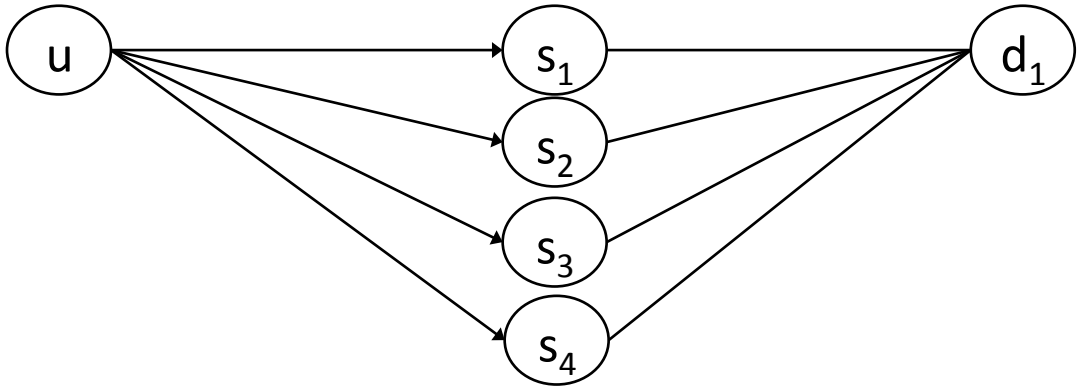
Approach 2

Limit use of TP-cookies per web site category and within categories
Use for a limited number of times

- Categories define type of web site
  - nytimes.com => news
  - Readily available (e.g., Alexa)

- Spatial tracking threshold $\mathbf{L_s}$
  - Limits spatial tracking across web sites within categories

Approach 2



$L_s = 2$

Category
News
News
News
Search
Email

u-> $\mathbf{d_1}$: ads.com, www.swissinfo.ch, c1

u-> $\mathbf{d_1}$: ads.com, www.nytimes.com, c1

u-> $\mathbf{d_1}$: ads.com, www.ft.com, c2 — Because 3 > L_S

u-> $\mathbf{d_1}$: ads.com, www.google.com , c3 — Because search != news

u-> $\mathbf{d_1}$: ads.com, mail.google.com , c4 — Because email != search and email != news

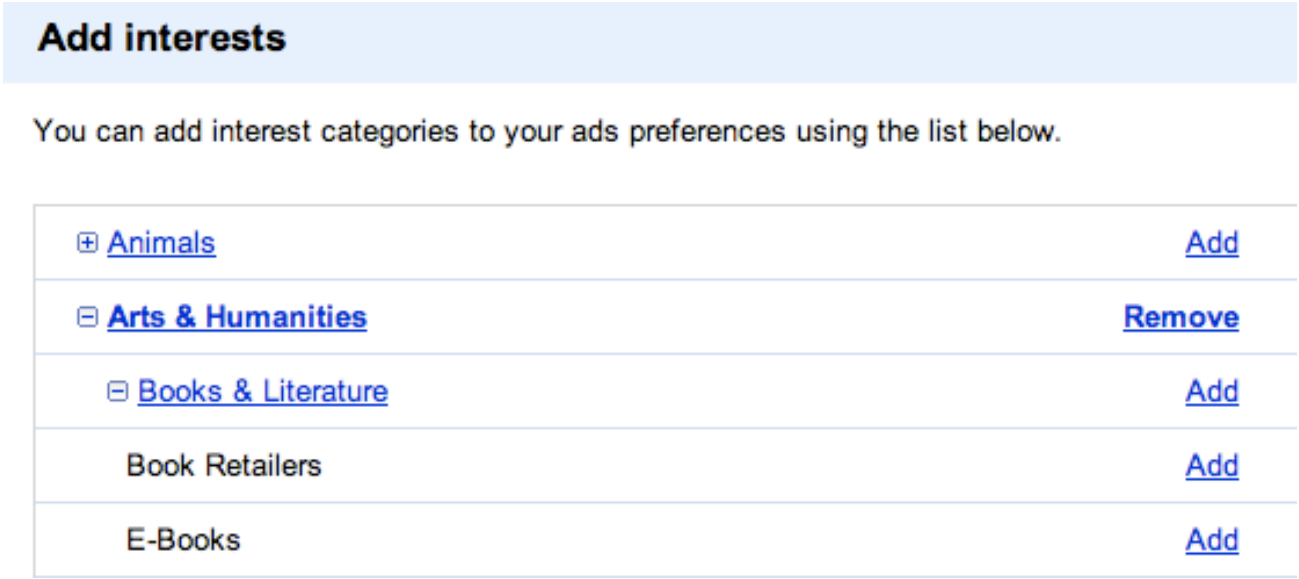
Approach 3

Limit use of TP-cookies based on URLs and user preferences
Use for a limited number of times

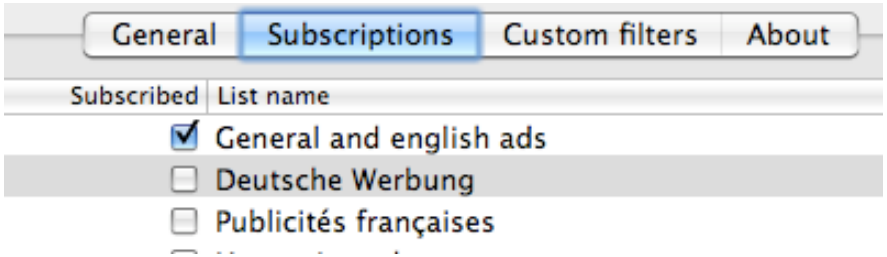
- URLs
  - Leak information through referrer
  - google.com/search?q=julien
- Preferences on web site categories
  - Privacy: What users do not want to share
  - Advertising: What users want to get

Setting up Preferences

Google Ad preference manager



Rely on online social communities



Approach 3

$$\sum_{b_i \in H(B)} w_1(b_i) \cdot w_2(b_i) < L_s$$

URLs (w_1)	User Privacy Pref. (w_2)
0.1	0
0.9	0
0.1	1
1	1

u-> $\mathbf{d_1}$: ads.com, www.google.com, c1

u-> $\mathbf{d_1}$: ads.com, www.google.com/search?q=computers, c1

u-> $\mathbf{d_1}$: ads.com, www.facebook.com, c1

u-> $\mathbf{d_1}$: ads.com, www.facebook.com/search?q=nevena , c2

$L_s = 1$

Because $0.1 + 1 > L_s$

Outline

1. Online Advertising
   - Privacy Implications
   - Existing Solutions
2. Proposed Solution
   - Privacy friendly Cookie management
   - User centric
3. Preliminary Evaluation
   - Firefox Extension

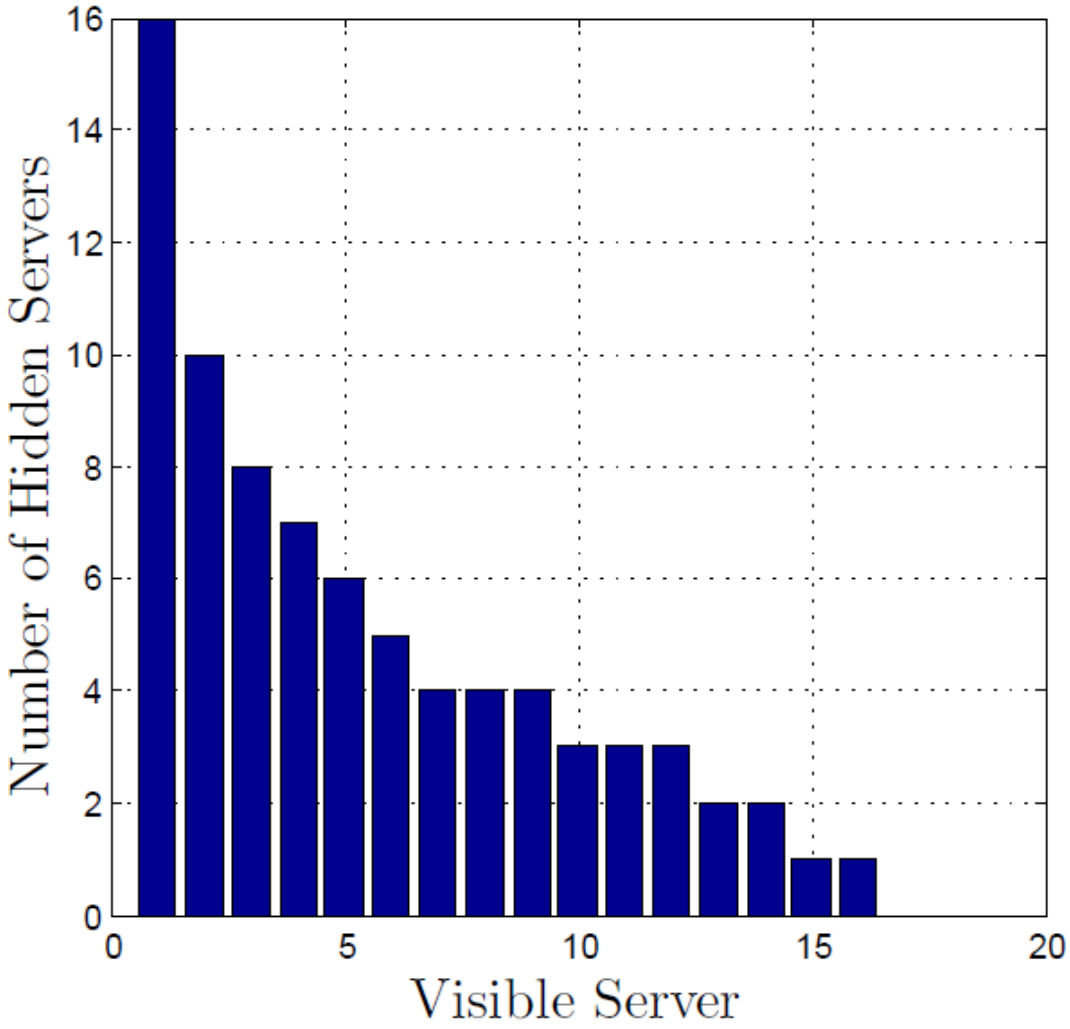
Implementation

- Firefox extension: **PrivaCookie**
  - Proof of concept code
  - Get it on http://icapeople.epfl.ch/freudiger
- TP cookie detection
  - Compare originating URL with current URL
- Local cookie table
  - Link cookies with hidden server that caused its assignment and visible server hosting ads
  - (Cookie, visible server, hidden server)

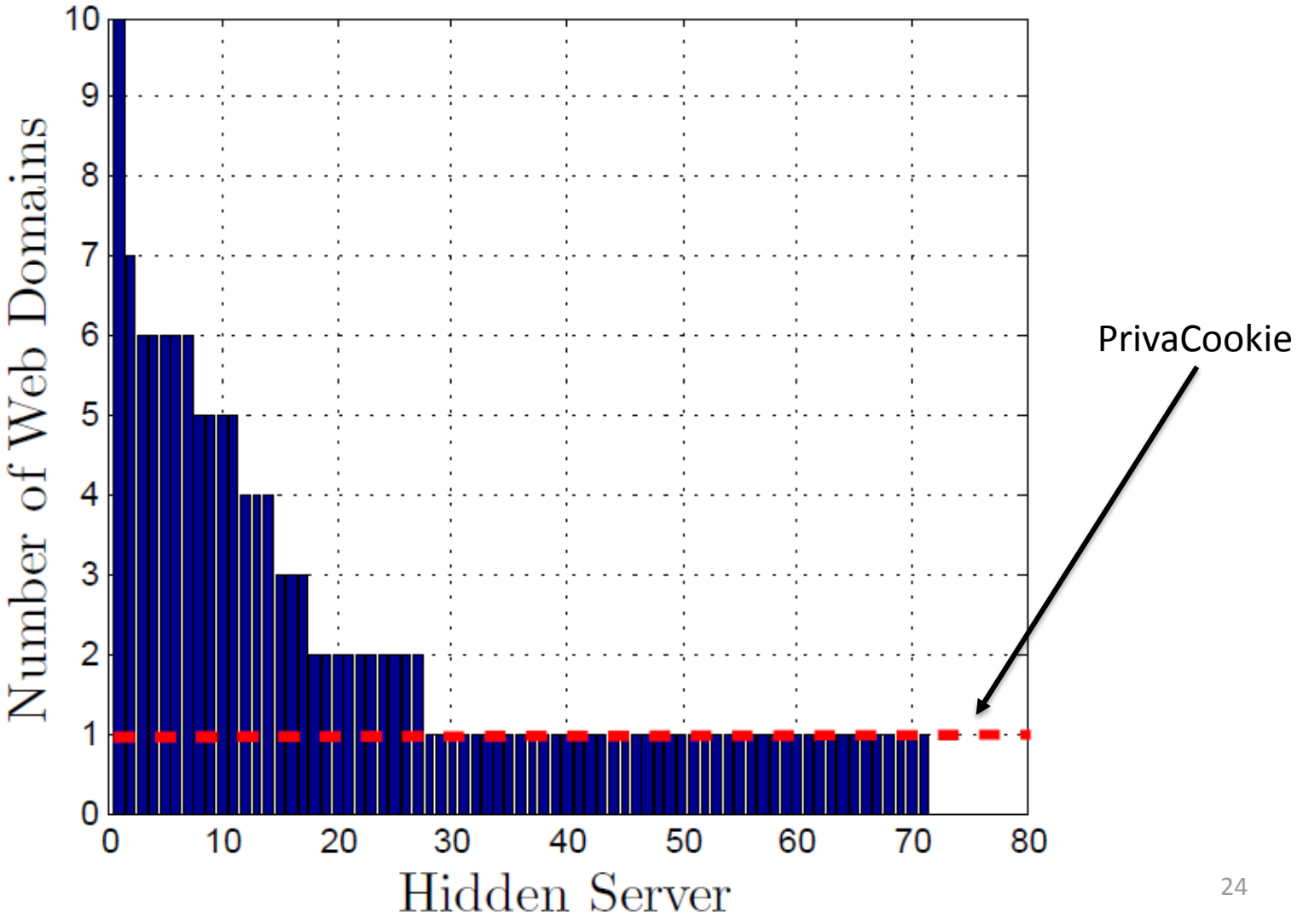
Study

- Chose 10 pages from each of the top 20 domains
- Firefox extension **pagestats**
  - Runs browser in batch mode with list of web sites
  - A total of 200 pages

Number of hidden servers for each of the top 20 domains



Number of visible servers for each hidden server



Top 10 associated visible servers connected with the most popular advertisers

Hidden Server	Visible Servers									
	Yahoo	Ebay	AOL	IMDB	Orkut	Msn	Myspace	HI5	Blogspot	Rapidshare
doubleclick	$c_1 \mid c_{1,1}$		$c_1 \mid c_{1,2}$	$c_1 \mid c_{1,3}$	$c_1 \mid c_{1,4}$	$c_1 \mid c_{1,5}$	$c_1 \mid c_{1,6}$	$c_1 \mid c_{1,7}$	$c_1 \mid c_{1,8}$	
quantaserve			✓			✓	✓		✓	
atmdt	✓	✓	✓	✓		✓				
advertising			✓	✓						
yieldmanager	✓		✓		✓		✓	✓		✓

Extension caused 81 additional cookies assignments

Tracking Countermeasures

- Track based on IP
  - Anonymizer/Tor
- Track with
  - Cache cookies
  - Browser history
  - Plugins (e.g., Flash cookies)
  - Proposed policies also apply to those cases
- Cooperative tracking?

Conclusion

- We propose a solution for trading-off privacy & traceability
  - Protects user privacy
  - Allows for targeted online advertising
  - No changes required from advertisers
  - Puts users in control
- Key idea: Maintains a collection of cookies in parallel
- **Future Work:**
  - Implement approach 2 & 3
  - Implement Javascript support
  - Consider other parameters in approach 3

URL Weight

- Parse URL for n-grams
  - "search"
  - "id"
  - "username"
- Can be done automatically before visiting URL