

ONLINE PRIVACY

Ben Livshits, Microsoft Research

Overview of Today's Lecture

2

- Some of the current problems in online privacy
- Tracking mechanisms
 - ▣ Cookies
 - ▣ Beacons
 - ▣ Browser fingerprinting
- Dangers of third-party tracking
- Ad ecosystem and user targeting
- Solutions for tracking prevention
- RePriv: combining personalization and privacy

Web privacy concerns

- Data is often collected silently
 - ▣ Web allows large quantities of data to be collected inexpensively and unobtrusively
- Data from multiple sources may be merged
 - ▣ Non-identifiable information can become identifiable when merged
- Data collected for business purposes may be used in civil and criminal proceedings
- Users are often given no explicit choice

HTTP Request + Cookie

4

```
GET /retail/searchresults.asp?qu=beer HTTP/1.0
Referer: http://www.us.buy.com/default.asp
User-Agent: Mozilla/4.75 [en] (X11; U; NetBSD
  1.5_ALPHA i386)
Host: www.us.buy.com
Accept: image/gif, image/jpeg, image/pjpeg, */*
Accept-Language: en
Cookie: buycountry=us; dcLocName=Basket;
  dcCatID=6773; dcLocID=6773; dcAd=buybasket;
  loc=; parentLocName=Basket; parentLoc=6773;
  ShopperManager%2F=ShopperManager%2F=66FUQULL0
  QBT8MMTVSC5MMNKBJFWDVH7; Store=107;
  Category=0
```

Referer Logging Issues

5

- GET methods result in values in URL
- These URLs are sent in the referer header to next host
- Somewhat contrived example:

```
http://www.ebay.com/cgi_bin/order?name=Bill+Clinton&address=here+there&credit+card=234876923234&PIN=1234& -> index.html
```

Tracking Mechanics: Cookies

6

- An HTTP cookie, originally invented by Lou Montulli and John Giannandrea at Netscape in 1994, is extremely useful for the web
- Cookies are the easiest way to offer "stateful" user interfaces such as user accounts and logins, multi-page forms, or online shopping carts
- Cookies also allow sites to store a unique ID in your browser, and to track you
- Many people have learned to block, limit or delete their cookies
- Categories of cookies
 - ▣ Persistent cookie – cookie replayed until expiration date
 - ▣ First-party cookie – cookie associated with the site the user requested
 - ▣ Third-party cookie – cookie associated with an image, ad, frame, or other content from a site with a different domain name that is embedded in the site the user requested

Tracking Mechanics: Beacons

7

- Often invisible 1x1 images
- Work just like banner ads from ad networks, but you can't see them unless you look at the code behind a web page
- Also embedded in HTML formatted email messages, MS Word documents, etc.

Yahoo!'s Practices Regarding Web Beacons

Yahoo! may collect information through web beacons about your web browsing activities such as the address of the page you are visiting, the address of the referrer page you previously visited, the time you are viewing the page, your browsing environment and your display settings. We may use the information we collect through web beacons:

- To understand traffic patterns and the number of visitors to the branded Yahoo! network of websites, websites within the Yahoo! Network Plus, and other non-Yahoo! websites that we partner with.
- To understand how you use and interact with Yahoo! products and services, including, but not limited to, the use of Yahoo! Mail outside of a browser-based experience.
- To improve Yahoo! products and services.
- To optimize your browsing experience.
- To provide anonymous individual and/or aggregate auditing, research, modeling and reporting for our advertisers and other partners. No personally identifiable information about you is shared with our advertisers and other partners as part of these services.

Tracking Mechanics: Fingerprinting

8



A research project of the [Electronic Frontier Foundation](#)

Panopti**cl**ick

How Unique – and Trackable – Is Your Browser?

Is your browser configuration rare or unique? If so, web sites may be able to track you, *even if you limit or disable cookies.*

Panopti**cl**ick tests your browser to see how unique it is based on the **information** it will share with sites it visits. Click below and you will be given a uniqueness score, letting you see how easily identifiable you might be as you surf the web.

Only **anonymous data** will be collected by this site.



A paper reporting the statistical results of this experiment is now available: [How Unique Is Your Browser?](#), Proceedings of the Privacy Enhancing Technologies Symposium (PETS 2010), Springer Lecture Notes in Computer Science.

Panopticlick Results

Panopticlick

How Unique – and Trackable – Is Your Browser?

Your browser fingerprint appears to be unique among the 1,865,596 tested so far.

Currently, we estimate that your browser has a fingerprint that conveys at least 20.83 bits of identifying information.

Browser Characteristic	bits of identifying information	one in x browsers have this value	value
User Agent	17.51	186559.6	Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/535.8 (KHTML, like Gecko) Chrome/17.0.938.0 Safari/535.8
HTTP_ACCEPT Headers	3.83	14.26	text/html, */* ISO-8859-1,utf-8;q=0.7,*;q=0.3 gzip,deflate,sdch en-US,en;q=0.8
Browser Plugin Details	20.83+	1865596	<p>Plugin 0: BitDefender QuickScan; BitDefender QuickScan Web Netscape Plugin; npqscan.dll; (npqscan; application/x-bitdefender-quickscanner;). Plugin 1: Chrome PDF Viewer; ; pdf.dll; (Portable Document Format; application/pdf; pdf) (Portable Document Format; application/x-google-chrome-print-preview-pdf; pdf). Plugin 2: Default Plug-in; Provides functionality for installing third-party plug-ins; default_plugin; (; *;). Plugin 3: Google Update; Google Update; npGoogleUpdate3.dll; (; application/x-vnd.google.update3webcontrol.3;) (; application/x-vnd.google.oneclickctrl.9;). Plugin 4: Microsoft Office 2010; Office Authorization plug-in for NPAPI browsers; NPAUTHZ.DLL; (14.0.4730.1010; application/x-msoffice14; *;). Plugin 5: Microsoft Office 2010; The plug-in allows you to open and edit files using Microsoft Office applications; NPSPWRAP.DLL; (SharePoint Plug-in for Firefox; application/x-sharepoint;). Plugin 6: Native Client; ; ppGoogleNaClPluginChrome.dll; (Native Client Executable; application/x-nacl; nexa). Plugin 7: Picasa; Picasa plugin; npPicasa3.dll; (3.1; application/x-picasa-detect; pinstall). Plugin 8: QuickTime Plug-in 7.7; The QuickTime Plug-in allows you to view a wide variety of multimedia content in Web pages. For more information, visit the QuickTime Web site.; npqtplugin.dll; (SDP stream descriptor; application/sdp; sdp) (SDP stream descriptor; application/x-sdp; sdp) (RTSP stream descriptor; application/x-rtsp; rtsp,rt) (QuickTime Movie; video/quicktime; mov,qt,mqv) (AutoDesk Animator (FLC); video/flc; flc,flc,cel) (WAVE audio; audio/x-wav; wav,bwf) (WAVE audio; audio/wav; wav,bwf). Plugin 9: QuickTime Plug-in 7.7; The QuickTime Plug-in allows you to view a wide variety of multimedia content in Web pages. For more information, visit the QuickTime Web site.; npqtplugin2.dll; (AIFF audio; audio/aiff; aiff,aif,aifc,odda) (AIFF audio; audio/x-aiff; aiff,aif,aifc,odda) (uLaw/AU audio; audio/basic; au,snd,ulw) (MIDI; audio/mid; mid,midi,smf,kar) (MIDI; audio/x-midi; mid,midi,smf,kar) (MIDI; audio/midi; mid,midi,smf,kar) (QUALCOMM PureVoice audio; audio/vnd.qcelp; qcp). Plugin 10: QuickTime Plug-in 7.7; The QuickTime Plug-in allows you to view a wide variety of multimedia content in Web pages. For more information, visit the QuickTime Web site.; npqtplugin3.dll; (GSM audio; audio/x-gsm; gsm) (AMR audio; audio/amr; AMR) (AAC audio; audio/aac; aac,adts) (AAC audio; audio/x-aac; aac,adts) (CAF audio; audio/x-caf; caf) (AC3 audio; audio/ac3; ac3) (AC3 audio; audio/x-ac3; ac3) (MPEG media; video/x-mpeg; mpeg,mpeg,m1s,m1v,m1a,m75,m15,m2,mpeg,mpv,mpa). Plugin 11: QuickTime Plug-in 7.7; The QuickTime Plug-in allows you to view a wide variety of multimedia content in Web pages. For more information, visit the QuickTime Web site.; npqtplugin4.dll; (MPEG media; video/mpeg; mpeg,mpeg,m1s,m1v,m1a,m75,m15,m2,mpeg,mpv,mpa) (MPEG audio; audio/mpeg; mpeg,mpeg,m1s,m1a,m2,mpeg,mpa,m2a) (MPEG audio; audio/x-mpeg; mpeg,mpeg,m1s,m1a,m2,mpeg,mpa,m2a) (3GPP media; video/3gpp; 3gp,3gpp). Plugin 12: QuickTime Plug-in 7.7; The QuickTime Plug-in allows you to view a</p>

Third-Party Tracking

10

- A third party is typically an advertiser or ad network
- Their content is placed alongside primary (first-party) content
- Requests go to their site and result in
 - ▣ Referred often containing the URL and user identifying information to be sent to the site
 - ▣ An ID that is stored in the cookie for cross-correlation
 - ▣ Date, time, etc.

Clickstreams

11

- In the language of computer science, clickstreams – browsing histories that companies collect – are not anonymous at all; rather, they are pseudonymous.
- The latter term is not only more technically appropriate, it is much more reflective of the fact that at any point after the data has been collected, the tracking company might try to attach an identity to the pseudonym (unique ID) that your data is labeled with.
- Thus, identification of a user affects not only future tracking, but also retroactively affects the data that's already been collected. Identification needs to happen only once, ever, per user.

Arvind Narayanan, Stanford

Magnitude of the Problem

12

Privacy leakage vs. Protection measures: the growing disconnect

Balachander Krishnamurthy
AT&T Labs-Research
bala@research.att.com

Konstantin Naryshkin
Worcester Polytechnic Institute
konary@wpi.edu

Craig E. Wills
Worcester Polytechnic Institute
cew@cs.wpi.edu

ABSTRACT

Numerous research papers have listed different vectors of personally identifiable information leaking via traditional and mobile Online Social Networks (OSNs) and highlighted the ongoing aggregation of data about users visiting popular Web sites. We argue that the landscape is worsening and existing proposals (including the recent U.S. Federal Trade Commission's report) do not address several key issues. We examined over 100 popular non-OSN Web sites across a number of categories where tens of millions of users representing diverse demographics have accounts, to see if these sites leak private information to prominent aggregators. Our results raise considerable concerns: we see leakage in sites for every category we examined; fully 56% of the sites directly leak pieces of private information with this result growing to 75% if we also include leakage of a site user's ID. Sensitive search strings sent to healthcare Web sites and travel itineraries on flight reservation sites are leaked in 9 of the top 10 sites studied for each category. The community needs a clear understanding of the shortcomings of existing privacy protection measures and the new proposals. The growing disconnect between the protection measures and increasing leakage and linkage suggests that we need to move beyond the losing battle with aggregators and examine what roles first-party sites can play in protecting privacy of their users.

1. INTRODUCTION

Recently, multiple vectors of private information leakage via Online Social Networks (OSN) and the two-decade long aggregation of data about users visiting popular Web sites have been reported. The problem of privacy has worsened significantly in spite of the various proposals and reports by researchers, government agencies, and privacy advocates. The ability of advertisers and third-party aggregators to collect a vast amount of increasingly personal information about users who visit various Web sites has been steadily growing. Numerous stories have expressed alarm about the situation with legislators and privacy commissioners in different countries paying closer attention to the problem [14]. The awareness about the steady erosion of privacy on the part of users is growing slowly. The potential economic impact as a result of loss of brand value has forced some companies to start paying closer attention to complaints from users and privacy advocates.

In this paper we argue that the privacy landscape is worsening as there is a growing disconnect between steadily increasing leakage to and linkage by aggregators with existing and proposed protection measures. We show that beyond the egregious leakage of private information via OSNs and their more recent mobile counterparts, a key part of the Internet with tens of millions of users representing diverse demographics with accounts on popular non-OSN Web sites also suffer from private information leakage to prominent aggregators. Additionally, less well-understood notions of linkage are typically not addressed by most of the proposed privacy solutions. One such privacy issue arises from the existence of globally unique IDs such as an OSN ID or reused email addresses that could be used to link together pieces of seemingly distinct information. Beyond the intrinsic identifying nature of these IDs, they aid in linking together other information, such as cookies from a home and work computer. New proposals, such as the recent United States Federal Trade Commission's December 2010 report [10], fail to address several key issues.

Our earlier work focused on longitudinal data gathering by aggregators on the Web [15], leakage of personal information via popular OSNs [13] and the more recently mobile OSNs [16]. However, there has been no attention paid thus far to another segment of the Internet where sites encourage and allow users to create accounts so that they could have a richer interaction experience. Many popular Web sites allowed users to establish profiles long even before the advent of OSNs. There are significant demographics that are present in non-OSN Web sites that may not be on OSN sites and their private information is also of interest to aggregators. On many of these sites, users create profiles with varying amounts of personal information, but typically less than what they supply on OSN sites. Unlike OSNs, these Web sites already have content and do not depend on users to create content; users could however add comments or tags. Surprisingly, there is considerable overlap in the nature of personal information that users provide across these sites. We should also note that the degree of sensitivity to different aspects of their personal information varies across users as is the potential for identifiability (ability to link a unit of personal information with a specific user).

- Recorded interactions with 120 popular sites for information leakage to third parties
- Found that
 - ▣ 56% leaked some form of private information
 - ▣ 48% leaked a user identifier

Linking User Names Across Services

13

How Unique and Traceable are Usernames?

Daniele Perito, Claude Castelluccia, Mohamed Ali Kaafar, Pere Manils
INRIA Rhône-Alpes
{perito,ccastel,kaafar,manils}@inrialpes.fr

ABSTRACT

Suppose you find the same username on different online services, what is the probability that these usernames refer to the same physical person? This work addresses what appears to be a fairly simple question, which has many implications for anonymity and privacy on the Internet. One possible way of estimating this probability would be to look at the public information associated to the two accounts and try to match them. However, for most services, these information are chosen by the users themselves and are often very heterogeneous, possibly false and difficult to collect. Furthermore, several websites do not disclose any additional public information about users apart from their usernames (e.g., discussion forums or Blog comments), nonetheless, they might contain sensitive information about users.

This paper explores the possibility of linking users profiles only by looking at their usernames. The intuition is that the probability that two usernames refer to the same physical person strongly depends on the “entropy” of the username string itself. Our experiments, based on crawls of real web services, show that a significant portion of the users’ profiles can be linked using their usernames. To the best of our knowledge, this is the first time that usernames are considered as a source of information when profiling users on the Internet.

1. INTRODUCTION

Online profiling is a serious threat to users privacy. In particular, the ability to trace users by linking multiple identities from different public profiles may be of great interest to profilers, advertisers and the like. Indeed, it might be possible to gather information from different online services and combine it to sharpen the knowledge of users identities. This knowledge may then be exploited to perform efficient social phishing or targeted spam, and

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
Copyright 2010 ACM X-XXXXX-XX-X/XXX/XX ...\$10.00.

might be as well used by advertisers or future employers seeking information. As it has been colloquially put by a judge of the US Supreme Court in a recent case about warrantless GPS tracking¹: “When it comes to privacy, the whole may be more revealing than its parts.”

Recent works [4, 3] showed how it is possible to retrieve users information from different online social networks (OSN). All of these works mainly exploit flaws in the OSN’s API design (e.g., Facebook friend search). Other approaches [17] use the topology of social network friend graphs to de-anonymize its nodes.

In this paper, we propose a novel methodology that uses usernames -an easy to collect information- rather than social graphs to tie user online identities. Our technique only assumes knowledge of usernames and it is widely applicable to all web services that publicly expose usernames. Our purpose is to show that users’ pseudonyms allow simple, yet efficient tracking of online activities.

Recent scraping services’ activities illustrate well the threats introduced by the ability to match up user’s pseudonyms on different social networks [2]. For instance, PeekYou.com has lately applied for a patent for a way to match people’s real names to pseudonyms they use on blogs, OSN services and online forums [14]. The methodology relies on public information collected for an user, that might help in matching different online identities. The algorithm empirically assigns weights to each of the collected information so as to deem different identities to be the same. However, the algorithm is ad-hoc and not robust to false or mismatching information. In light of these recent developments, it is desirable that the research community investigates the capabilities and limits of these profiling techniques. This will, in turn, allow for the design of appropriate countermeasures to protect users’ privacy.

In general, profiling unique identities from multiple public profiles is a challenging task, as information from public profiles is often incorrect, misleading or altogether missing [11]. Techniques designed for the purpose of profiling need to be robust to these occurrences.

Contributions.

The contributions of this paper are manifold. First, we introduce the problem of linking multiple online identities

¹<http://www.eff.org/press/archives/2010/08/06-0>

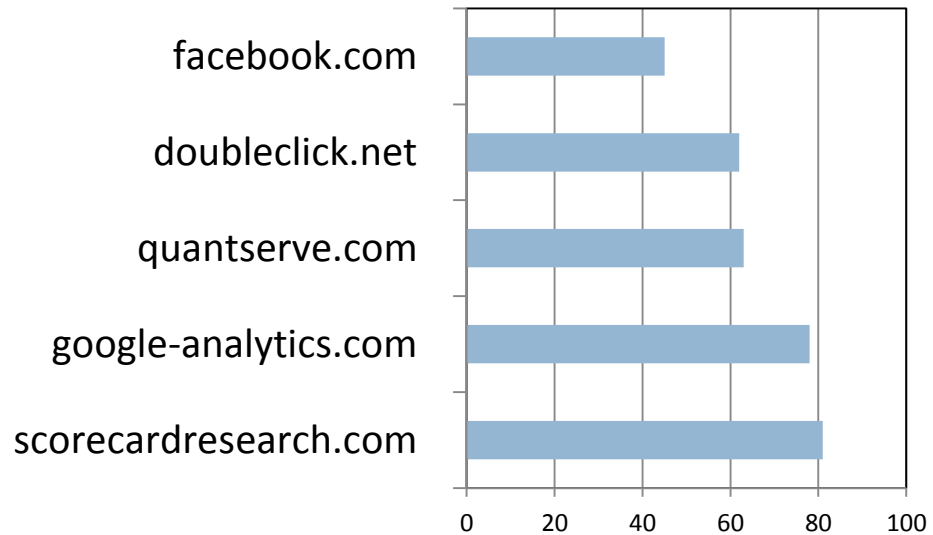
- Suppose you find the same username on different online services, what is the probability that these usernames refer to the same physical person?
- Our experiments, based on crawls of real web services, show that a significant portion of the users’ profiles can be linked using their usernames.
- To the best of our knowledge, this is the first time that usernames are considered as a source of information when profiling users on the Internet.

□

Recent Stanford Experiments

14

- Picked 185 popular sites
 - Used FourthParty web measurement platform to create an account and interact with the site
 - Explored content that dealt with user identity, such as profile and settings pages
 - After collecting data, searched Request-URIs and Referer headers for known personal information
- User name/ID leaked in 113 websites or 61%



More Results from the Stanford Study

15

- Viewing a local ad on the Home Depot website sent the user's first name and email address to 13 companies
- Entering the wrong password on the Wall Street Journal website sent the user's email address to 7 companies
- Changing user settings on the video sharing site Metacafe sent first name, last name, birthday, email address, physical address, and phone numbers to 2 companies
- Signing up on the NBC website sent the user's email address to 7 companies
- Signing up on Weather Underground sent the user's email address to 22 companies.
- The mandatory mailing list page during CNBC signup sent the user's email address to 2 companies.
- Clicking the validation link in the Reuters signup email sent the user's email address to 5 companies.
- Interacting with Bleacher Report sent the user's first and last names to 15 companies.
- Interacting with classmates.com sent the user's first and last names to 22 companies.

Privacy Policies?

16

- Many first-party websites make what would appear to be incorrect, or at minimum misleading, representations about not sharing PII. Here are some examples:
 - The Home Depot:
 - Personal Information Disclosure: The Home Depot will not trade, rent or sell your personal information, without your prior consent, except as otherwise set out herein. [Does not describe sharing with third-parties for advertising or analytics.]
 - The Wall Street Journal:
 - We will not sell, rent, or share your Personal Information with these third parties for such parties' own marketing purposes, unless you choose in advance to have your Personal Information shared for this purpose. Information about your activities on our Online Services and other non-personally identifiable information about you may be used to limit the online ads you encounter to those we believe are consistent with your interests. Third-party advertising networks and advertisers may also use cookies and similar technologies to collect and track non-personally identifiable information such as demographic information, aggregated information, and Internet activity to assist them in delivering advertising on our Online Services that is more relevant to your interests.
-

Players in the Online Space: Ad Scenario

17

- Ad networks
- Hosts – sites on which ads are placed
- Users – some are concerned about their privacy

Ad Targeting

18

- The better (more relevant) ads are, the more they appeal to the user
- The more they appeal to the user, the higher the click-through rates (CTR) become
- The more click the advertising network gets, the more they get paid (pay-per-click)
- How do we create more relevant ads?
- Need to know what the user finds relevant
- How can we find that out?
- One option is to do user profiling/modeling
- Followed by ad targeting

Tracking Prevention Solutions

19

1. Browser privacy modes
2. Opting out of cookie-based tracking
3. "Do Not Track (DNT)
4. Tracking Protection Lists (TPLs)

Browser Privacy Modes

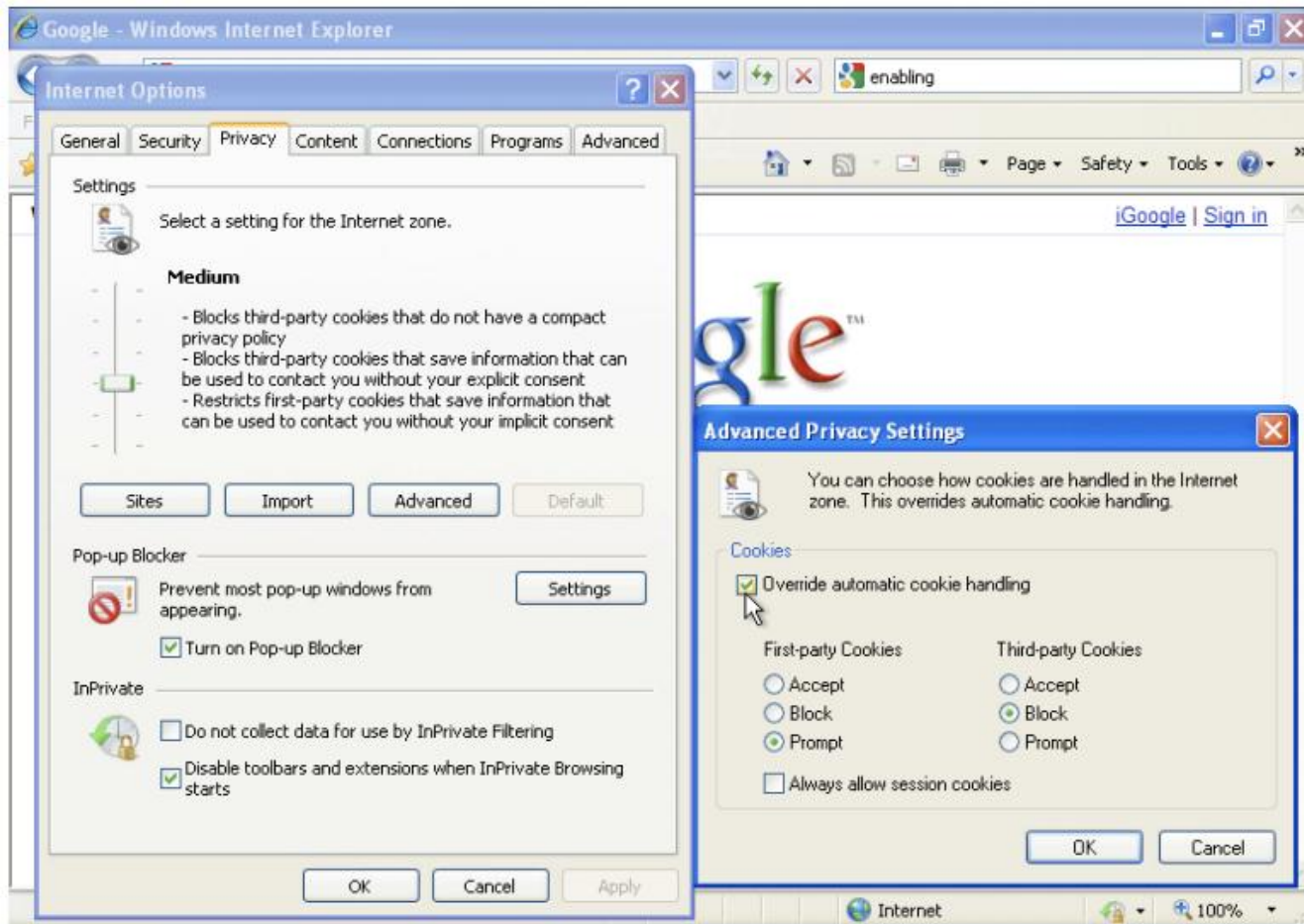
- Prevent access to persistent user data
- Prevent storing persistent data
- Cleanse referers

Privacy Mode Comparison	Chrome's Incognito	IE8's InPrivate Browsing	Firefox 3.5's Private Browsing	Safari's Private Browsing ²
Visited sites are not stored in the browser history	✓	✓	✓	✓
Downloaded files are not stored in the download history	✓		✓	✓
Form field data (including passwords) is not stored	✓	✓	✓	✓
Addresses typed into the address bar are not stored	✓	✓	✓	✓
Visited links are not stored	✓	✓	✓	✓
Search queries are not stored in the browser	✓	✓	✓	✓
Cached files are deleted at the end of the browsing session	✓	✓	✓	✓
Existing third-party cookies cannot be read	✓	✓	✓	✓

Privacy Mode Comparison	Chrome's Incognito	IE8's InPrivate Browsing	Firefox 3.5's Private Browsing	Safari's Private Browsing ²
New cookies are deleted at the end of the session	✓	✓	✓	✓
Blocks referring URL from being sent. ³		✓		
Mode can operate on a per-window basis.	✓	✓		
Mode can persist even when user quits and re-starts browser.				

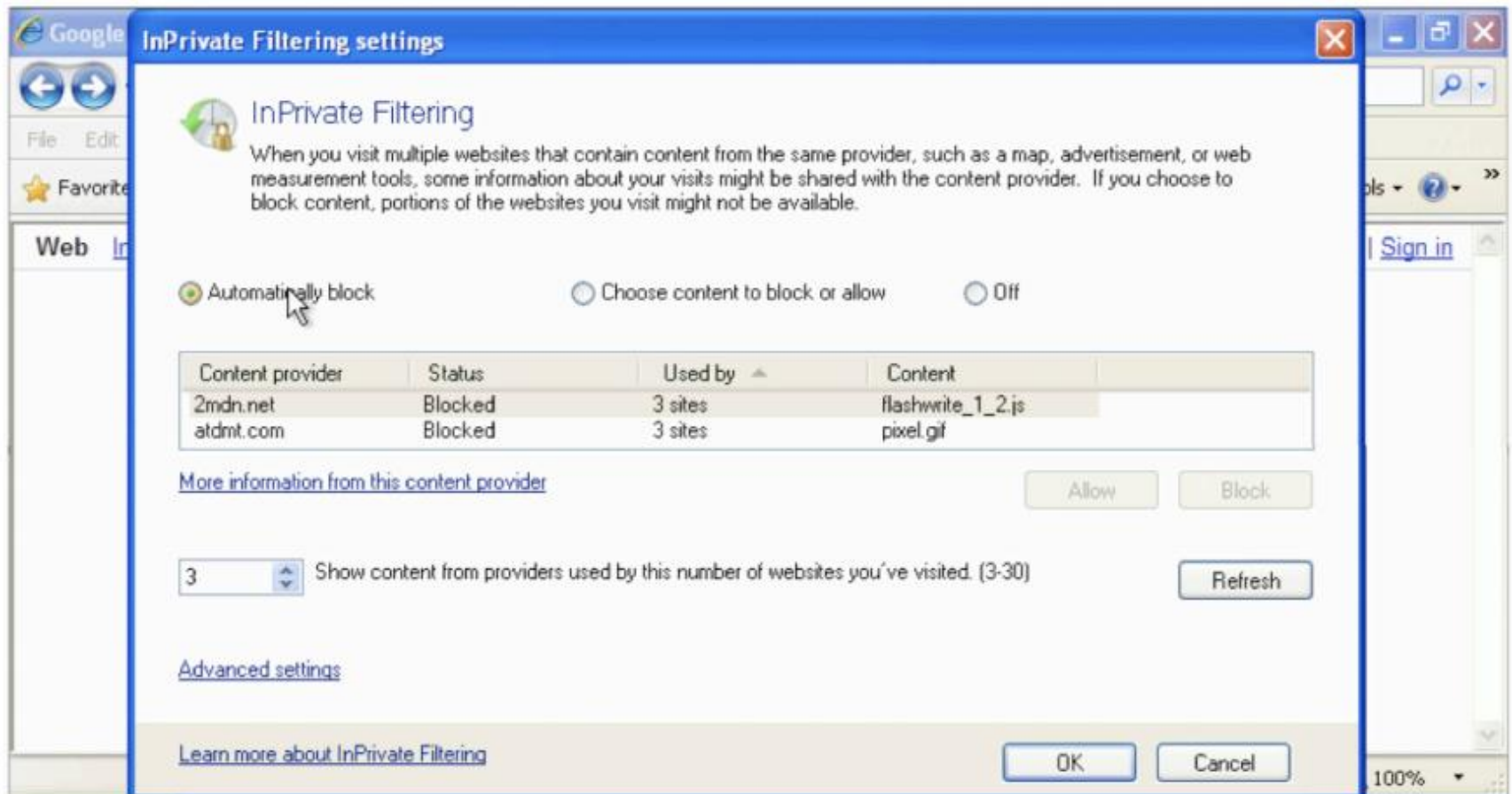
Controlling Cookie Access

21



InPrivate Filtering in IE8/IE9

22



Opting out of Cookie-based Tracking

23

- Instead of preventing cookie access, explicitly set opt-out cookies
- Many ad networks provide mechanisms for this
- There are tools to help you set the right cookie: SelectOut.org

247-real-media	33across	accuen-media	acxiom
adap-tv	adara-media	adbrite	adbuyer
adcentric	adchemy	adconion	adconion
adgear	adinterax	adition-technologies	adjug
adjugler	admeld	admotion	adnetik
adnologies	adotube	adperium	adroit-interactive
adshuffle	adspeed	adtech	advertising-aol
aggregateknowledge	akamai	almondnet	anonymous-media
aol-behavioral	aol-sponsored	appnexus	atlas-technology
audiencescience	beencounter	bidplace	bizo
bluekai	brandnet	brightroll	brilig
btbuckets	buysight	buzzlogic	bv-media
casale-media	chango	channel-intelligence	choice-stream
clearspring	clickdistrict	cobalt-group	cognitive-match
contextin	convertro	cox-digital-solutions	cpmstar
cpx-interactive	crimson-tangerine	critico	cross-pixel-media
dapper	datalogix	dataxu	datran-media
demand-media	demdex	dotomi	double-verify
e-planning	think-realtime	effective-measure	efficient-frontier
eloqua	emc	engage-bdr	exelate-media
experian-marketing-services	exponential-interactive	eyewonder	facilitate-digital
fetchback	fireclick	flashtalking	forbes-media
fox-audience	freewheel	full-circle-studies	glam-media
google	groupon	hurra-communications	i-behavior
infectious-media	inflection-point-media	insight-express	intent-media
interclick	invite-media	jumptap	keyade
lijit	liverrail	lotame	lucid-media
magnetic	maxpoint-interactive	media-innovation-group	media6degrees
mediaforge	mediamath	mediamind	mediaplex
meebo	microsoft	millennial-media	mindset-media
mindshare	mixpo	monster	mybuys
mythings-media	navegg	netmining	newtention
next-performance	nextag	nielsen	nugg-ad
omniture	openx	optimax-media-delivery	outbrain
owneriq	oxamedia	peerset	pointroll
precisionclick	predictad	proximic	publishers-clearing-house
pubmatic	pulse360	quantcast	quinstreet
quisma	radiumone	rapleaf	red-aril
reedge	rewardtv	richrelevance	ringleader-digital
rocket-fuel	rovion	safecount	sagemetrics

Manipulating Opt-Out Cookies

24

The screenshot displays a web browser's cookie management interface for the domain **adchemy.com**. A cookie named **adc_optout** is selected. A dialog box is open, showing the following details:

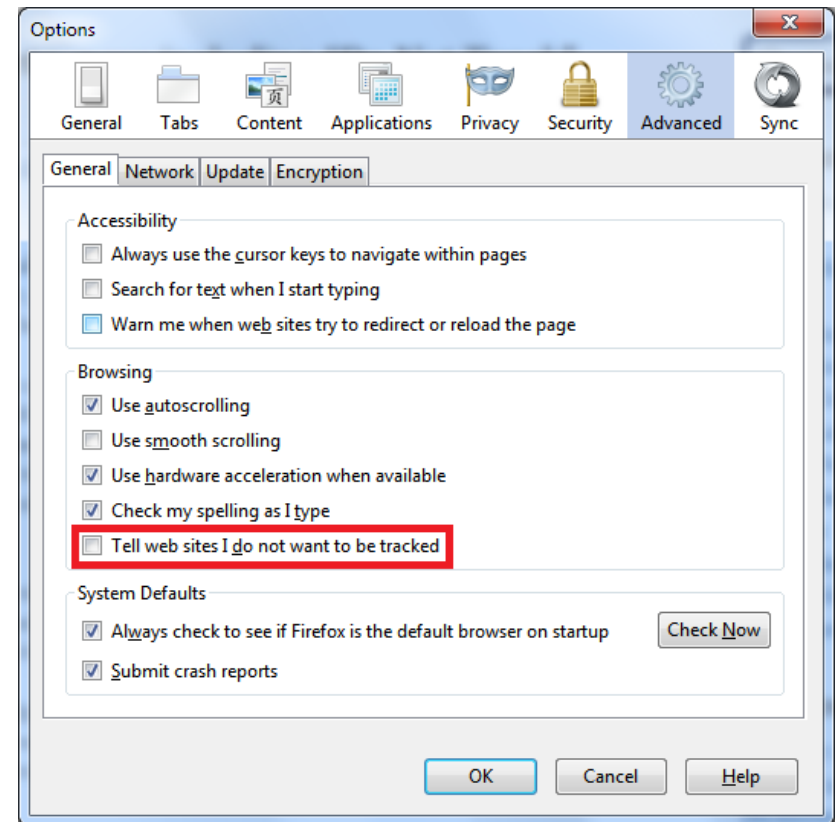
- Domain:** .adchemy.com
- Path:** /
- Name:** adc_optout
- Store ID:** 0
- Value:** opted_out
- Expires:** 2030-12-31 19:00
- Session
- Host-Only
- Read-Only
- Secure
- Set Cookie** button

At the bottom of the dialog, there are buttons for **Exceptions**, **Delete All**, and **Delete**. A red banner at the top right of the dialog reads **✓ Opt-Out**. The background shows a sidebar with 'Quick' and 'Stand' sections and a main content area with partially visible text: 'bles', 'ental', 'nt, we', and 'across'.

"Do Not Track (DNT)

25

- The Do Not Track proposal is to include a simple, machine-readable header indicating that you don't want to be tracked. The header that would be inserted is **DNT : 1**
- Because this signal is a header, and not a cookie, users will be able to clear their cookies at will without disrupting the functionality of the Do Not Track flag
- It's important to note that there is no "list" that consumers need to sign up for. Early discussion of Do Not Track included proposals about a list-based registry of users, similar to the Do Not Call Registry. This proposal does not collect data on consumers in a central list



DNT: Fear, Uncertainty, and Doubt

26

RECENT UPDATES

BLOG POSTS

EFF IN THE NEWS | November 15, 2011

[Advertisers Can't Be Trusted to Self-regulate on Data Collection, Says EFF](#)

DEEPLINKS BLOG | November 14, 2011

[The DAA's Self-Regulatory Principles Fall Far Short of Do Not Track](#)

DEEPLINKS BLOG | October 21, 2011

[An EFF Guide to the Silicon Valley Human Rights Summit](#)

EFF IN THE NEWS | October 6, 2011

[Kindle Fire's Silk browser raises issues over website tracking history of users](#)

DEEPLINKS BLOG | September 22, 2011

[EFF Advocates for User Privacy in W3C Workshop on Do Not Track](#)

EFF IN THE NEWS | August 18, 2011

[Do you know about your digital fingerprint?](#)

EFF IN THE NEWS | August 18, 2011

[Tracking Your Every Move on the Internet](#)

DEEPLINKS BLOG | July 7, 2011

[EFF Urges Senators to Recognize Need for Updated Privacy Laws](#)

EFF IN THE NEWS | May 27, 2011

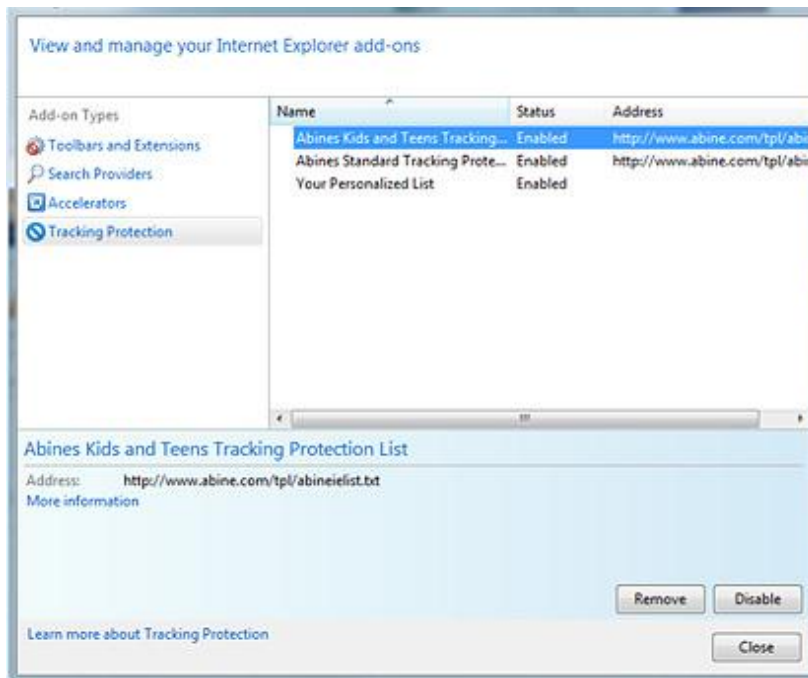
['Like' Button Follows Web Users](#)

EFF IN THE NEWS | May 27, 2011

[US Senate Sinks Its Teeth Into Online Privacy Reform](#)

Tracking Protection Lists (TPLs)

27



Abine, The Online Privacy Company, is the leading provider of online privacy solutions for consumers. Abine's products and services allow regular people to regain control over their personal information while continuing to browse, interact, and shop online.

Abine's Tracking Protection List blocks many online advertising and marketing technologies that can track and profile you as you browse the Web. This list is updated weekly to keep you safer and more private.

[Visit the Abine website for more information about this Tracking Protection List.](#)



EasyPrivacy Tracking Protection List is based on the popular EasyPrivacy subscription for Adblock Plus and is managed by the well-known EasyList project, which serves nearly ten million daily users and has a large support forum with dozens of experienced members able to assist resolving any issues that may arise.

[Visit the EasyList website for more information about this Tracking Protection List.](#)



PrivacyChoice maintains a comprehensive database of tracking companies, including domains used by nearly 300 ad networks and platforms, tracking methods, summaries of key policies, oversight, and opt-out and opt-in processes.

PrivacyChoice has created Tracking Protection Lists based on this data. You have the option of installing two lists. The first list blocks companies that are not subject to oversight by the NAI and the second list blocks all tracking company domains in the PrivacyChoice database. These lists will be automatically updated with new tracking domains discovered through continuous website scanning and user panels.

[Visit the PrivacyChoice website for more information about this Tracking Protection List.](#)



TRUSTe is the leading online privacy certification and services provider. TRUSTe's TRUSTed Tracking Protection List enables relevant and targeted ads from companies that demonstrate respectful consumer privacy practices and comply with TRUSTe's high standards and direct oversight. TRUSTe helps users get good ads, without compromising personal privacy.

[Visit the TRUSTe website for more information about this Tracking Protection List.](#)

Tracking Protection Lists (TPLs)

How do they work?

- The websites you visit often contain content from third parties. In order to load this content, certain information about your computer, including your IP address and the address of the webpage you're viewing, is sent to each of the third parties. If a site is listed as a "do not call" site on a TPL, Internet Explorer 9 will block third-party content from that site, unless you visit the site directly by clicking on a link or typing its web address. By limiting "calls" to third-party websites, Internet Explorer 9 limits the information these third-party sites can collect about you.

Do TPLs only block third-party calls?

- TPLs can include "do not call" or "OK to call" entries that permit calls to specific third-party sites. Please be aware that if there are conflicts between "do not call" and "Ok to call" TPLs, the "Ok to call" rules will govern. You should review carefully the TPLs that you choose to download to ensure that you want to allow calls to each of the sites included in any "Ok to call" list.
-

Privacy in the News

- Concerns about tracking
- Personal data siloed away
- Browser features help
- Legislative pressure



Question of the Day

What are some of the reasons for the outrage caused by third-party tracking?

RePriv



Re-Envisioning In-Browser
Personalization & Privacy

[Oakland S&P 2011]

Ben Livshits

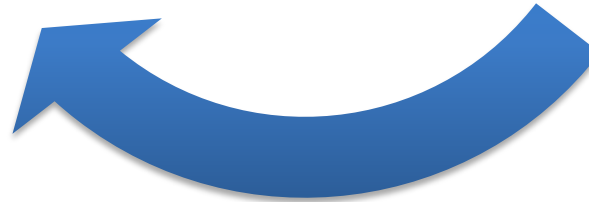
Microsoft Research

users want a
highly
personalized
web
experience



**Share data to get
personalized
results**

**Privacy
concerns**



Browser: Personalization & Privacy



Browsing history

Top: Computers: Security
Top: Arts: Movies
Top: Sports: Hockey
Top: Science: Math
Top: Recreation: Outdoors



Distill



User interest profile

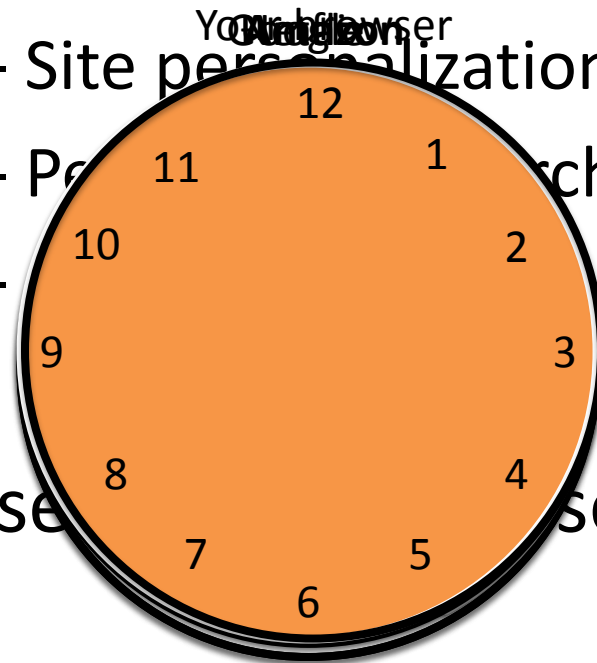
- Broad applications:

– Site personalization

– Personalized search

– User interface

- Control information release



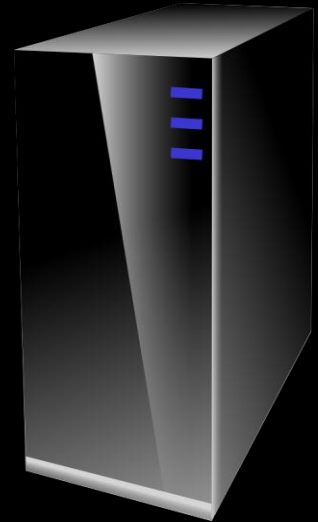
bn.com would like to learn your top interests.
We will let them know you are interested in:

- Science
- Technology
- Outdoors

Accept

Decline

RePriv Protocol



GET /index.html HTTP 1.1
Host: www.example.com
Accept: repriv ...

HTTP/1.1 **300 Multiple Choices**
index.html
index.html?top-n&level=m

POST /index.html HTTP 1.1
Host: www.example.com
Content-Length: x
category1=c1&...

HTTP/1.1 200 OK

Personalized page content

Would you like to install an extension called “Bing Personalizer” that will:

- Watch mouse clicks on bing.com
- Modify appearance of bing.com
- Store personal data in browser

Accept

Decline

Contributions of RePriv

RePriv

- An in-browser framework for collecting & managing personal data to facilitate personalization.

Core Behavior Mining

- Efficient in-browser behavior mining & controlled dissemination of personal data.

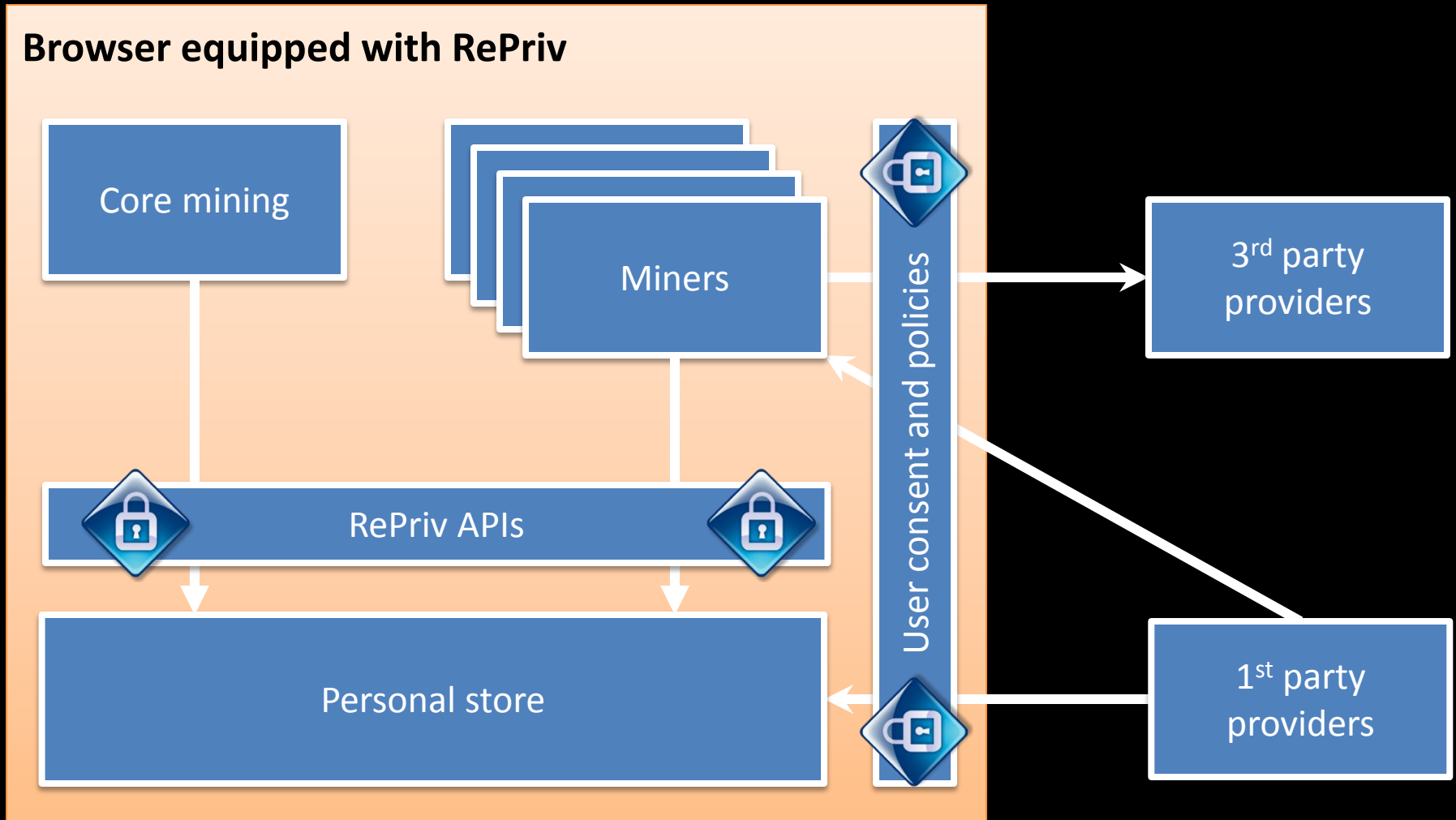
RePriv miners

- A framework for integrating verified third-party code into the behavior mining & dissemination of RePriv.

Real-world Evaluation

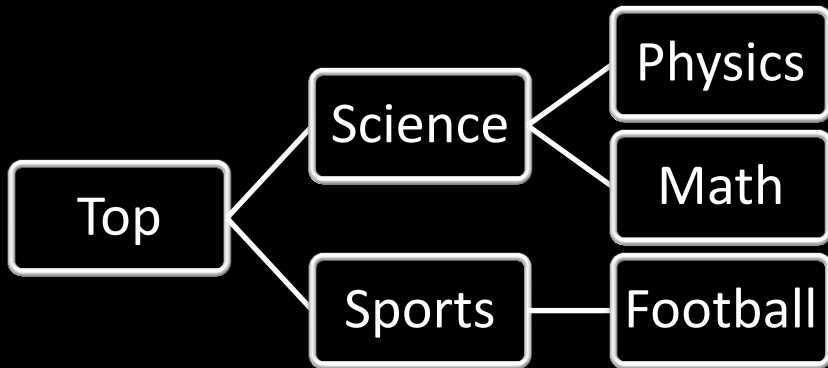
- Evaluation of above mechanisms on real browsing histories & two in-depth case studies.

RePriv Architecture



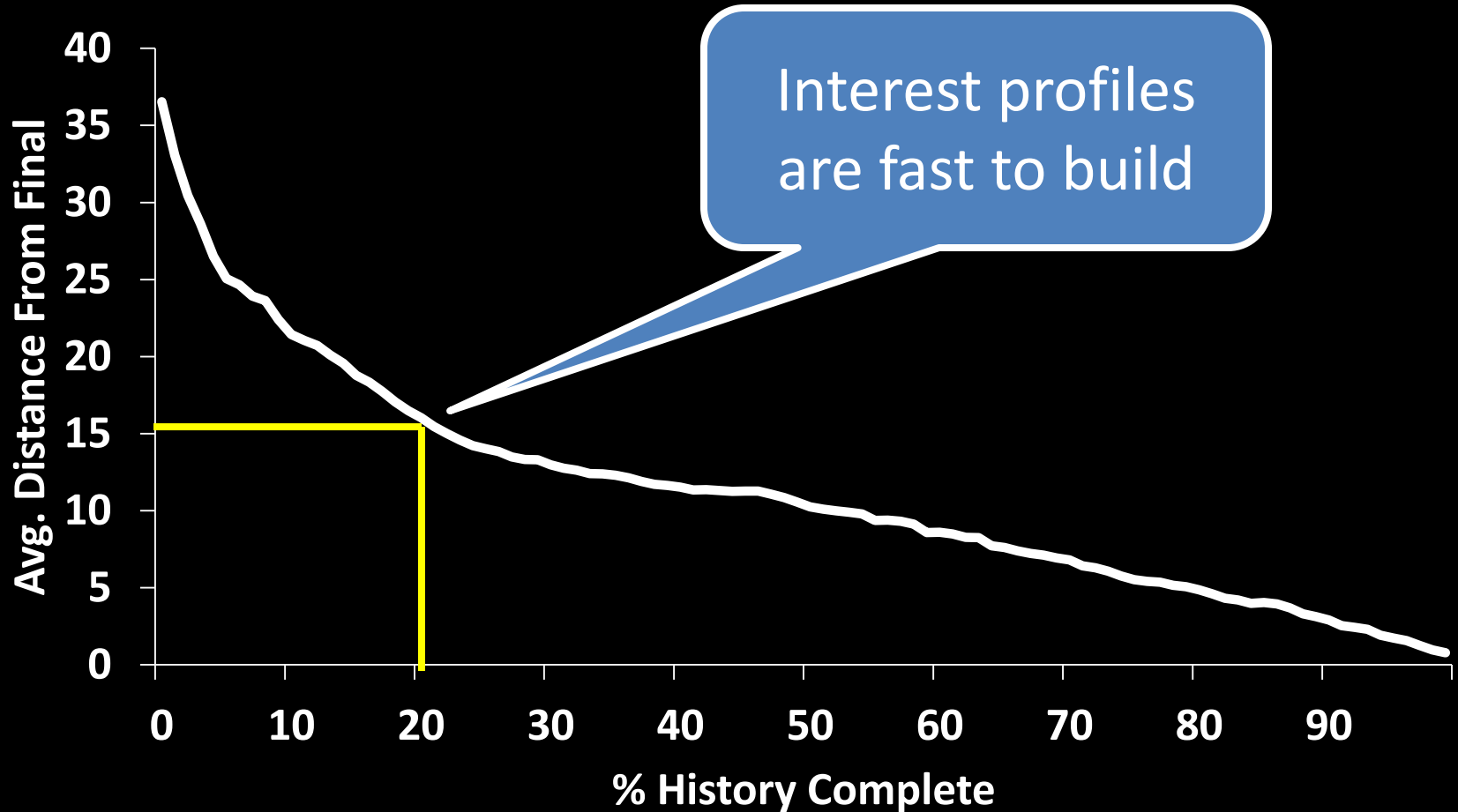
Core Mining

- Taxonomy from first two levels of ODP taxonomy
 - ~450 categories total
 - 20 top-level categories
 - Overlap exists



- Naïve Bayes
 - All categories equally likely
 - Training: min(3000, # pages) sites per category
 - Attribute words occur in at least 15% of docs for ≥ 1 category
- Classification is fast enough: $O(c \cdot n)$
 - n is # words in document
 - c is # document categories

Global Mining Convergence



RePriv

- An in-browser framework for collecting & managing personal data to facilitate personalization.

Core Behavior Mining

- Efficient in-browser behavior mining & controlled dissemination of personal data.

RePriv miners

- A framework for integrating verified third-party code into the behavior mining & dissemination of RePriv.

Real-world Evaluation

- Evaluation of above mechanisms on real browsing histories & two in-depth case studies.

Miner Name	C# LoC	Fine LoC	Verif. Time
TwitterMiner	89	36	6.4
BingMiner	78	35	6.8
NetflixMiner	112	110	7.7
GlueMiner	213	101	9.5

assume ExtensionId "twitterminer"

assume CanCommunicateXHR "twitter.com" Nil

assume CanUpdateStore("twitter.com" "twitterminer")

Netflix Example

```
let doGetMovies genre cdom =
```

...

```
let flixEnts = GetStoreEntriesByTopic
```

```
  myprov "movie" in
```

```
let genreFlix = bind myprov flixEnts
```

```
  (filterByGenre genre) in
```

```
  ExtensionReturn cdom myprov genreFlix
```

14 lines of Fine code

```
...
let doGetMovies genre cdom =
  ...
  let flixEnts = GetStoreEntriesByTopic
    myprov "movie" in
  let genreFlix = bind myprov flixEnts
    (filterByGenre genre) in
  ExtensionReturn cdom myprov genreFlix
...
assume CanReadDOMClass "netflix.com" "rv5"
assume CanCaptureEvents "onclick" (P "netflix.com" "netflixminer")
assume CanServeInformation "fandango.com" (P "netflix.com" "netflixminer")
assume CanServeInformation "amazon.com" (P "netflix.com" "netflixminer")
assume CanServeInformation "metacritic.com" (P "netflix.com" "netflixminer")
assume CanHandleSites "netflix.com"
assume CanReadStore (P "netflix.com" "netflixminer")
assume CanReadLocalFile "moviegenres.txt"
...
```

RePriv

- An in-browser framework for collecting & managing personal data to facilitate personalization.

Core Behavior Mining

- Efficient in-browser behavior mining & controlled dissemination of personal data.

RePriv miners


- A framework for integrating verified third-party code into the behavior mining & dissemination of RePriv.

Real-world Evaluation

- Evaluation of above mechanisms on real browsing histories & two in-depth case studies.

Privacy-Aware News Personalization


Map RePriv interest taxonomy to del.icio.us topics



Query personal store for top interests



Ask del.icio.us API for “hot” stories in appropriate topic areas from nytimes.com

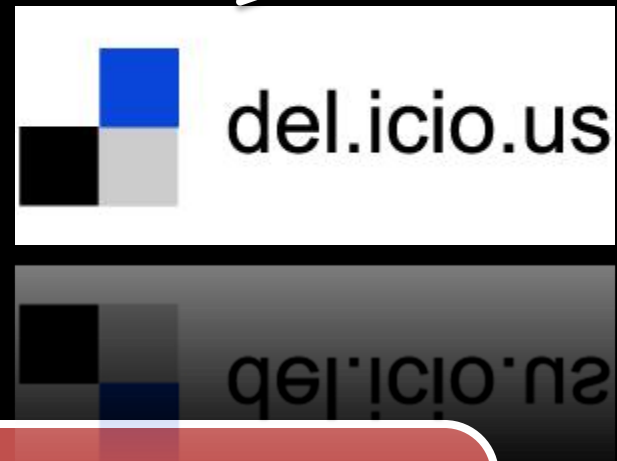
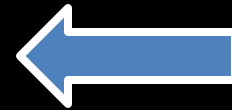


Replace nytimes.com front page with del.icio.us stories

Privacy Policy



Change "href" attribute of anchor elements on nytimes.com



Change TextContent of selected anchor and div elements on nytimes.com

User profile:

- Games/Card_Games
- Games/Conventions
- Games/Video_Games

Do Video Games Equal Less Crime?

That's one theory for the continuing fall in crime, despite the recession.

Gamers Finally Get Their Wheaties Box ...sort of

Dr Pepper is featuring the Halo 3 player Tom Taylor, who goes by Tsquared, on the labels, which will appear on about 175 million 20-ounce bottles from January to April.

Sony To Shut its SF Metreon PlayStation store

Sony is closing down its one-and-only U.S. PlayStation store at the Metreon mall in San Francisco. The recession is clearly to blame, but it's happening at time when Microsoft - which opened and shut its own Microsoft store at the Metreon - is going to open a chain of its own stores.

Microsoft Takes on Cable With Xbox Streaming Video

If talks with Disney work out, the game console could stream ESPN content, making it that much easier to watch TV without cable.

Some Video Gamers Leery of Obama's Views

Gamers are worried that the president-elect's positions on video games may signal new regulations or restrictions on the industry.

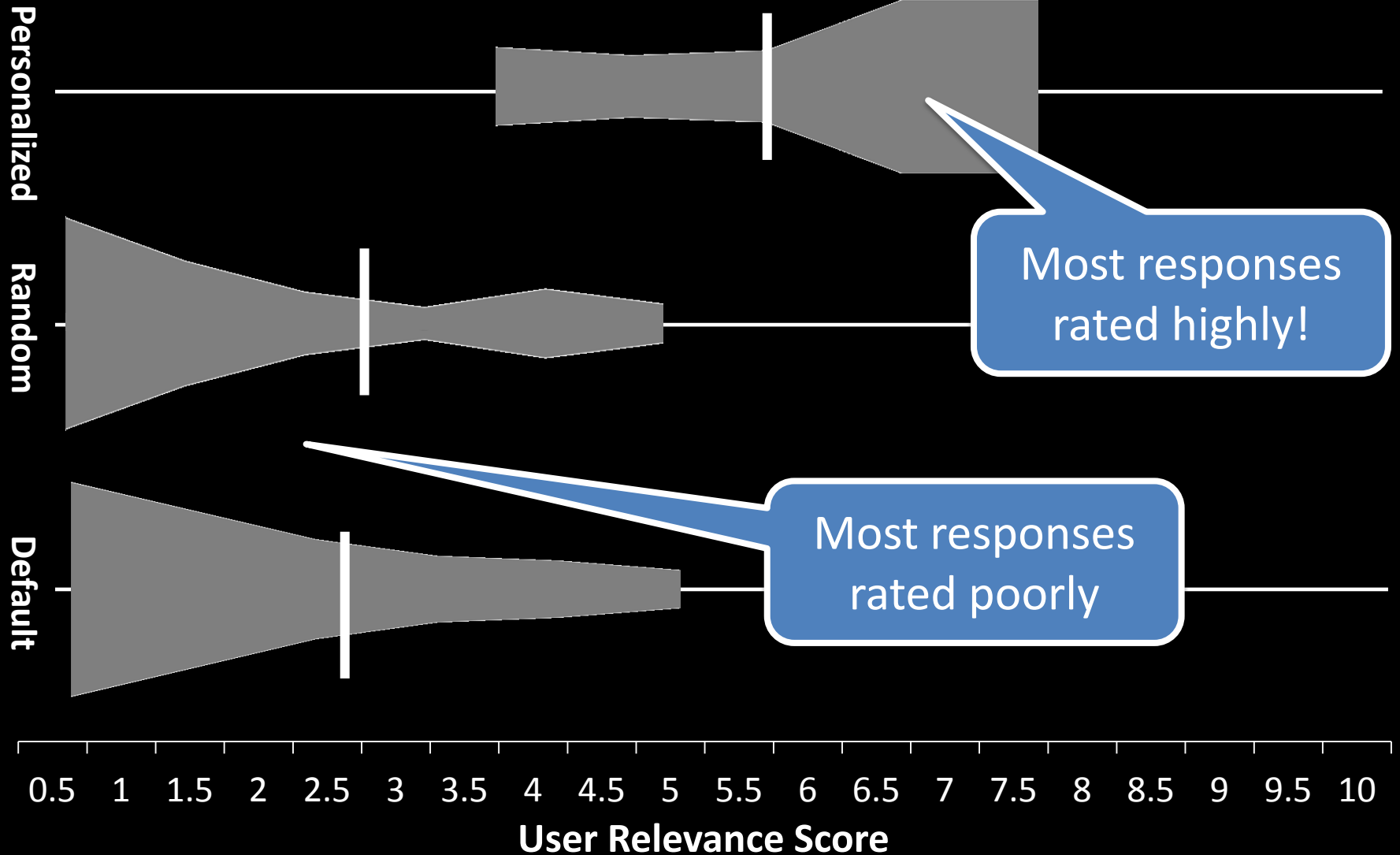
Relevance: (required)

Technology
Technology
Science/C
Science/P

- 2,2
- Ov
- Ty



News Personalization: Effectiveness



Most responses rated highly!

Most responses rated poorly

RePriv Summary

- Existing solutions require privacy sacrifice
- RePriv is a browser-based solution
 - User retains control of personal information
 - High-quality information mined from browser use
 - General-purpose mining useful & performant
 - Flexibility with rigorous guarantees of privacy
- Personalized content & privacy can coexist
- See our Oakland papers and W2SP papers

1. Introduction

The motivation of this work comes from the observation that in today's web there are two distinct groups, users and service providers such as Amazon, Google, Microsoft, Facebook and the like, as they can so that they can better target their ads or provide personalization. Users might welcome content, advertisements as long as it does not compromise their privacy. In today's web, for service providers, personalization is limited. Even if sites like Amazon sometimes require authentication, services are limited. A user might only spend a few minutes on a site. A user might only spend a few minutes on a site. A user might only spend a few minutes on a site. A user might only spend a few minutes on a site.

1. Let the browser know the user's history
2. Let the browser know the user's history

Summary

52

- Some of the current problems in online privacy
- Tracking mechanisms
 - ▣ Cookies
 - ▣ Beacons
 - ▣ Fingerprinting
- Dangers of third-party tracking
- Ad ecosystem and user targeting
- Solutions for tracking prevention
- RePriv: combining personalization and privacy