Privacy

Presenter: Yinzhi Cao
Lehigh University

Some contents are borrowed from the following sources.

- http://www.cs.umd.edu/projects/privacy-curriculum/Differential_Privacy.pptx
- https://www.cs.utexas.edu/~shmat/courses/cs361s/webtrack.ppt

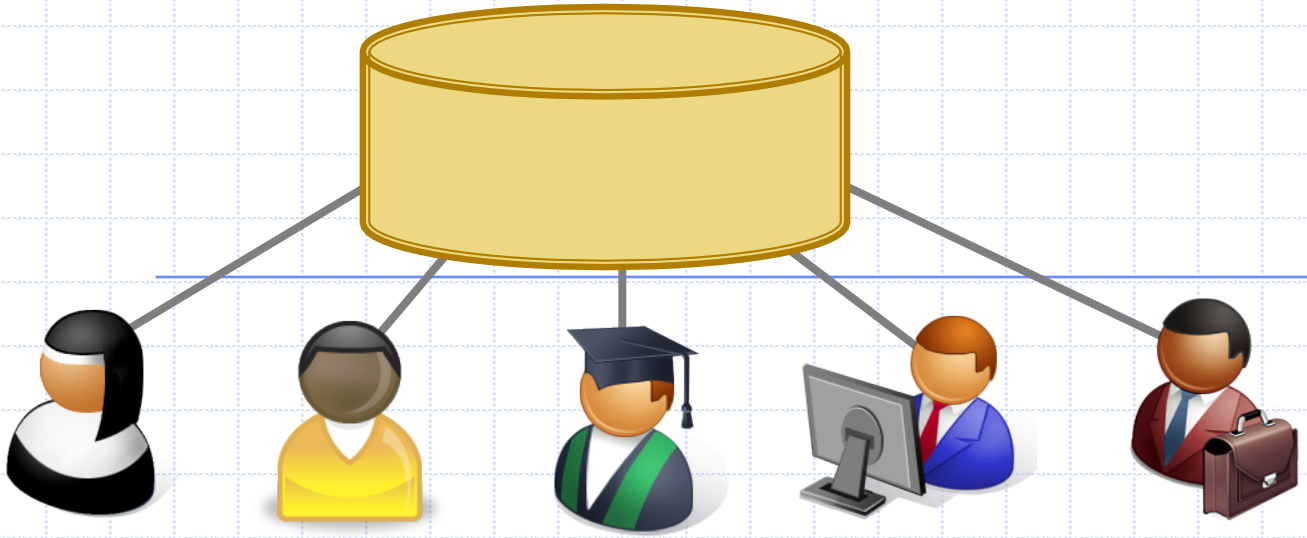
Outline

- Differential Privacy
- Web Privacy
- TrackingFree

General Setting

Medical data
Query logs
Social network data
…

Data mining
Statistical queries



General Setting

How can you allow meaningful usage of such datasets while preserving individual privacy?

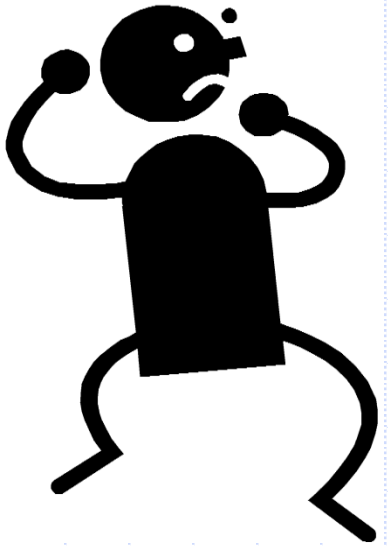
Blatant Non-Privacy

Blatant Non-Privacy

- Leak individual records
- Can link with public databases to re-identify individuals
- Allow adversary to reconstruct database with significant probablity

Attempt 1: Crypto-ish Definitions

I am releasing some useful statistic f(D), and nothing more will be revealed.



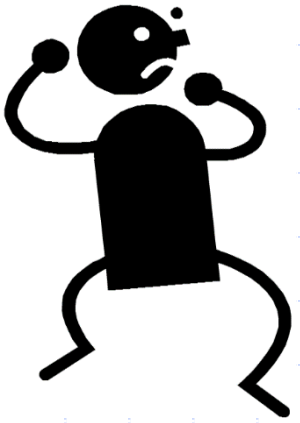
What kind of statistics are safe to publish?

How do you define privacy?

Attempt 2:

I am releasing researching findings showing that people who smoke are very likely to get cancer.

You cannot do that, since it will break my privacy. My insurance company happens to know that I am a smoker…



Attempt 2: Absolute Disclosure Prevention

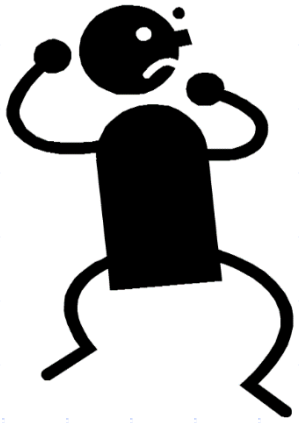
"If the release of statistics *S makes it possible to* determine the value [of private information] more accurately than is possible without access to *S*, a disclosure has taken place." [Dalenius]

An Impossibility Result

[informal] It is not possible to design any non-trivial mechanism that satisfies such strong notion of privacy.[Dalenius]

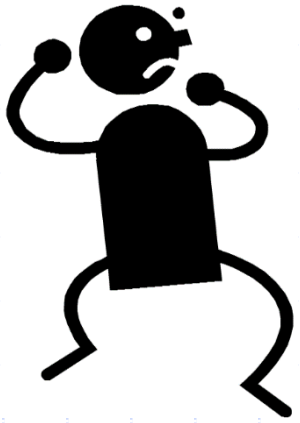
Attempt 3: "Blending into Crowd" or k-Anonymity

K people purchased A and B, and all of them also purchased C.



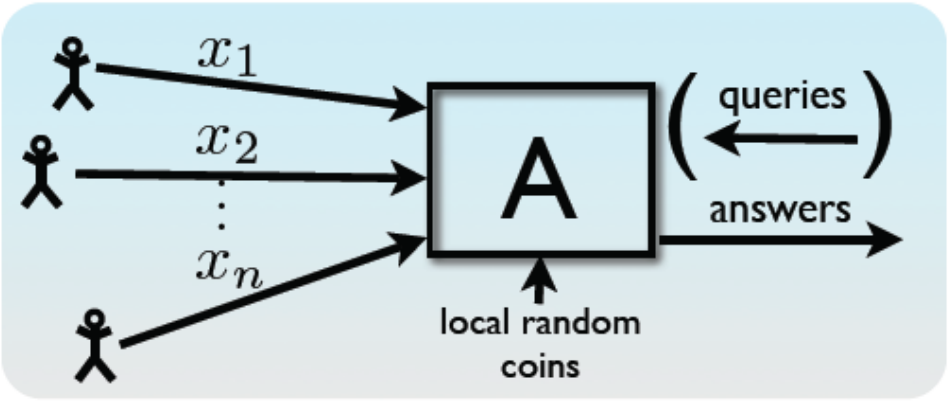
Attempt 3: "Blending into Crowd" or k-Anonymity

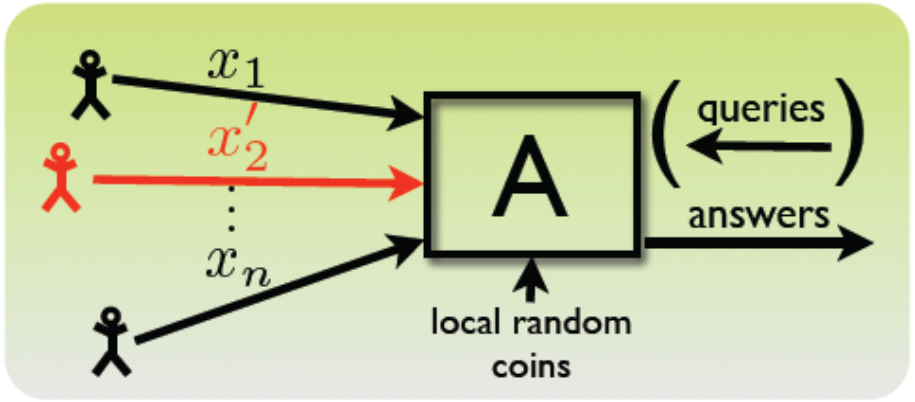
K people purchased A and B, and all of them also purchased C.



I know that Elaine bought A and B…

Attempt 4: Differential Privacy

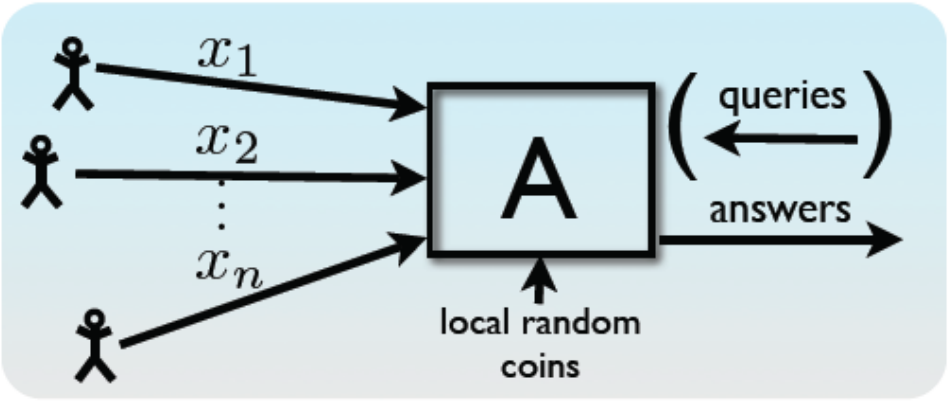


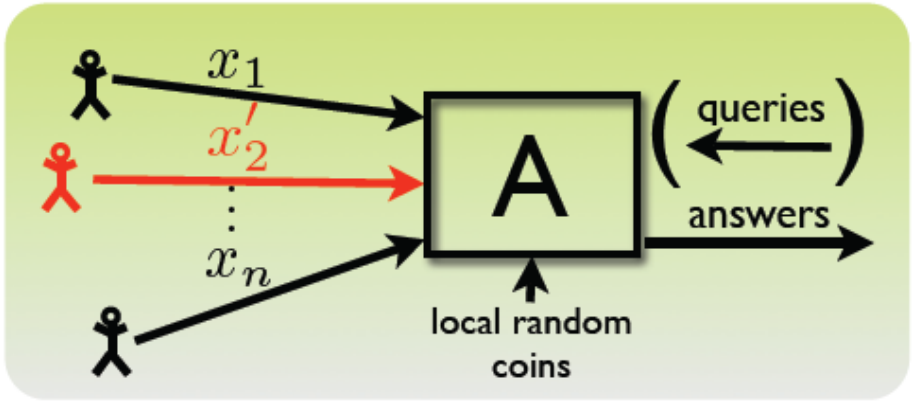


x' is a neighbor of x
if they differ in one row

From the released statistics, it is hard to tell which case it is.

Attempt 4: Differential Privacy





x' is a neighbor of x
if they differ in one row

For all neighboring databases x and x'
For all subsets of transcripts:

$\Pr[A(x) \in S] \leq e^{\varepsilon} \Pr[A(x') \in S]$

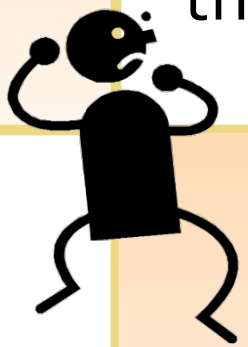
Attempt 4: Differential Privacy

I am releasing researching findings showing that people who smoke are very likely to get cancer.

1

Please don't blame me if your insurance company knows that you are a smoker, since I am doing the society a favor.

2



Oh, btw, please feel safe to participate in my survey, since you have nothing more to lose.

3

Since my mechanism is DP, whether or not you participate, your privacy loss would be roughly the same!

4

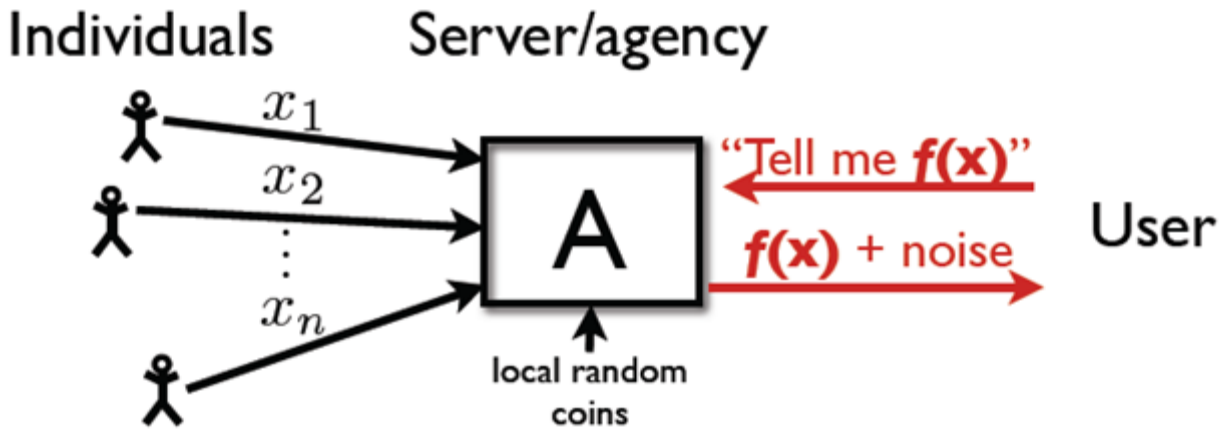
Notable Properties of DP

- Adversary knows arbitrary auxiliary information
  - No linkage attacks
- Oblivious to data distribution
- Sanitizer need not know the adversary's prior distribution on the DB

Techniques for Achieving DP

- **Output perturbation**
- Input perturbation
- Perturbation of intermediate values
- Sample and aggregate

Method: Output Perturbation



- Global Sensitivity:

$$GS_f = max\|f(x) - f(x')\|_1$$

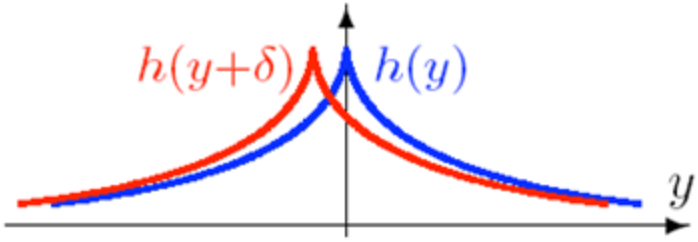
x,x′ neighbors

Example: $GS_{avg} = \frac{1}{n}$

Method: Output Perturbation

$A(x) = f(x) + \mathrm{Lap}(\frac{GS_f}{\epsilon})$ is ϵ-DP

Laplace distribution $\mathsf{Lap}(\lambda)$ has density $h(y) \propto e^{-\frac{\|y\|_1}{\lambda}}$



Sliding property of $\mathsf{Lap}\left(\frac{\mathsf{GS}_f}{\varepsilon}\right)$: $\frac{h(y)}{h(y+\delta)} \leq e^{\varepsilon \cdot \frac{\|\delta\|}{\mathsf{GS}_f}}$ for all y, δ

Proof idea:

$A(x)$: blue curve

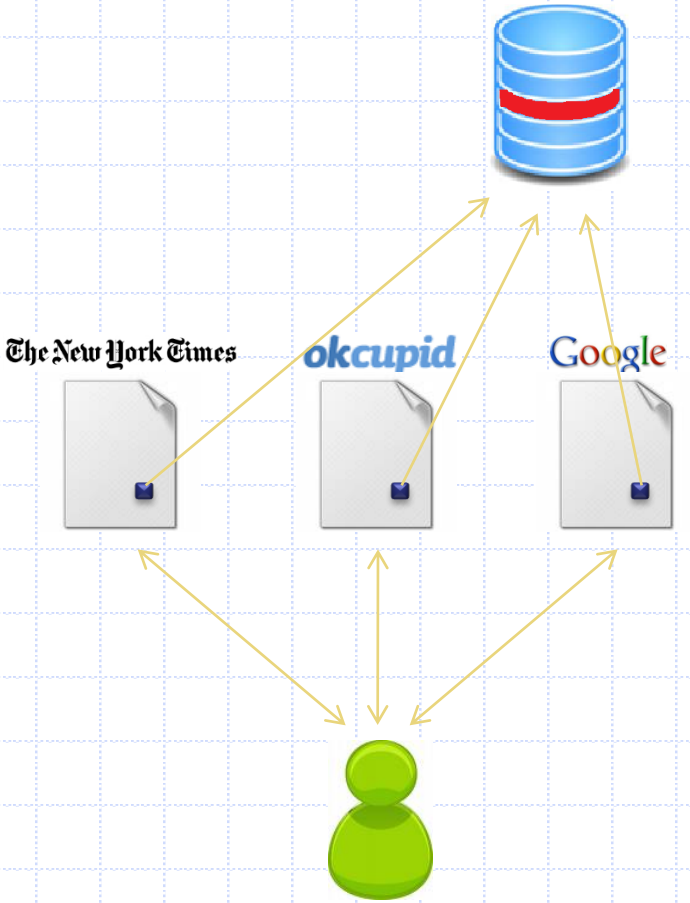
$A(x')$: red curve

$\delta = f(x) - f(x') \leq \mathsf{GS}_f$

Web Privacy

It's the Internet! Of course they know you're a dog. They also know your favorite brand of pet food and the name of the cute poodle at the park that you have a crush on!

Third-Party Tracking

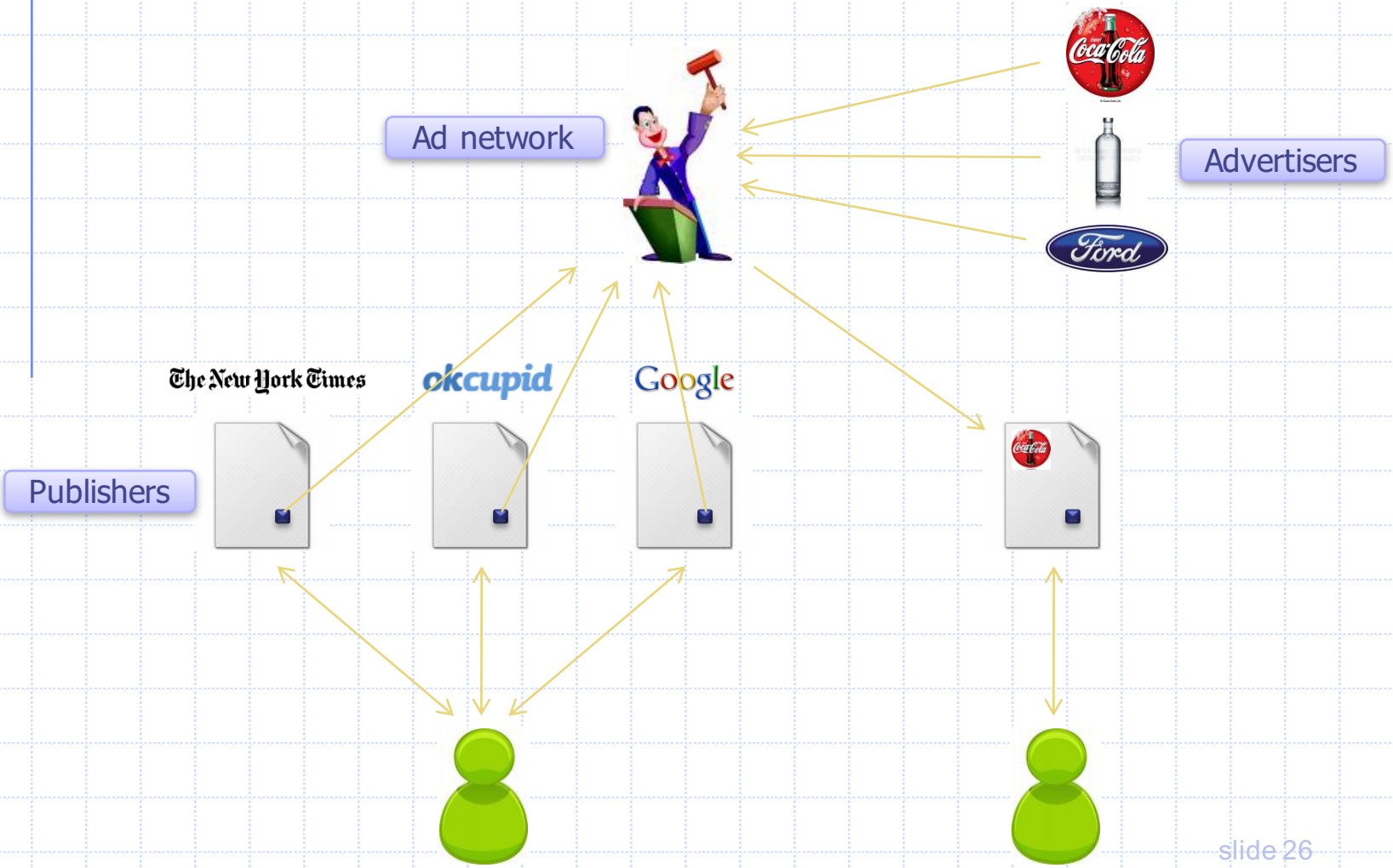


Third-party cookies:

- Disabled by default (Safari)
- Can be disabled by user (many browsers)
- Cannot be disabled (Android)

… but there are many other tracking technologies

Behavioral Targeting



Partial List of Ad Networks

24/7 Real Media	33Across	Acerno	Acxiom Relevence-X	AdAdvisor	AdBrite
Adify	AdInterax (Yahoo!)	AdJuggler	AdShuffle	ADTECH (AOL)	Advertising.com (AOL)
Aggregate Knowledge	Akamai	AlmondNet	Atlas (Microsoft)	AudienceScience	Bizo
Blue Kai	BlueLithium (Yahoo!)	Bluestreak	BrightRoll	BTBuckets	Burst Media
Casale Media	Chitika	ChoiceStream	ClickTale	Collective Media	comScore VoiceFive
Coremetrics	Cossette	Criteo	Effective Measure	Eloqua	Eyeblaster
eXelate	EyeWonder	e-planning	Facilitate Digital	FetchBack	Flashtalking
Fox Audience Network	FreeWheel	Google	Hurra	interCLICK	Lotame
Navegg	NextAction	NexTag	Mediaplex (ValueClick Media)	Media 6 Degrees	Media Math
Microsoft	MindSet Media	Nielsen Online	nugg.ad	Omniture	OpenX
Outbrain	PointRoll	PrecisionClick	Pulse 360	Quantcast	Quigo (AOL)
richrelevance	Right Media (Yahoo!)	Rocket Fuel	Safecount *	ScanScout	Smart Adserver
Snoobi	Specific Media	TACODA (AOL)	Tatto Media	Tealium	TradeDoubler
Traffic Marketplace	Tribal Fusion / Exponential	TruEffect	Tumri	Turn	Undertone Networks / Zedo
ValueClick Media	Vizu	Weborama	WebTrends	Yahoo!	[x+1]

Tracking Is Pervasive

64

independent tracking mechanisms in an average top-50 website

Sticky Tracking

Subverting same origin policy
(publisher also runs an ad network)
`ad.hi5.com = ad.yieldmanager.com`

Flash cookies 

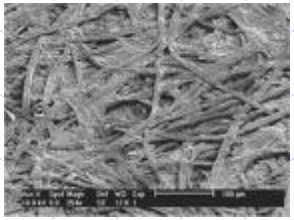
Browser fingerprinting 

History sniffing

Tracking Technologies

- HTTP Cookies
- HTTP Auth
- HTTP Etags
- Content cache
- IE userData
- HTML5 protocol and content handlers
- HTML5 storage
- Flash cookies
- Silverlight storage
- TLS session ID & resume
- Browsing history
- window.name
- HTTP STS
- DNS cache

Everything Has a Fingerprint



Fingerprinting Web Browsers

- User agent
- HTTP ACCEPT headers
- Browser plug-ins
- MIME support
- Clock skew
- Installed fonts
- Cookies enabled?
- Browser add-ons
- Screen resolution

A research project of the Electronic Frontier Foundation

Panopticlick

How Unique — and Trackable — Is Your Browser?

Is your browser configuration rare or unique? If so, web sites

Your browser fingerprint **appears to be unique** among the 3,435,834 tested so far

web.

Only anonymous data will be collected by this site.

TEST ME

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.

Learn about Panopticlick and web tracking. The Panopticlick Privacy Policy. Learn about the Electronic Frontier Foundation.

Panopticlick Example

Plugin 0: Adobe Acrobat; Adobe Acrobat Plug-In Version 7.00 for Netscape; nppdf32.dll; (Acrobat Portable Document Format; application/pdf; pdf) (Acrobat Forms Data Format; application/vnd.fdf; fdf) (XML Version of Acrobat Forms Data Format; application/vnd.adobe.xfdf; xfdf) (Acrobat XML Data Package; application/vnd.adobe.xdp+xml; xdp) (Adobe FormFlow99 Data File; application/vnd.adobe.xfd+xml; xfd). Plugin 1: Adobe Acrobat; Adobe PDF Plug-In For Firefox and Netscape; nppdf32.dll; (Acrobat Portable Document Format; e PDF in XML For) (XML Version applicat d). Plugin 2: Google Update; Google Update; npGoogleOneClick8.dll; (; application/x-vnd.google.oneclickctrl.8;). Plugin 3: MicrosoftÂ® Windows Media Player Firefox Plugin; np-mswmp; np-mswmp.dll; (np-mswmp; application/x-ms-wmp; *) (; application/asx; *) (; video/x-ms-asf-plugin; *) (; application/x-mplayer2; *) (; video/x-ms-asf; asf,asx,*) (; video/x-ms-wm; wm,*) (; audio/x-ms-wma; wma,*) (; audio/x-ms-wax; wax,*) (; video/x-ms-wmv; wmv,*) (; video/x-ms-wvx; wvx,*). Plugin 4: Move Media Player; npmnqmp 07103010; npmnqmp07103010.dll; (npmnqmp; application/x-vnd.moveplayer.qm; qmx,qpl) (npmnqmp; application/x-vnd.moveplay2.qm;) (npmnqmp; application/x-vnd.movenetworks.qm;). Plugin 5: Mozilla Default Plug-in; Default Plug-in; npnul32.dll; (Mozilla Default Plug-in; *; *). Plugin 6: Shockwave Flash; Shockwave Flash 10.0 r32; NPSWF32.dll; (Adobe Flash movie; application/x-shockwave-flash; swf) (FutureSplash movie; application/futuresplash; spl). Plugin 7: Windows Genuine Advantage; 1.7.0059.0; npLegitCheckPlugin.dll; (npLegitCheckPlugin; application/WGA-plugin; *).

84% of browser fingerprints are unique
With Flash or Java, 94% are unique

How Websites Get Your Identity

Third party is sometimes the site itself

Leakage of identifiers

```
GET http:/ /ad.doubleclick.net/adj/...
Referer: http:/ /submit.SPORTS.com/...?email=jdoe@email.com
Cookie: id=35c192bcfe0000b1...
```

Security bugs

Remember XSUH (cross-site URL hijacking)?

Third party buys your identity

Syphilis – NHS Choices

Syphilis – NHS Choices

http://www.nhs.uk/conditions/syphilis/pages/introduction.aspx

Google

Feedback

Home | About | Contact | Communities | Tools | Video | Choose and Book

Log in or create an account

NHS choices Your health, your choices

Enter a search term

Search

Health A-Z | Live Well | Carers Direct | Health news | Find and choose services

Syphilis

Share Save Easy print Like 5

Overview | Map of Medicine | Medicines info | Clinical trials

Syphilis | Symptoms | Causes | Diagnosis | Treatment | Complications | Prevention

Introduction

Syphilis is a bacterial infection that is usually passed on through having sex with someone who is infected. It can also be passed from an infected mother to her unborn child and, in rare cases, can be caught through injecting drugs.

It is extremely rare to catch syphilis through a blood transfusion in the UK as blood donors are carefully screened.

Three stages of disease

Stage 1 (primary syphilis). Symptoms of syphilis begin with a painless but highly infectious sore on the genitals or sometimes around the mouth. If somebody else comes into close contact with the sore, typically during sexual contact, they can also become infected. The sore lasts two to six weeks before disappearing.

Stage 2 (secondary syphilis). Secondary symptoms, such as a skin rash and sore throat, then develop. These symptoms may disappear within a few weeks, after which you experience a latent (hidden) phase with no symptoms, which can last for years. After this, syphilis can progress to its third, most dangerous stage.

Stage 3 (tertiary syphilis). At this stage, it can cause serious damage to the body.

The primary and secondary stages are when you are most infectious to other people. In the latent phase (and usually around two years after becoming infected), syphilis cannot be passed onto others but can still cause symptoms.
See Symptoms of syphilis for more information on the

Useful links

NHS Choices links

Video: gay healthcare
Video: condom negotiation
Live Well: condoms
Live Well: drugs
Health A-Z: HIV and AIDS
Health A-Z: STIs
Find sexual health services
Infections you can catch through oral sex

External links

British Association for Sexual Health and HIV
Brook: for under-25s
FPA: sexual health
Health Protection Agency: syphilis
Lab Tests Online: syphilis test
Men's Health Forum

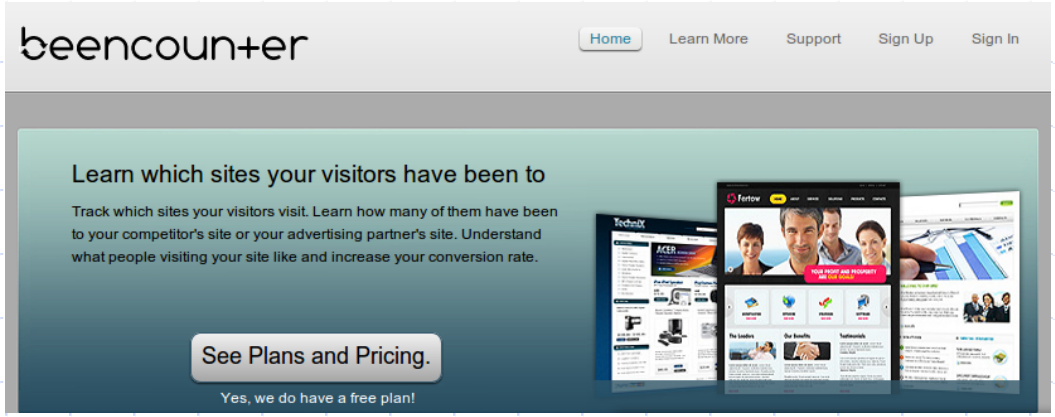
Screening and testing for gays and lesbians

Research shows that gay men and lesbians are less likely to have NHS screening and testing than heterosexuals. But it's important.

History Sniffing

How can a webpage figure out which sites you visited previously?

- Color of links
  - CSS :visited property
  - getComputedStyle()
- Cached Web content timing
- DNS timing



Do Not Track

Basics

HTTP header

- DNT: 1

Standardization

Browser support in FF4, IE9

Beginning to see adoption (AP, NAI)… or not

Privacy protections

No tracking across sites

- Who is the "third" party?

Can't be based on domain
Example: amazonaws.com, ad.hi5.com …

No intrusive tracking

Limits on regular log data

Exceptions for fraud prevention, etc.

DNT Adoption Issues

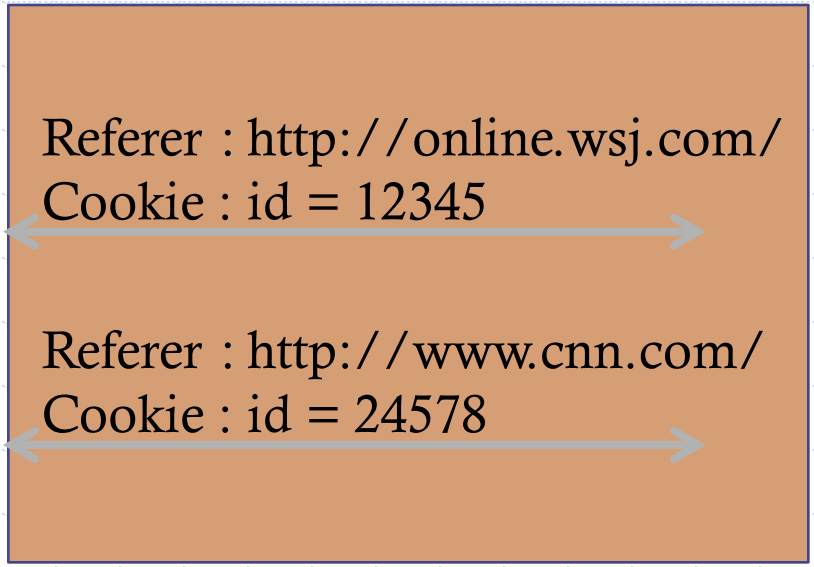
"But the NAI code also recognizes that companies sometimes need to continue to collect data for operational reasons that are separate from ad targeting based on a user's online behavior. For example, online advertising companies may need to gather data to prove to advertisers that an ad has been delivered and should be paid for; to limit the number of times a user sees the same ad; or to prevent fraud."

Translation: we're going to keep tracking you, but we'll simply call it "operational reasons."

TrackingFree

Goals and Challenges

- Anti-tracking Completeness
- Functionality/compatibility
- Performance

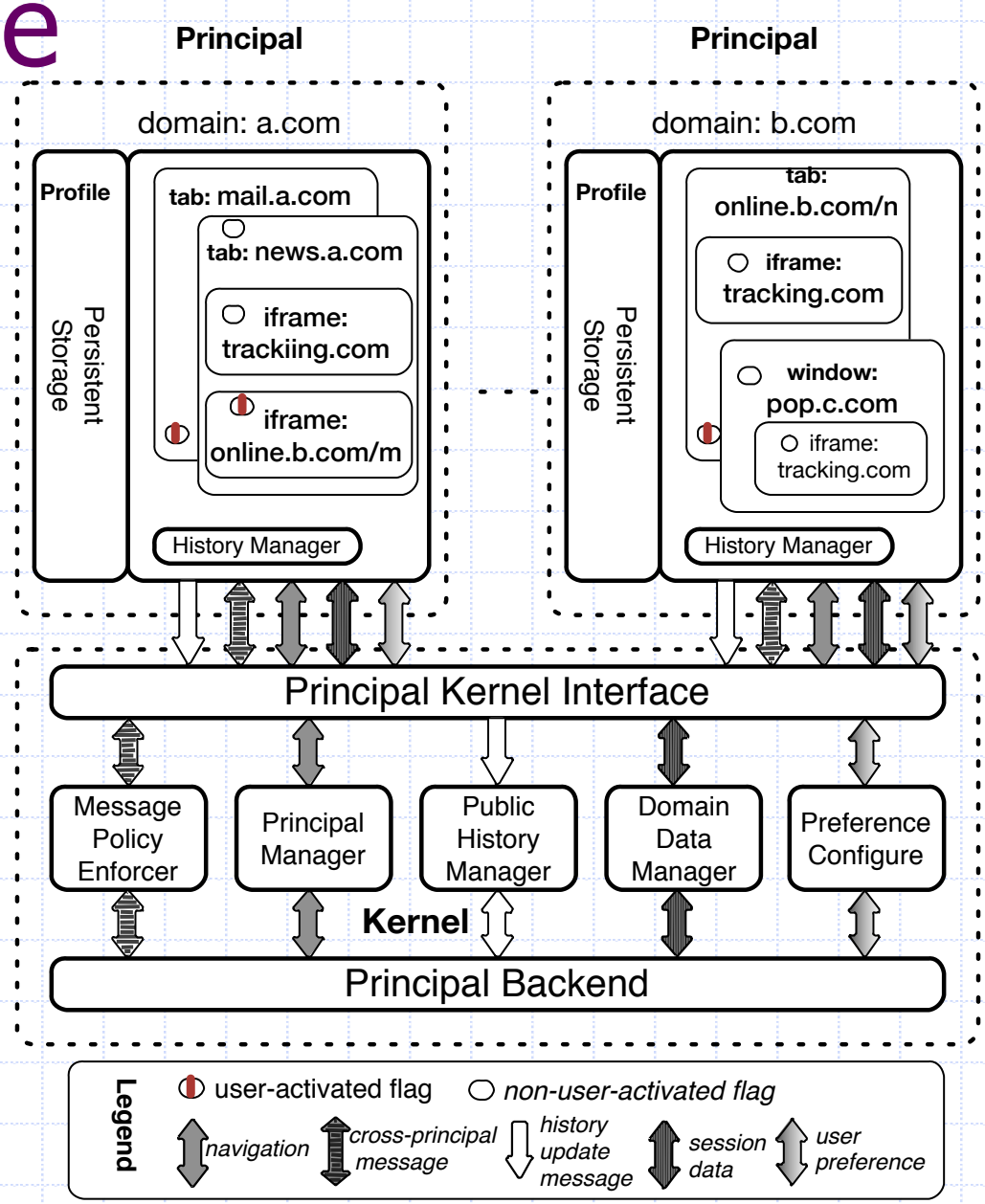


Core Idea : TrackingFree partitions client-side states into multiple isolation units so that the identifiers still exists but not unique any more!

Out-of-scope threats

- TrackingFree doesn't address following threats:
  - Within-Site Tracking.
  - Tracking by exploiting browser vulnerabilities
  - Stateless tracking.

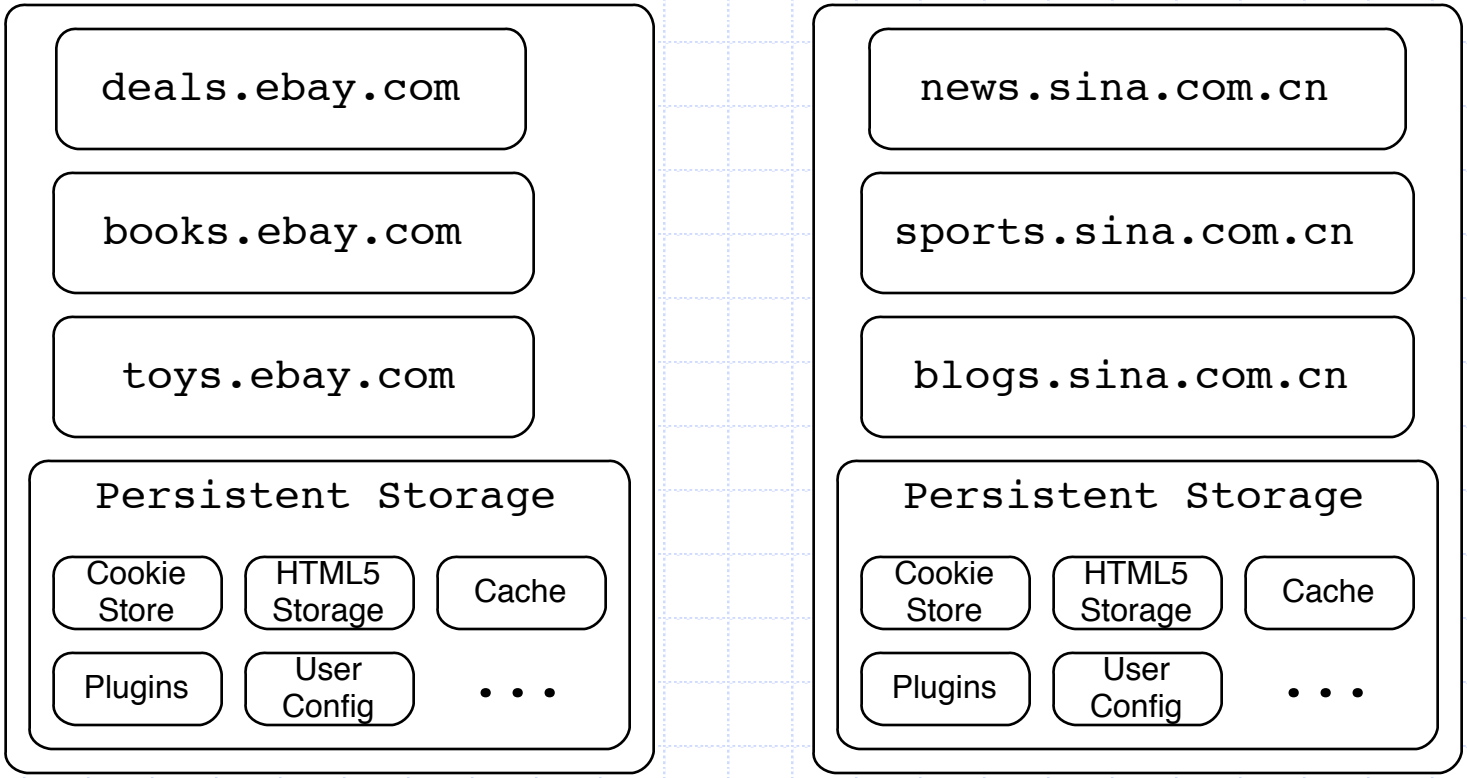
Architecture



Contents Allocation Mechanism

- Initial Contents Allocation
  - Handles those top frames that are navigated by users directly
- Derivative Contents Allocation
  - Handles those frames that are generated due to the contents on other frames, which we call child frame

Initial Contents Allocation

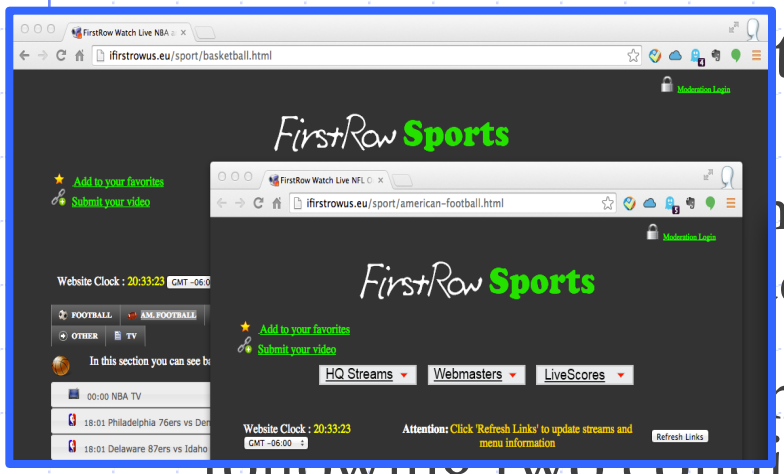


Derivative Contents Allocation

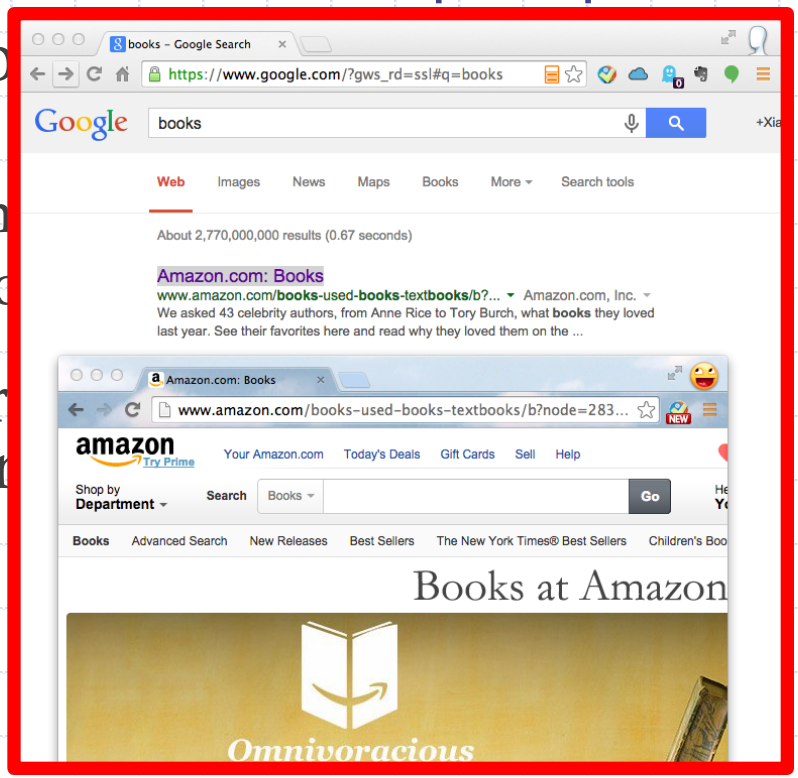
- Principal Switch
  - Should we switch principle for child frame?
- Principal Selection
  - How to choose target principal?

Principal Switch

Same principal

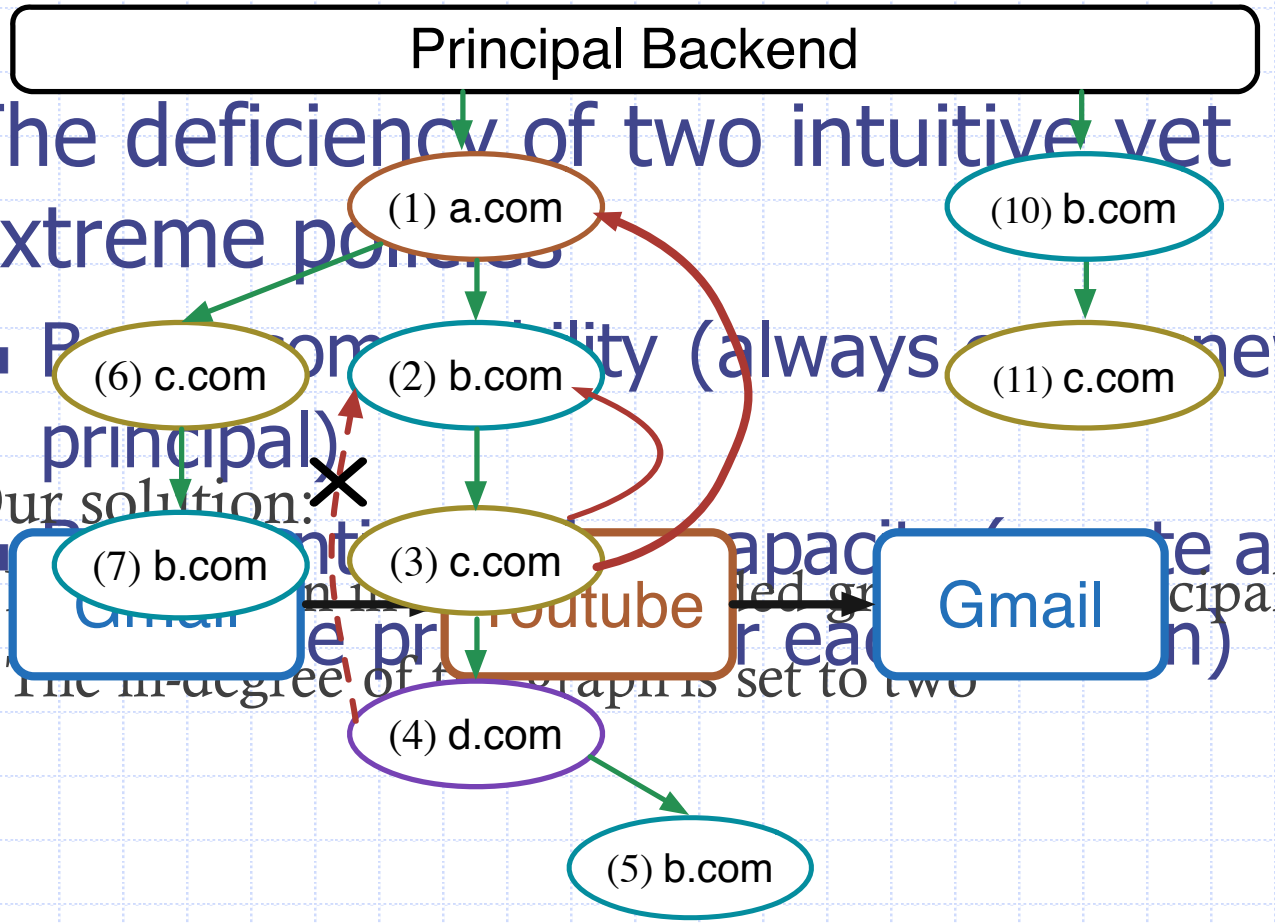


Different principal



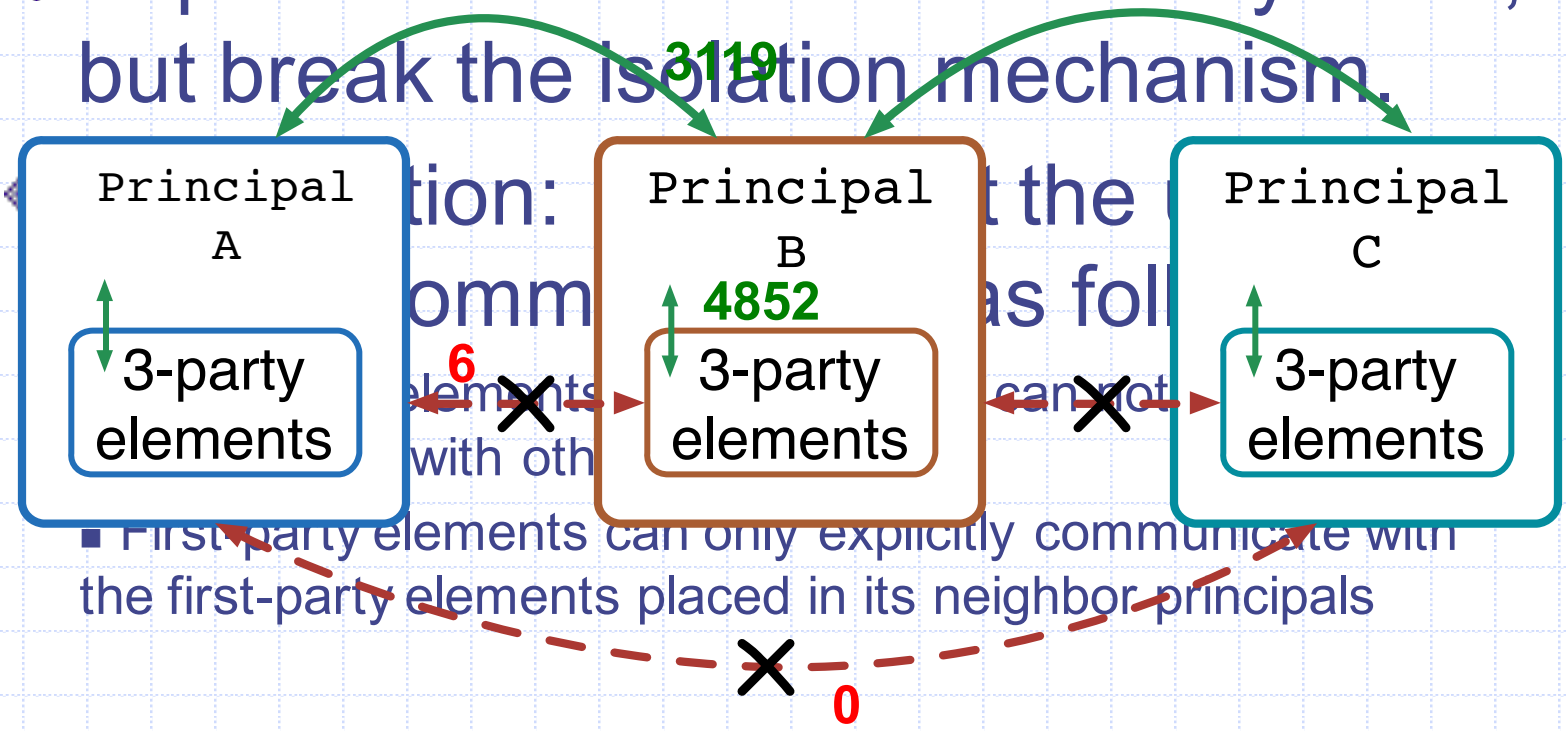
Principal Selection

- The deficiency of two intuitive yet extreme policies
  - P[illegible]om[illegible]ility (always [illegible] new principal)
- Our solution:
  - [illegible]
  - [illegible]
  - The in-degree of [illegible]



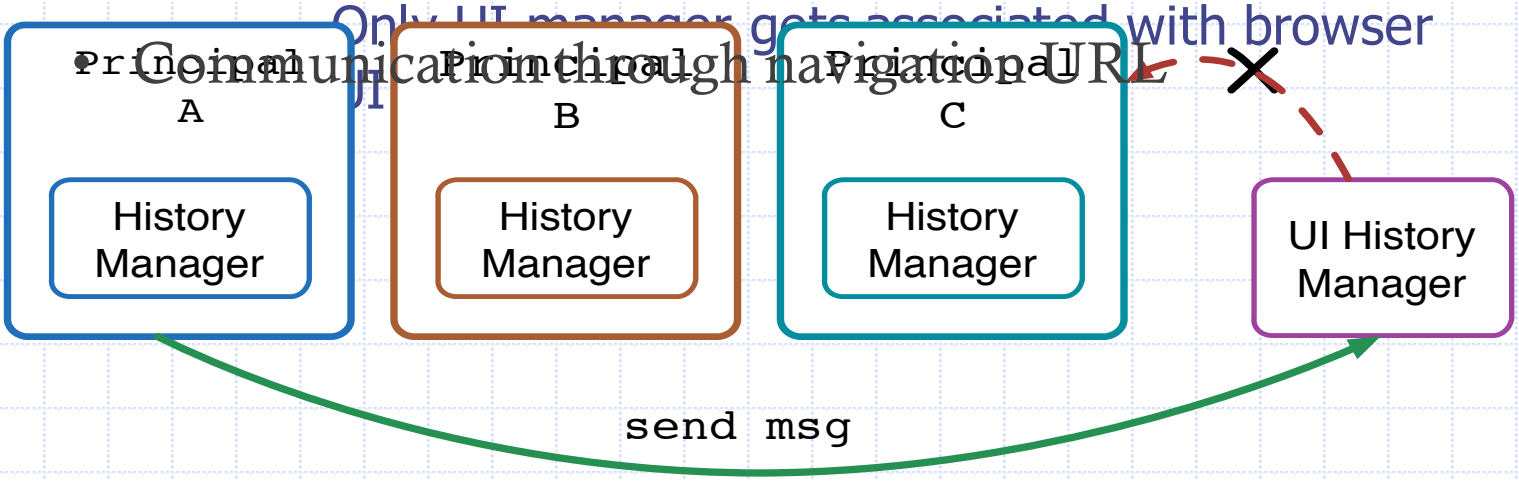
Principal Communication

- Explicit communication is widely used, but break the isolation mechanism.
- tion: t the
- omm as foll
  - elements can not with oth
  - First-party elements can only explicitly communicate with the first-party elements placed in its neighbor principals



Principal Communication

- Implicit Communication
  - History Sharing
    - UI history manager
    - Accepts information from other managers



Preference Configure

- User preference can be abused to store tracking identifier. (e.g. strict transport security)
- Completely isolating user preference affects user preference.
- Our solution:
  - Isolate user preference.
  - Apply user-initiated changes to all of the principals.
  - Monitor GUI message to determine user-initiated preference change.

Evaluation

- Anti-tracking capability
  - Formal proof
  - Experiments with real world websites
- Performance
  - Overhead (latency, memory, disk)
- Compatibility

Formal Proof

- Use Alloy to formally analyze TrackingFree 's anti-tracking ability.
  - Alloy is the most popular formal proof system

- Describe TrackingFree's behaviors on an existing Alloy Web model [Akhawe et al. CSF 2010].

- Formally verified trackers can correlate TrackingFree user's activities up to three principals without site collaboration.

Anti-tracking Capability with Real World Web Sites

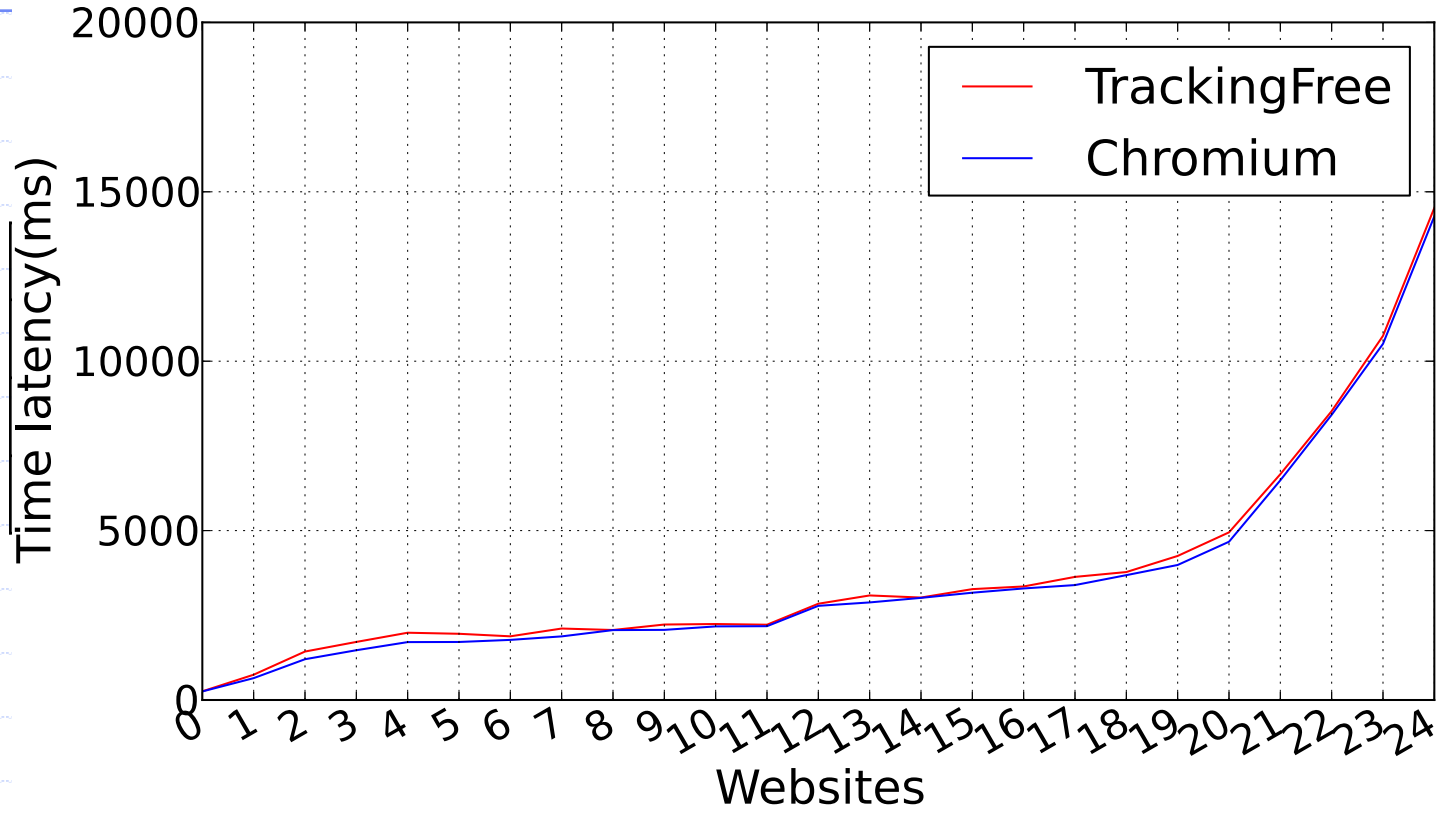
- Gathered tracking tokens on Alexa Top web sites by following the tracker detection of [Roesner et al. NSDI 2012].
- Detection based on the observation that each tracking request must contain the user's globally unique identifier.
- Some false negative, no false positive.

Anti-tracking Capability with Real World Web Sites

Tracking Host	Prevalence (# Domains)	Tracking Token(s)
b.scorecardresearch.com	133	UIDR
ad.doubleclick.ne	117	id, __gads
ib.adnxs.com	75	anj
p.twitter.com	70	__utma
cm.g.doubleclick.net	56	id
ad.yieldmanager.com	52	bx
bs.serving-sys.com	40	A4
cdn.api.twitter.com	40	__utmz
secure-us.imrworldwide.com	38	IMRID
adfarm.mediaplex.com	31	svid

Top 10 Tracking Hosts

Performance



(2). Address Bar Navigation Principal
Avg. Overhead 4.76%

(4). Bookmark Navigation Principal
Avg. Overhead 8.29%

Memory/Disk Overhead

Memory Overhead on 12 Web Pages (~25MB/Principal)

Memory	Chromium	TrackingFree	Increase
1 Principal	477.1(MB)	505(MB)	27.9(MB)
4 Principals	623.6(MB)	702.8(MB)	79.2(MB)
12 Principals	434.6(MB)	642.5(MB)	297.9(MB)

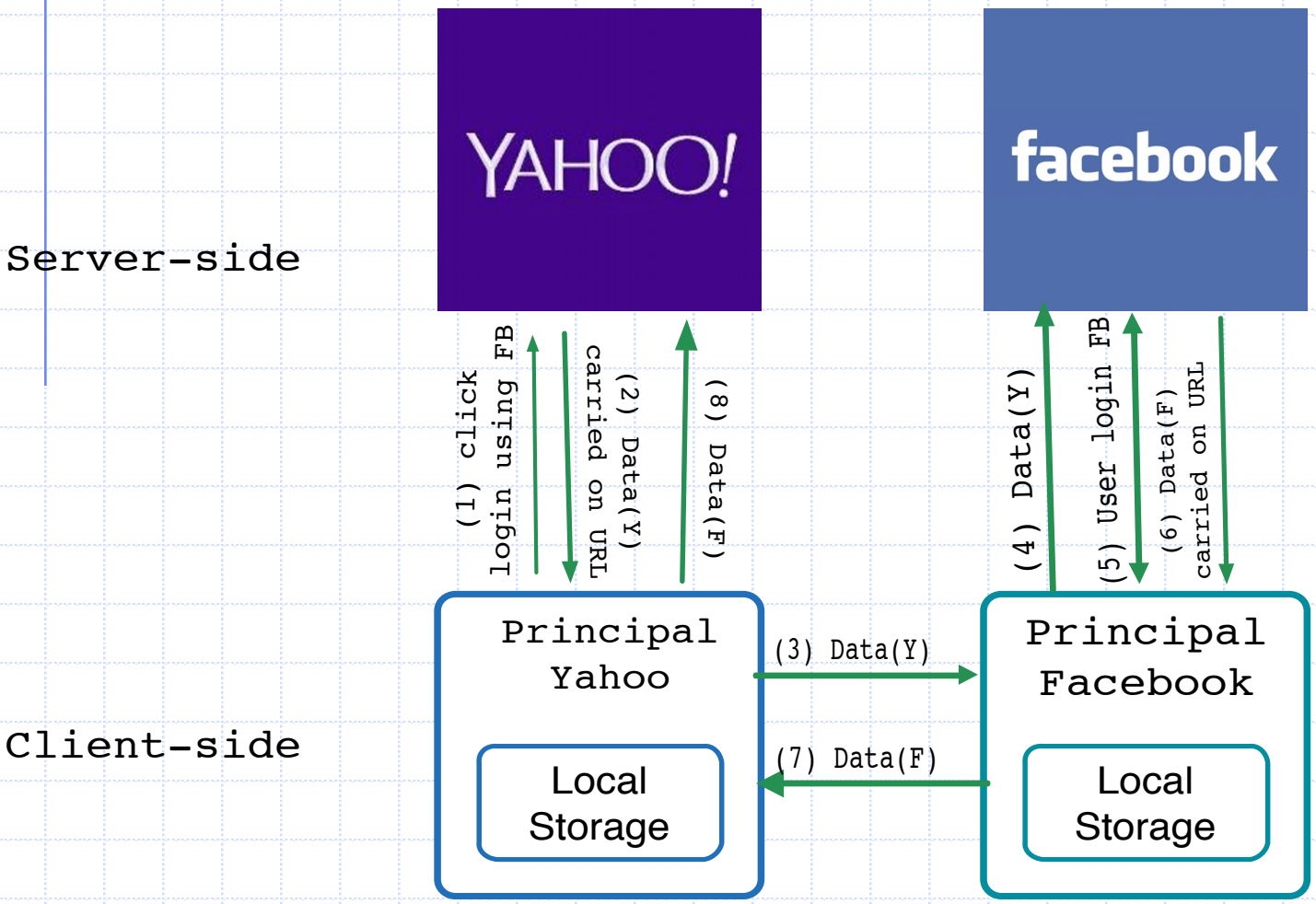
Disk Overhead on 12 Web Pages (~0.6MB/Principal)

Memory	Chromium	TrackingFree	Increase
1 Principal	21.3(MB)	21.8(MB)	0.5(MB)
4 Principals	22.5(MB)	25.9MB)	3.4(MB)
12 Principals	23.7(MB)	29.4(MB)	5.7(MB)

Compatibility

- Manually tested TrackingFree's compatibility on Alexa Top 50 websites
- Compatibility on first-party websites
  - Results: 50/50
- Compatibility on third-party services
  - Cross-site online payments (1/1)
  - Cross-site content sharing (31/31)
  - Single sign-on (35/36)
  - Overall results: 67/68

Case study: Logging Yahoo using Facebook Account



Summary

- We designed and implemented TrackingFree browser that completely protect users from third-party web tracking by isolating resources in different principals.
- We theoretically and experimentally proved TrackingFree's anti-tracking capability.
- TrackingFree incurs affordable overhead and compatibility cost.