

CSE P564: Computer Security and Privacy

Tracking, Anonymity, and more

Autumn 2024

David Kohlbrenner

dkohlbre@cs

UW Instruction Team: David Kohlbrenner, Yoshi Kohno, Franziska Roesner. Thanks to Dan Boneh, Dieter Gollmann, Dan Halperin, John Manferdelli, John Mitchell, Vitaly Shmatikov, Bennet Yee, and many others for sample slides and materials

Paper discussion

Poisoning Web-Scale Training Datasets is Practical

Discussion Topics

- What makes these ‘vulnerabilities’? (They seem really simple...)
- The paper is called “Practical”
 - Is it *Realistic* too?
- Why no hashes?
- Are there other non-ML related areas where these observations hold?

Lab 1 discussion

Briefly

Sploit 3!

Final project notes

- We'll go over it next week and release it
- Expectation is that each part individually should not take very long
 - Time commitment will depend heavily on how you approach it and how comfortable you are reading/debugging the code.
- All parts release at the same time, can start immediately.

A bit more on authentication

Graphical Passwords

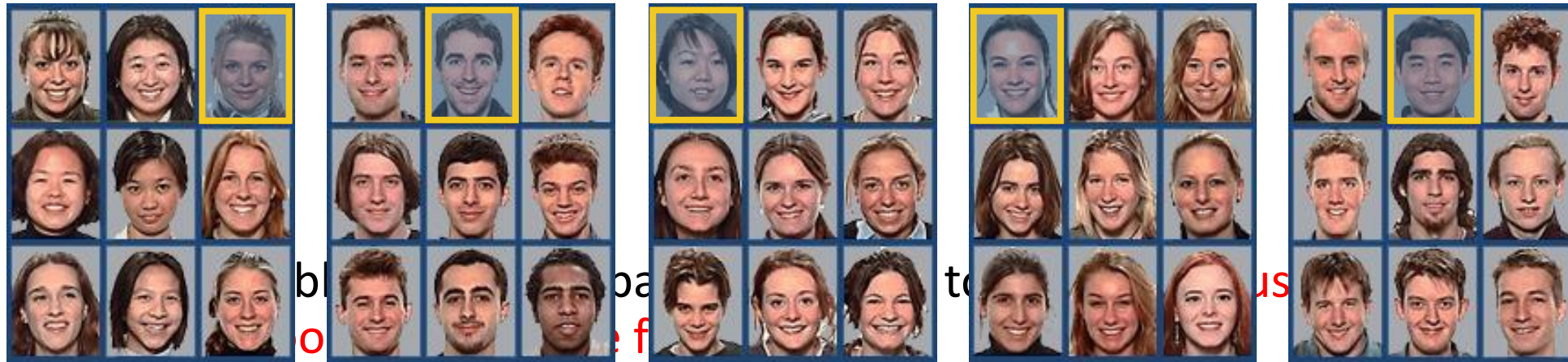
- Another variant: draw on the image (Windows 8)



- Problem: **users choose predictable points/lines**

Graphical Passwords

- Many variants... one example: Passfaces
 - Assumption: easy to recall faces



Unlock Patterns



- Problems:

- Predictable patterns (familiar pattern by now)
- Smear patterns
- Side channels: apps can use accelerometer and gyroscope to extract pattern!

Multi-Factor Authentication

1. Sign in with your **Google Account**
Email:
ex: pat@example.com
Password:
 Stay signed in

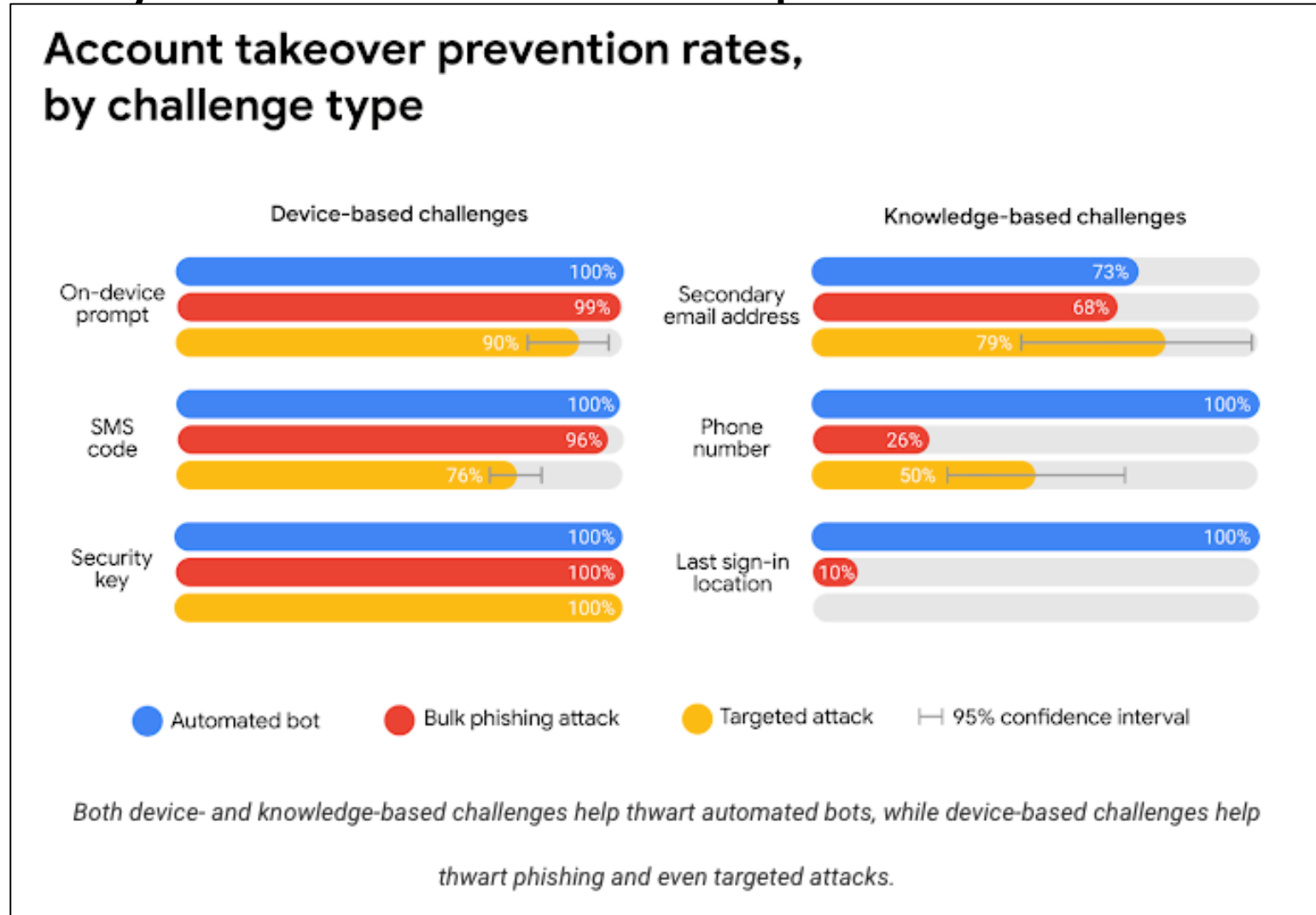
[Can't access your account?](#)

2. **Google accounts**
Enter verification code
To verify your identity on this computer, enter the verification code generated by your mobile application.
Enter code:
 Remember verification for this computer for 30 days.
[Other ways to get a verification code »](#)

Google Authenticator
966286
wileyc@acme.com

Turn on Login Approvals
What is Login Approvals?
Login Approvals is a security feature that requires you to enter a code that we text to your phone when you log in from an unrecognized computer. You can enable this feature in a few simple steps.
If you ever lose access to your phone, you can always return to a previously-recognized computer to regain access to your account.
Note: You'll need to have your mobile phone with you to complete this process.

Secondary Factors Do Help!



Phishing prevention

- The victim *believes they are on the correct site**
 - And then provides credentials
- Whatever our 2nd factor is, must not *be able* to be used incorrectly
 - SMS/email don't work here
 - Hardware tokens/passkeys/etc do

Hardware 2FA tokens (U2F/FIDO)



Passkeys (2024ish)

- An actual, deployed, genuine *password replacement*
 - *Also a 2fa replacement!*
 - *And a username replacement!*
- Basic goals:
 - Store some sort of key on user end-devices
 - Use that key to login to Stuff
 - Don't allow losing the key
 - Somehow make the key moving between devices Easy
 - Sync'd and managed by Apple/Google is the answer

Privacy and web tracking

A topic in flux

- Tracking via cookies
- Tracking via other methods
- Fingerprinting

Ads That Follow You

Advertisers (and others) track your browsing behaviors for the purposes of targeted ads, website analytics, and personalized content.

Third-Party Web Tracking

The image shows a collage of browser windows. On the left is 'The Onion - America's Finest News' with a Zappos ad. On the right is 'CNN.com - Breaking News' with a Zappos ad. In the center is a blue box with the following text:

Browsing profile for user 123:

- cnn.com
- theonion.com
- adult-site.com
- political-site.com

To the right of the list is a sad face emoji (☹️).

These ads allow **criteo.com** to link your visits between sites, **even if you never click on the ads.**

Gradescope

- Do you take any particular precautions about tracking?
 - For web browsing?
 - Phone apps?
 - Phone tracking?
- Why do you take or not take those actions?
 - Any you would like to but don't?

Marketing Technology Landscape

The Martech 5000

Total Solutions 8,000

Advertising & Promotion 922

Content & Experience 1,936

Social & Relationships 1,969

Commerce & Sales 1,314

Data 1,258

Management 601

Access all the data of this landscape & more at martech5000.com

2019

7,040 solutions



2018

6,829 solutions



2017

5,381 solutions



2016

3,874 solutions



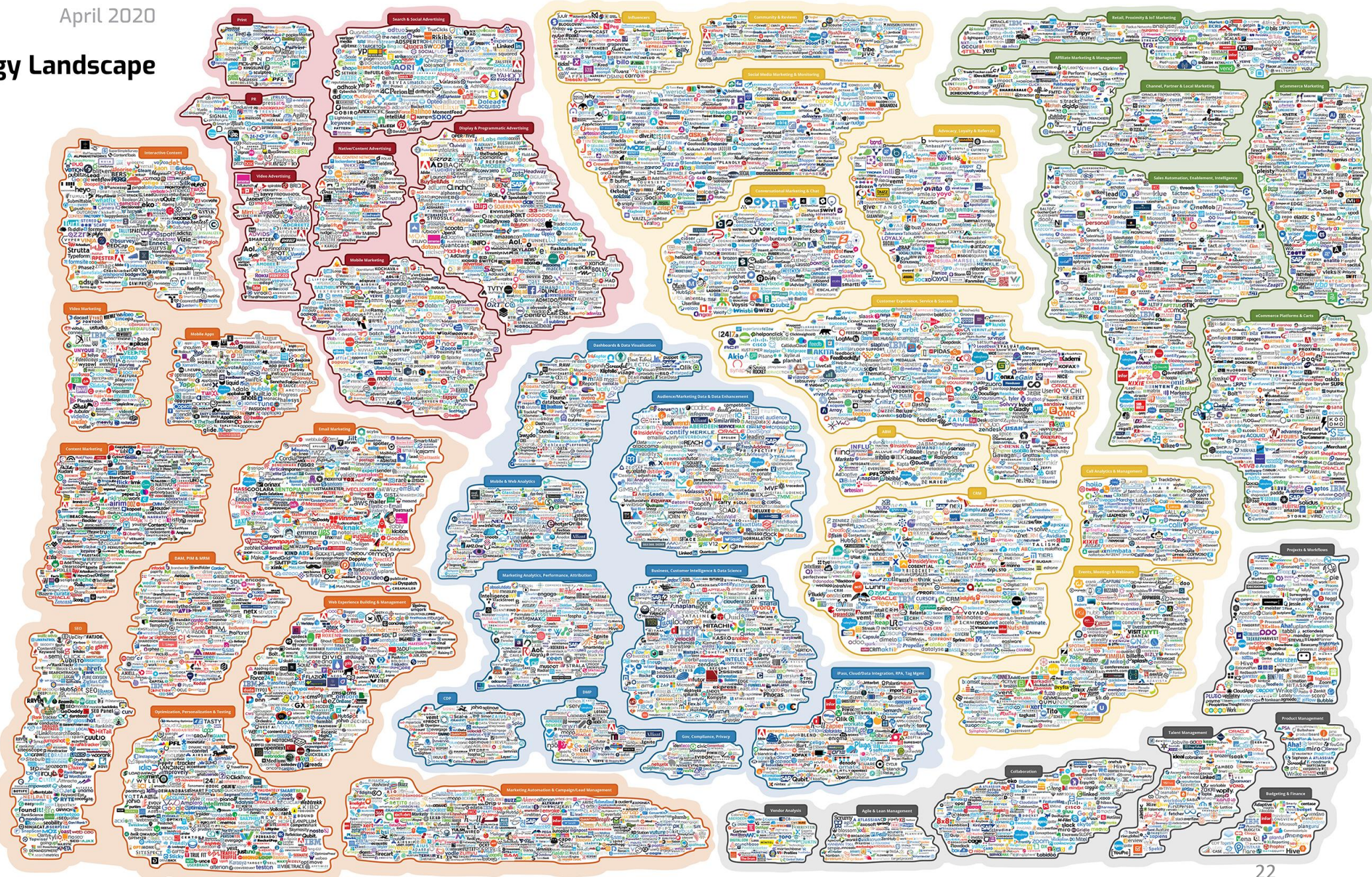
2015

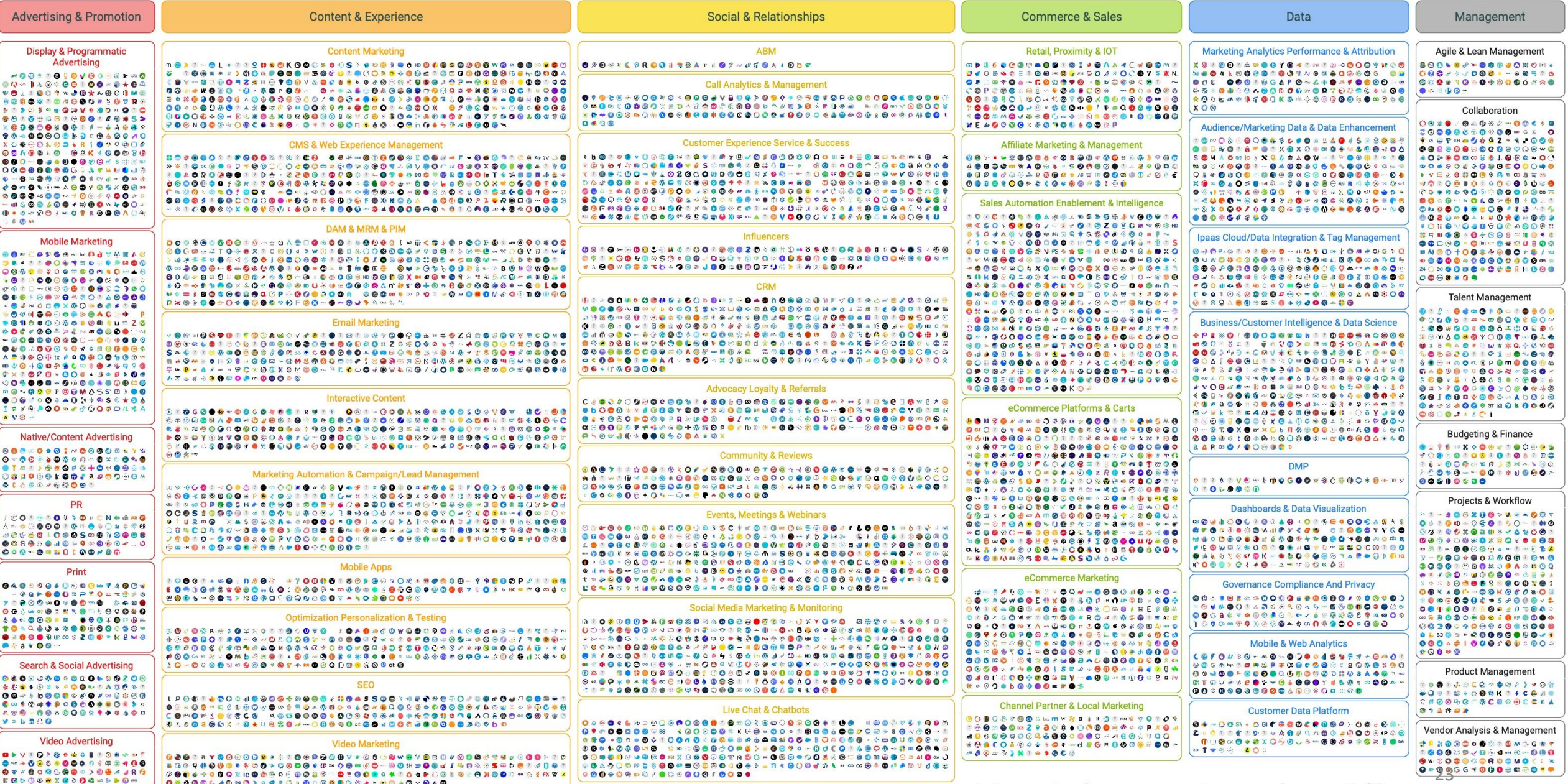
1,876 solutions



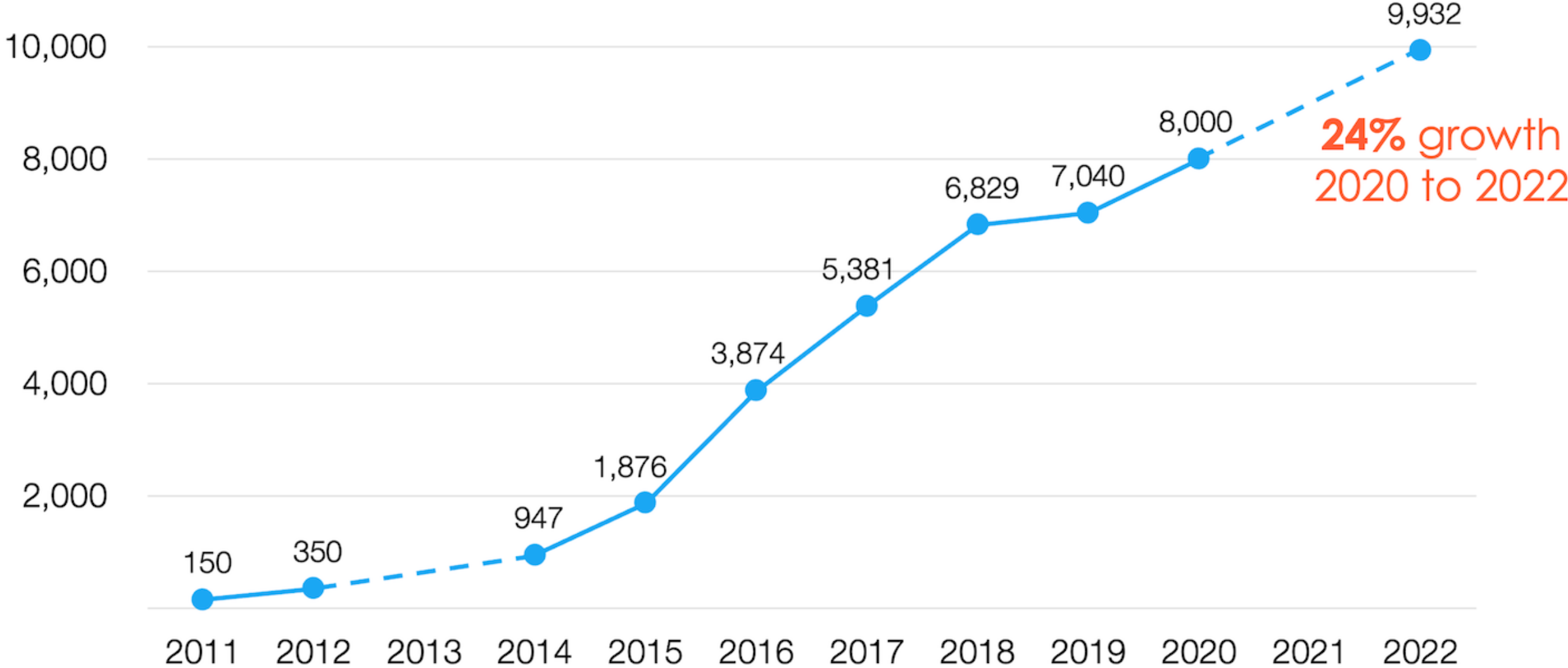
2014

947 solutions





6,521% growth 2011 to 2022



<https://chiefmartec.com/2022/05/marketing-technology-landscape-2022-search-9932-solutions-on-martechmap-com/>

Concerns About Privacy



House, Senate leaders nearing deal on landmark online privacy bill

The expected agreement vaults Congress closer to legislation that lawmakers have sought for decades



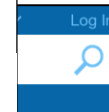
By [Cristiano Lima-Strong](#)

April 5, 2024 at 7:26 p.m. EDT

The file consists
identifies her as

On Friday, two bills were introduced in Washington in support of a Do Not Track mechanism that would give users control over how much of their data was collected by advertisers and other online companies.

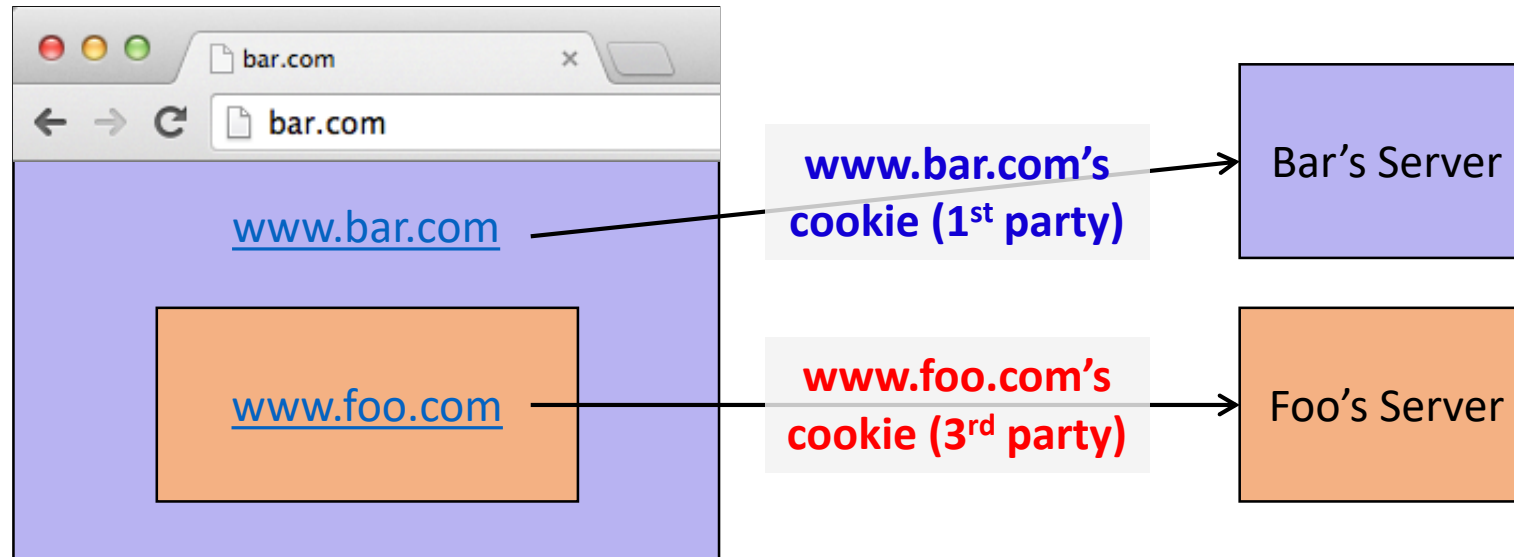
By JENNIFER VALENTINO-DEVRIES,
JEREMY SINGER-VINE and ASHKAN SOLTANI
December 24, 2012



als
ion

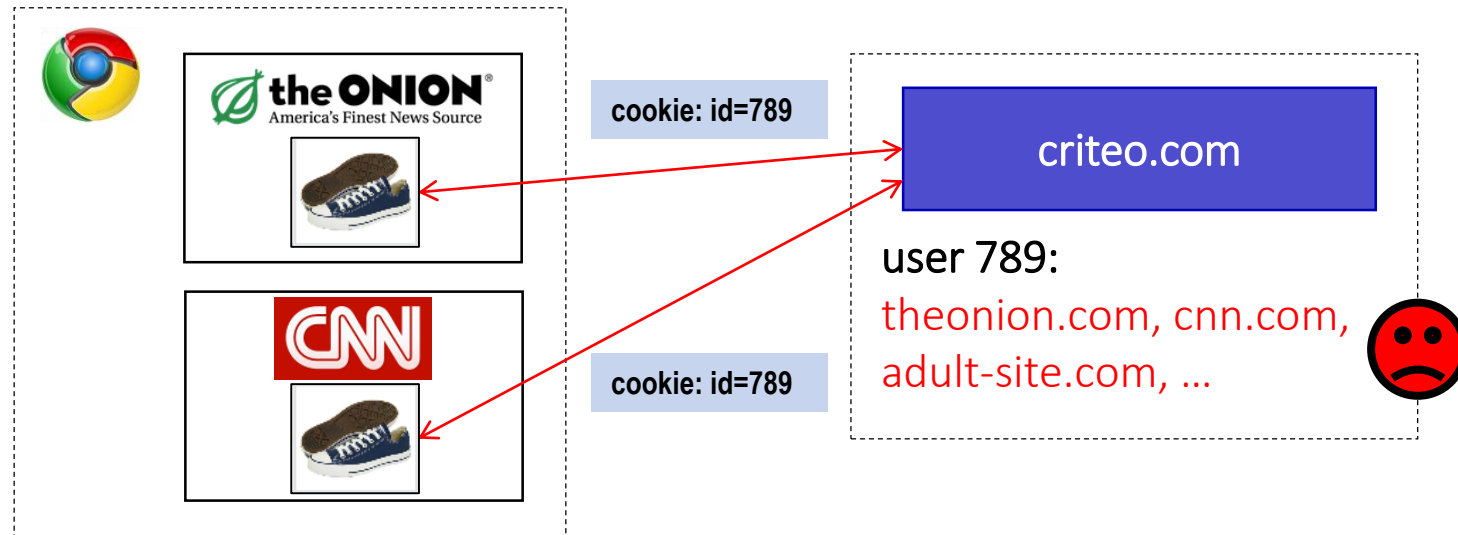
First and Third Parties

- **First-party cookie:** belongs to top-level domain.
- **Third-party cookie:** belongs to domain of embedded content (such as image, iframe).



Anonymous Tracking

Trackers **included in other sites** use **third-party cookies** containing unique **identifiers** to create browsing profiles.



Basic Tracking Mechanisms

- Tracking requires:
 - (1) re-identifying a user.
 - (2) communicating id + visited site back to tracker.

▼ Hypertext Transfer Protocol

```
▶ GET /pixel/p-3aud4J6uA4Z6Y.gif?labels=InvisibleBox&busty=2710 HTTP/1.1\r\n
Host: pixel.quantserve.com\r\n
Connection: keep-alive\r\n
Accept: image/webp,*/*;q=0.8\r\n
User-Agent: Mozilla/5.0 (Macintosh; Intel Mac OS X 10_9_2) AppleWebKit/537.36\r\n
Referer: http://www.theonion.com/\r\n
Accept-Encoding: gzip,deflate,sdch\r\n
Accept-Language: en-US,en;q=0.8\r\n
Cookie: mc=52a65386-f1de1-00ade-0b26e; d=ENkBRgGHD4GYEA35MMIL74MKiyDs1A2MQI1Q
```

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
- “Zombie” cookies that respawn (<http://samy.pl/evercookie>)

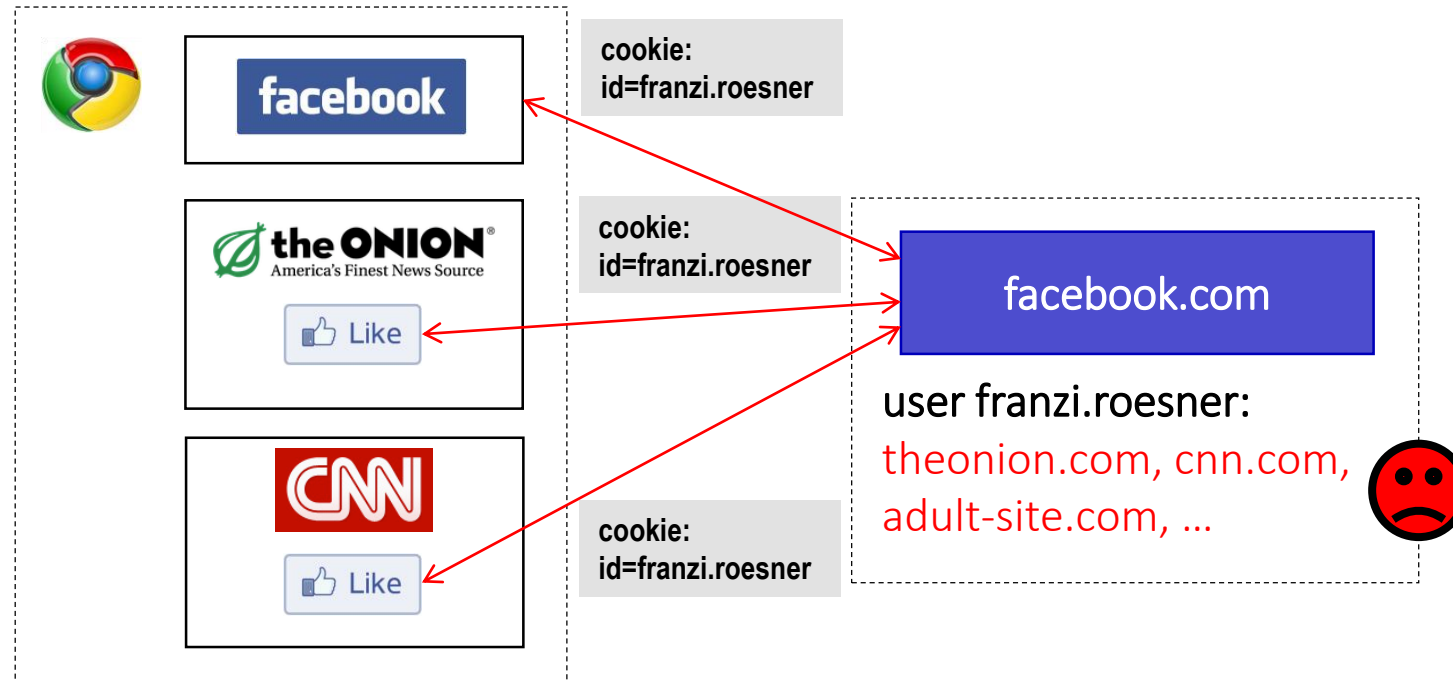
Other Trackers?



“Personal” Trackers



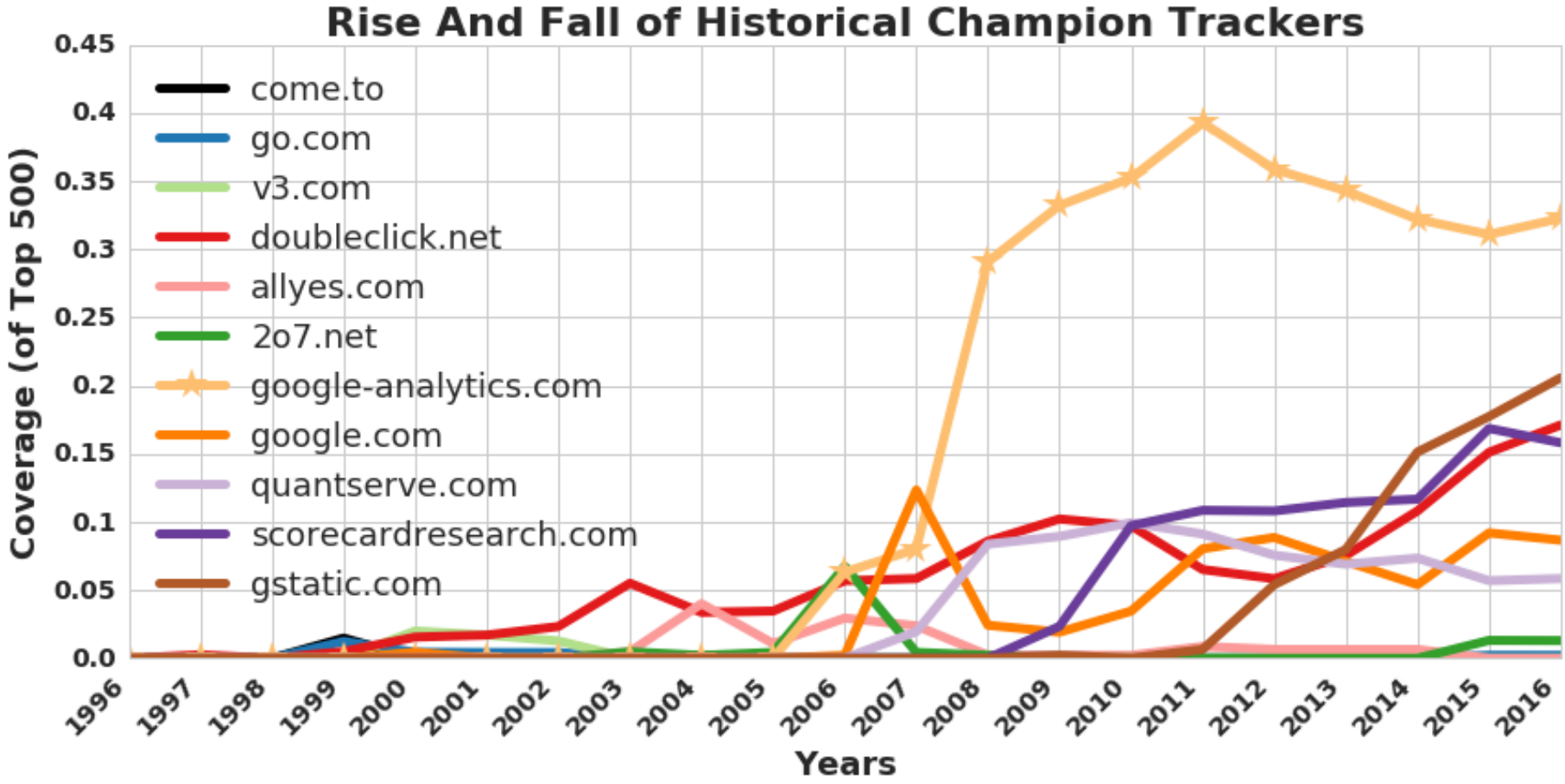
Personal Tracking



- Tracking is **not anonymous** (linked to accounts).
- Users **directly visit tracker's site** → evades some defenses.

1996-2016: More & More Tracking

- More trackers of more types, more per site, **more coverage**



Defenses to Reduce Tracking

- Do Not Track?

Send a 'Do Not Track' request with your browsing traffic

Do Not Track is not a technical defense:
trackers must honor the request.

Defenses to Reduce Tracking

- Do Not Track proposal?
- Private browsing mode?

Private browsing mode doesn't protect against network attackers fully.

You've gone incognito

Now you can browse privately, and other people who use this device won't see your activity. However, downloads and bookmarks will be saved. [Learn more](#)

Chrome won't save the following information:

- Your browsing history
- Cookies and site data
- Information entered in forms

Your activity might still be visible to:

- Websites you visit
- Your employer or school
- Your internet service provider

Defenses to Reduce Tracking

- Do Not Track proposal?
- Private browsing mode?
- Third-party cookie blocking?



3rd party cookies

- Chrome...

“By undermining the business model of many ad-supported websites, blunt approaches to cookies encourage the use of opaque techniques such as fingerprinting (an invasive workaround to replace cookies), which can actually reduce user privacy and control. We believe that we as a community can, and must, do better.”

Aug 2022: Remove 3rd party cookies by 2024

The state of 3rd party cookies

- Safari:
 - Blocks most - <https://webkit.org/blog/10218/full-third-party-cookie-blocking-and-more/>
- Chrome
 - No longer removing? https://privacysandbox.com/intl/en_us/news/privacy-sandbox-update/
- Firefox
 - Specific blocks/etc <https://developer.mozilla.org/en-US/blog/goodbye-third-party-cookies/>
- Others
 - Variety of behaviors, wide variation

Cookie ghostwriting

- No 3rd party cookies allowed ☹️
- Instead, `<script src=https://trackerdomain/cookiewriter.js/>`
- No longer in an iframe... what can they do?

Fingerprinting

- An alternative, popular, approach is *fingerprinting*
 - Website runs some javascript to measure browser/machine behavior
 - Generates an ID from this
 - ID is semi-consistent even across things like incognito mode
- Fingerprinting is unaffected by 3rd party cookie changes!

Anonymity



The New Yorker,
1993

"On the Internet, nobody knows you're a dog."

Privacy on Public Networks

- Internet is designed as a public network
 - Machines on your LAN may see your traffic, network routers see all traffic that passes through them
- Routing information is public
 - IP packet headers identify source and destination
 - Even a passive observer can figure out who is talking to whom
- Encryption does not hide identities
 - Encryption hides payload, but not routing information
 - Even IP-level encryption (tunnel-mode IPSec/ESP) reveals IP addresses of IPSec gateways
- Modern web: Accounts, web tracking, etc. ...

What is Anonymity?

- Anonymity is the state of being not identifiable within a **set of subjects**
 - You cannot be anonymous by yourself!
 - Big difference between anonymity and confidentiality
 - Hide your activities among others' similar activities
- Unlinkability of action and identity
 - For example, sender and email they send are no more related after observing communication than before
- Unobservability (hard to achieve)
 - Observer cannot even tell whether a certain action took place or not

Questions

Q1: Why might we **want** people to have anonymity on the Internet?

Q2: Why might we **not want** people to have anonymity on the Internet?

Applications of Anonymity (I)

- Privacy
 - Hide online transactions, Web browsing, etc. from intrusive governments, marketers and archivists
- Untraceable electronic mail
 - Corporate whistle-blowers
 - Political dissidents
 - Socially sensitive communications (online AA meeting)
 - Confidential business negotiations
- Law enforcement and intelligence
 - Sting operations and honeypots
 - Secret communications on a public network

Applications of Anonymity (II)

- Digital cash
 - Electronic currency with properties of paper money (online purchases unlinkable to buyer's identity)
- Anonymous electronic voting
- Censorship-resistant publishing

Part 1: Anonymity in Datasets

How to release an anonymous dataset?

A Face Is Exposed for AOL Searcher No. 4417749

By MICHAEL BARBARO and TOM ZELLER Jr.; Saul Hansell contributed reporting for this article.

Published: August 9, 2006

Buried in a list of 20 million Web search queries collected by AOL and recently released on the Internet is user No. 4417749. The number was assigned by the company to protect the searcher's anonymity, but it was not much of a shield.

No. 4417749 conducted hundreds of searches over a three-month period on topics ranging from "numb fingers" to "60 single men" to "dog that urinates on everything."


And search by search, click by click, the identity of AOL user No. 4417749 became easier to discern. There are queries for "landscapers in Lilburn, Ga," several people with the last name Arnold and "homes sold in shadow lake subdivision gwinnett county georgia."

It did not take much investigating to follow that data trail to Thelma Arnold, a 62-year-old widow who lives in Lilburn, Ga., frequently researches her friends' medical ailments and loves her three dogs. "Those are my searches," she said, after a reporter read part of the list to her.

 FACEBOOK

 TWITTER

 GOOGLE+

 EMAIL

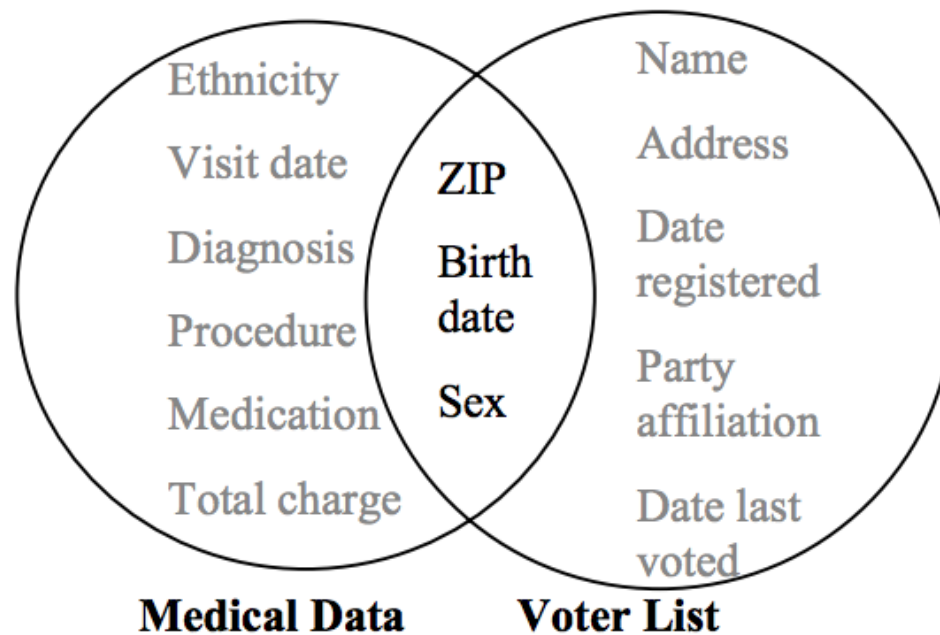
 SHARE

 PRINT

 REPRINTS

How to release an anonymous dataset?

- Possible approach: **remove identifying information from datasets?**



Massachusetts
medical+voter data
[Sweeney 1997]

Figure 1 Linking to re-identify data

k-Anonymity

- Each person contained in the dataset cannot be distinguished from at least $k-1$ others in the data.

Name	Age	Gender	State of domicile	Religion	Disease
Ramsha	29	Female	Tamil Nadu	Hindu	Cancer
Yadu	24	Female	Kerala	Hindu	Viral infection
Salima	28	Female	Tamil Nadu	Muslim	TB
Kaker	27	Male	Karnataka	Parsi	No illness
Joan	24	Female	Kerala	Christian	Heart-related
Bahuksana	23	Male	Karnataka	Buddhist	TB
Rambha	19	Male	Kerala	Hindu	Cancer
Kishor	29	Male	Karnataka	Hindu	Heart-related
John	17	Male	Kerala	Christian	Heart-related
John	19	Male	Kerala	Christian	Viral infection

k-Anonymity

- Each person contained in the dataset cannot be distinguished from at least $k-1$ others in the data.

Name	Age	Gender	State of domicile	Religion	Disease
*	$20 < \text{Age} \leq 30$	Female	Tamil Nadu	*	Cancer
*	$20 < \text{Age} \leq 30$	Female	Kerala	*	Viral infection
*	$20 < \text{Age} \leq 30$	Female	Tamil Nadu	*	TB
*	$20 < \text{Age} \leq 30$	Male	Karnataka	*	No illness
*	$20 < \text{Age} \leq 30$	Female	Kerala	*	Heart-related
*	$20 < \text{Age} \leq 30$	Male			
*	$\text{Age} \leq 20$	Male			
*	$20 < \text{Age} \leq 30$	Male			
*	$\text{Age} \leq 20$	Male			
*	$\text{Age} \leq 20$	Male	Kerala	*	Viral infection

Doesn't work for high-dimensional datasets (which tend to be sparse)

Robust De-anonymization of Large Sparse Datasets

Arvind Narayanan and Vitaly Shmatikov

The University of Texas at Austin

Netflix Challenge:

- Netflix released a (non-uniform) random sample of user's movie ratings
- Challenge was to build a better recommendation system
- Data was 'anonymous'
 - ID # only
 - Random selection of a given user's ratings
 - "noise" added (appears that there was no noise)

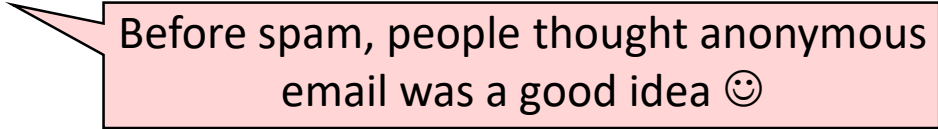
Result: No real anonymity

- Cross-correlate with IMBD ratings
- A handful (6 or fewer) ratings of non-top 500 movies is enough!

Part 2: Anonymity in Communication

Chaum's Mix

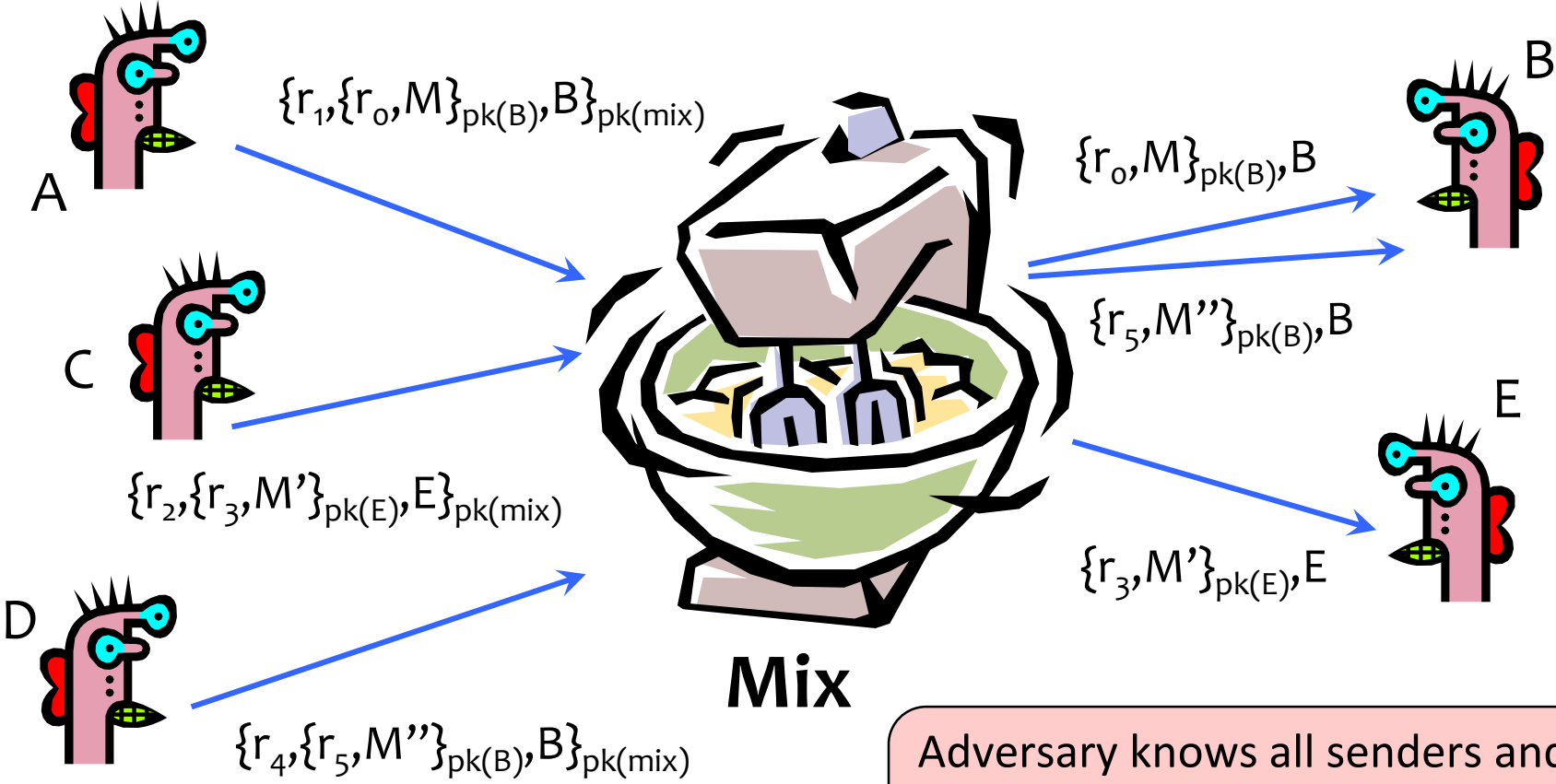
- Early proposal for anonymous email
 - David Chaum. “Untraceable electronic mail, return addresses, and digital pseudonyms”. Communications of the ACM, February 1981.



Before spam, people thought anonymous email was a good idea 😊

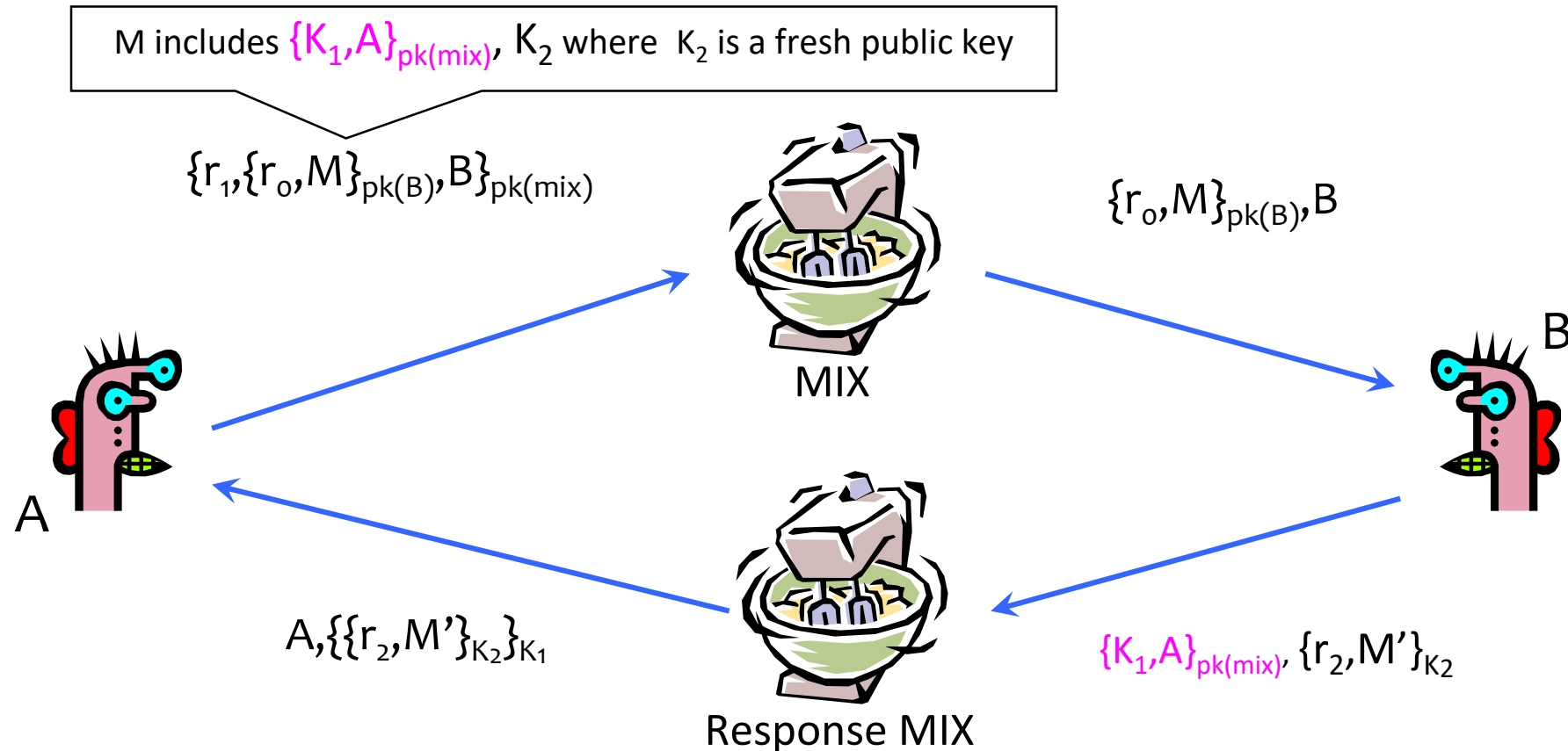
- Modern anonymity systems use Mix as the basic building block

Basic Mix Design



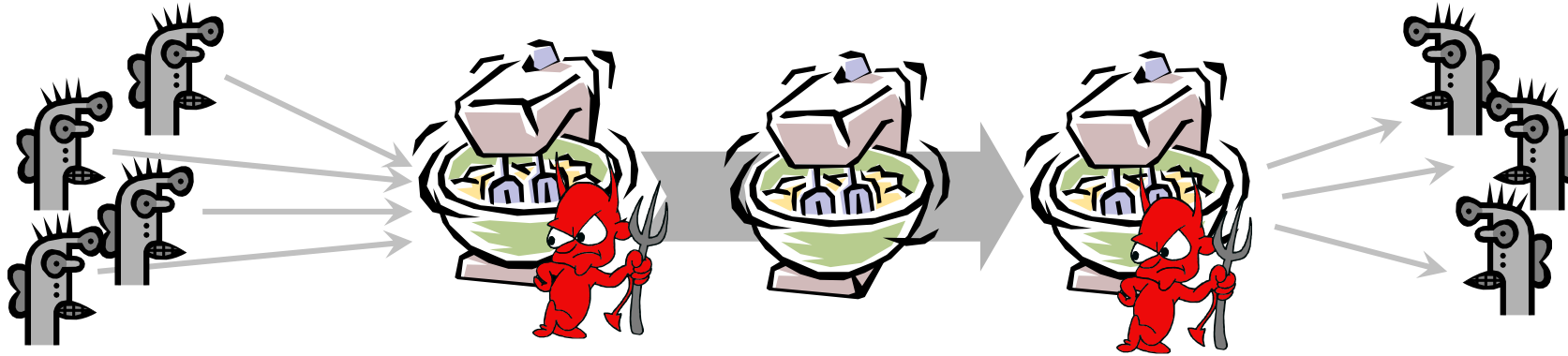
Adversary knows all senders and all receivers, but cannot link a sent message with a received message

Anonymous Return Addresses



Secrecy without authentication
(good for an online confession service 😊)

Mix Cascades and Mixnets



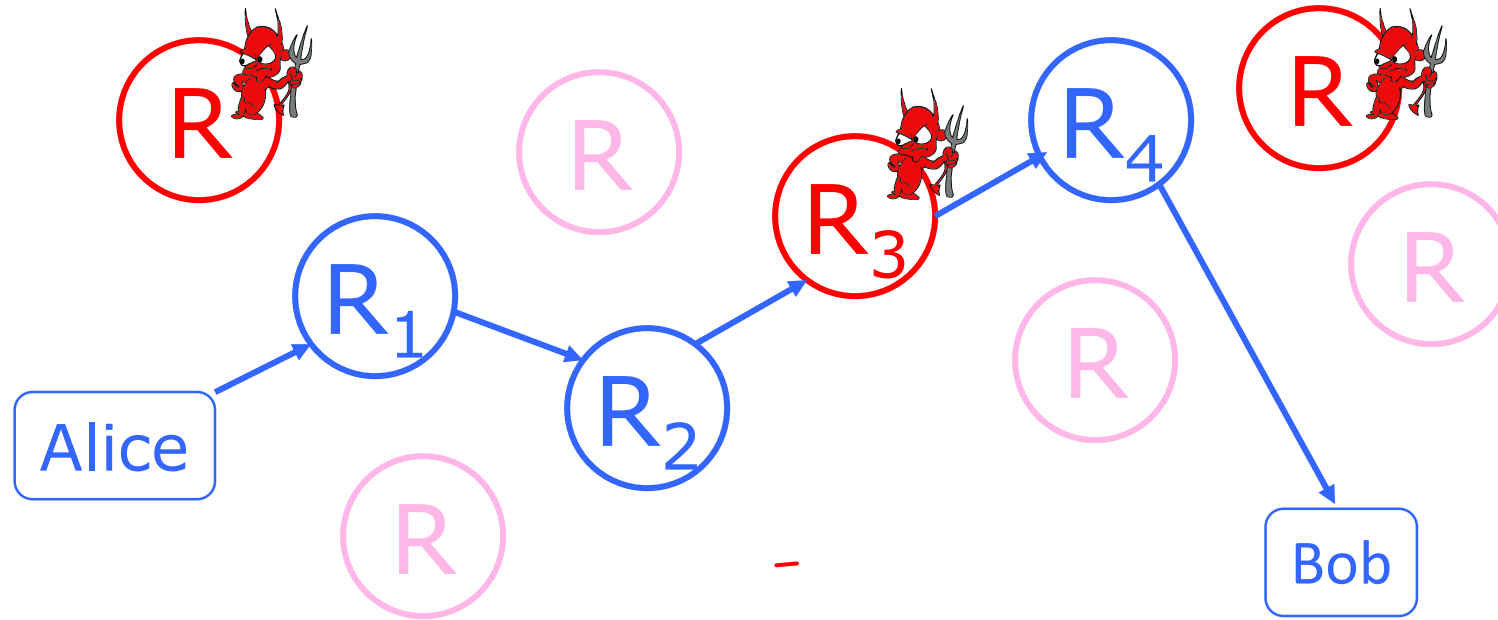
- Messages are sent through a **sequence of mixes**
 - Can also form an arbitrary network of mixes (“mixnet”)
- Some of the mixes may be controlled by attacker, but even a single good mix ensures anonymity
- Pad and buffer traffic to foil **correlation attacks**

Disadvantages of Basic Mixnets

- Public-key encryption and decryption at each mix are **computationally expensive**
- Basic mixnets have **high latency**
 - OK for email, not OK for anonymous Web browsing
- Challenge: **low-latency anonymity network**

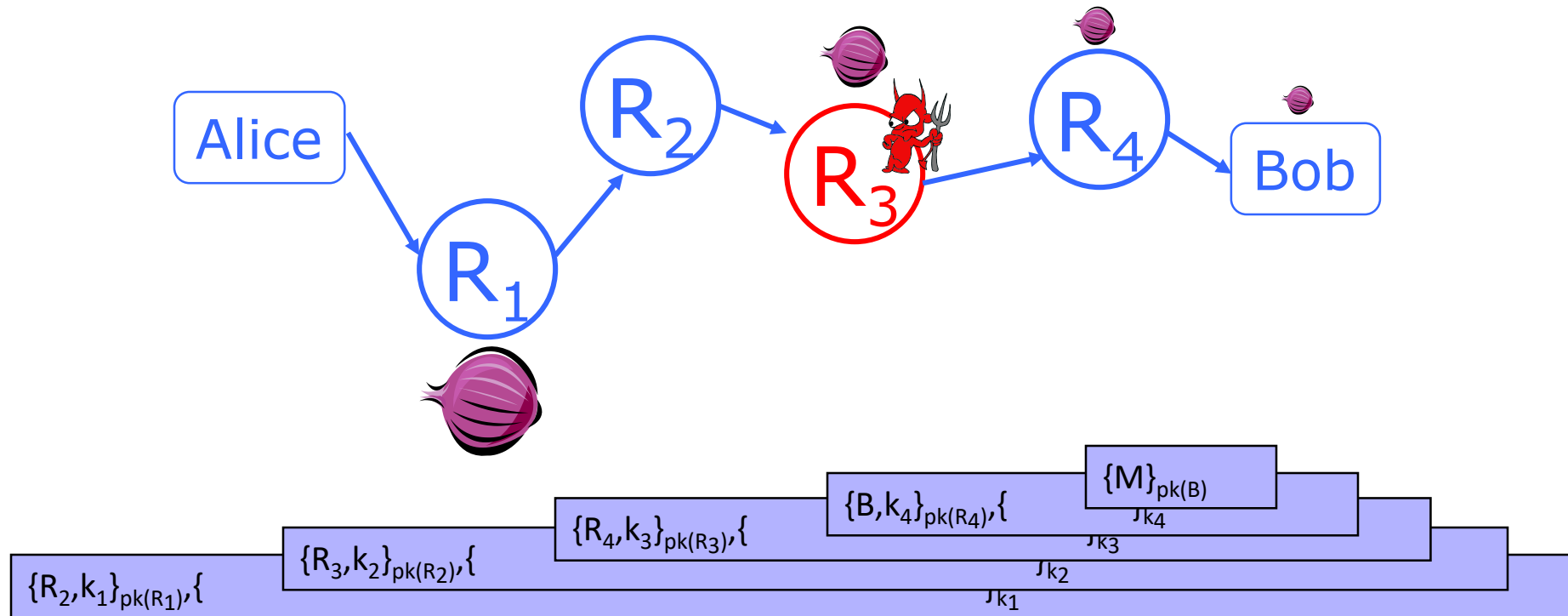
Another Idea: Randomized Routing

e.g., Onion Routing



- Sender chooses a random sequence of routers
 - Some routers are honest, some controlled by attacker
 - Sender controls the length of the path

Onion Routing



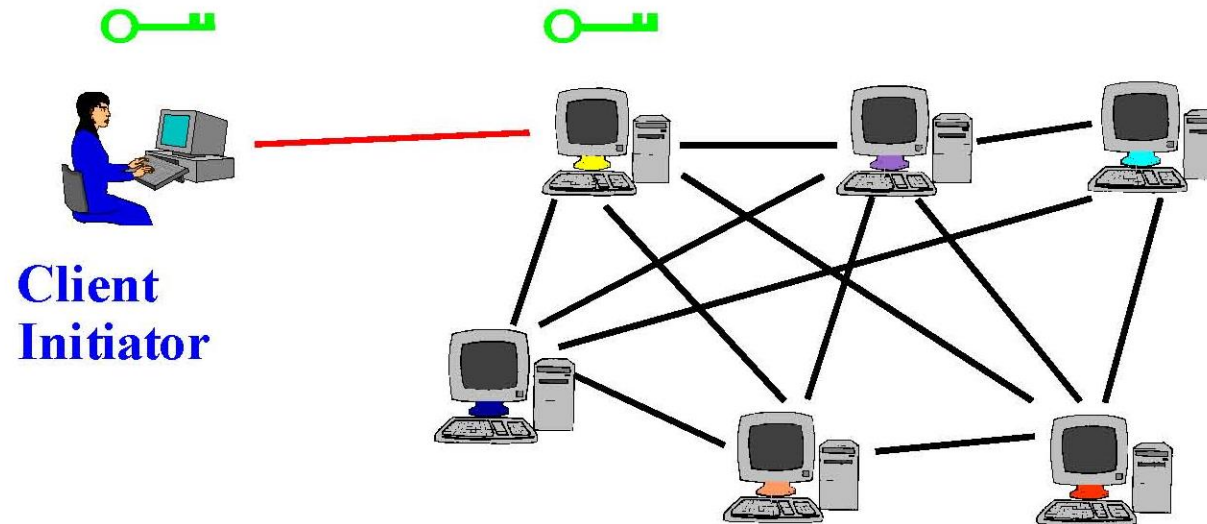
- Routing info for each link encrypted with router's public key
- Each router learns only the identity of the next router

Tor

- Second-generation onion routing network
 - <http://tor.eff.org>
 - Developed by Roger Dingledine, Nick Mathewson and Paul Syverson
 - Specifically designed for **low-latency** anonymous Internet communications
- Running since October 2003
- “Easy-to-use” client proxy
 - Freely available, can use it for anonymous browsing

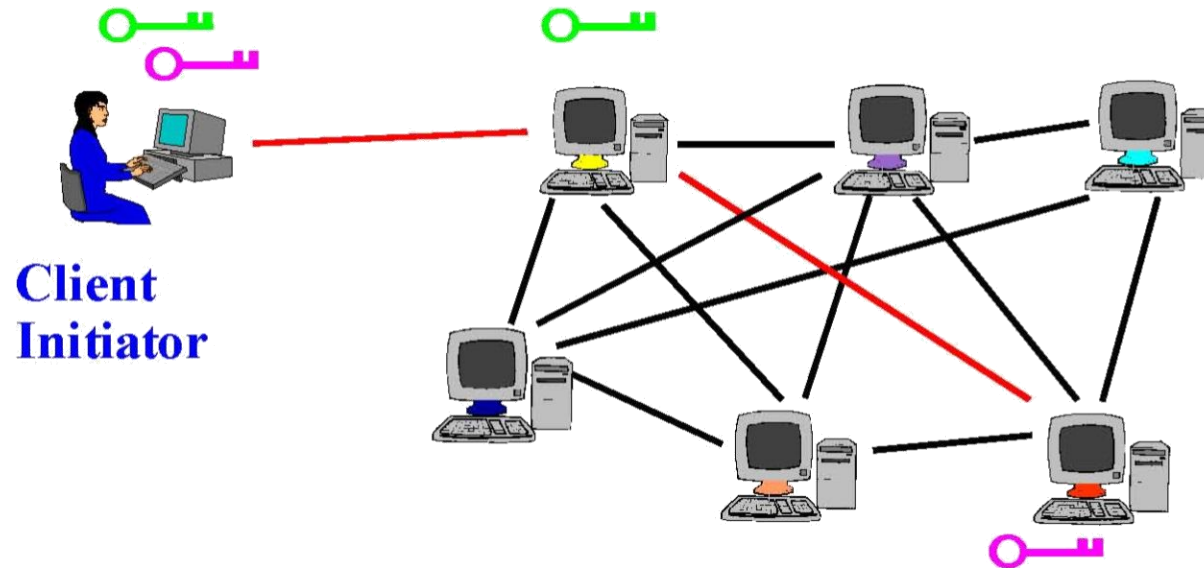
Tor Circuit Setup (1)

- Client proxy establishes a symmetric session key and circuit with Onion Router #1



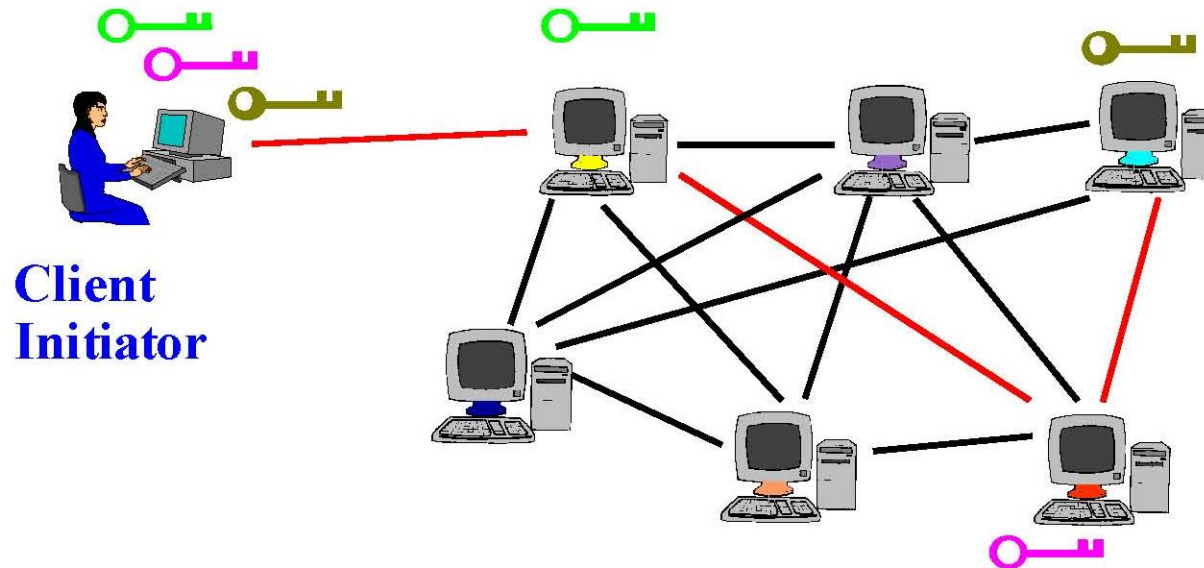
Tor Circuit Setup (2)

- Client proxy extends the circuit by establishing a symmetric session key with Onion Router #2
 - Tunnel through Onion Router #1



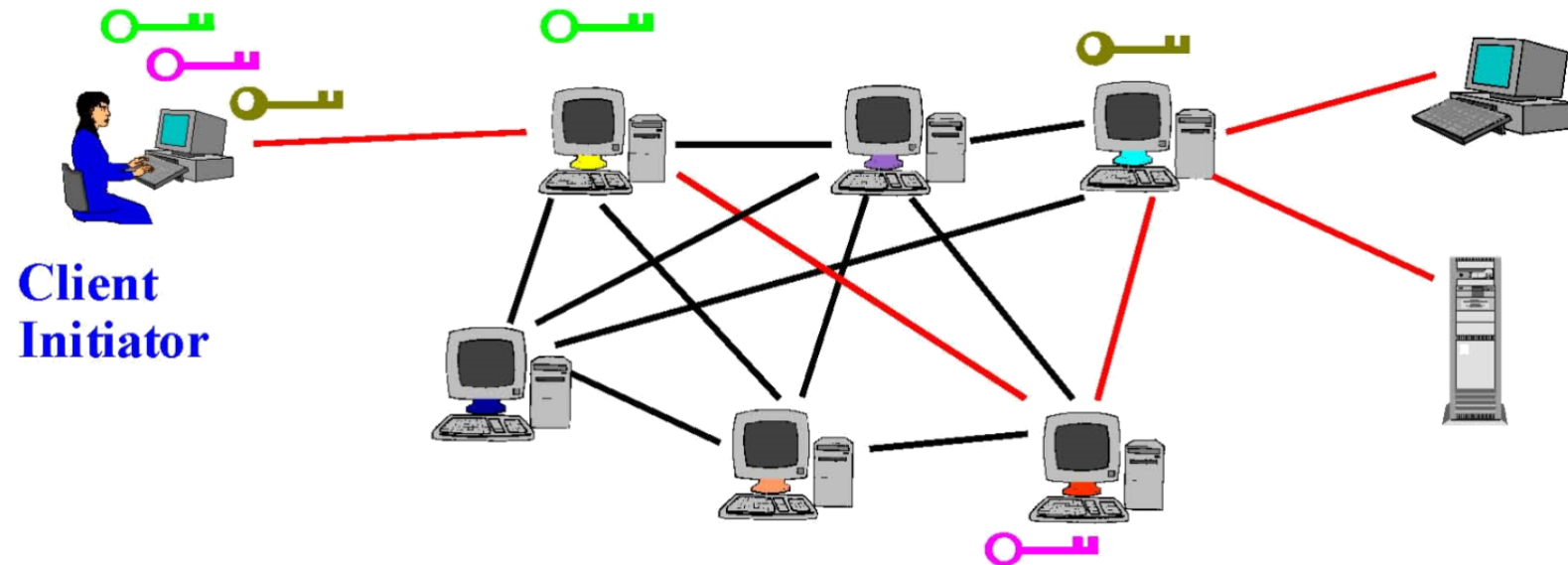
Tor Circuit Setup (3)

- Client proxy extends the circuit by establishing a symmetric session key with Onion Router #3
 - Tunnel through Onion Routers #1 and #2



Using a Tor Circuit

- Client applications connect and communicate over the established Tor circuit.



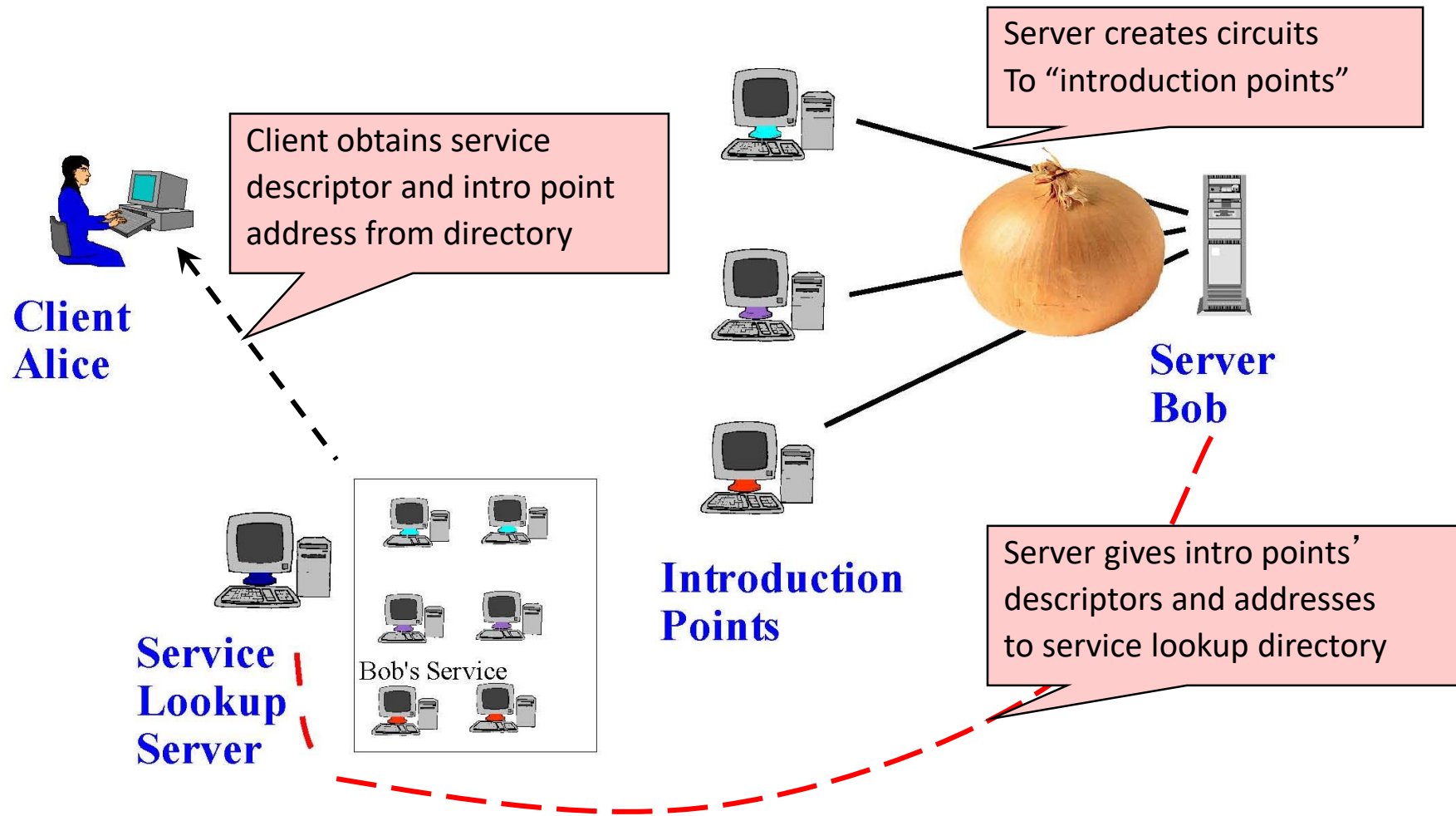
How do you know who to talk to?

- Directory servers
 - Maintain lists of active onion routers, their locations, current public keys, etc.
 - Control how new routers join the network
 - “Sybil attack”: attacker creates a large number of routers
 - Directory servers’ keys ship with Tor code

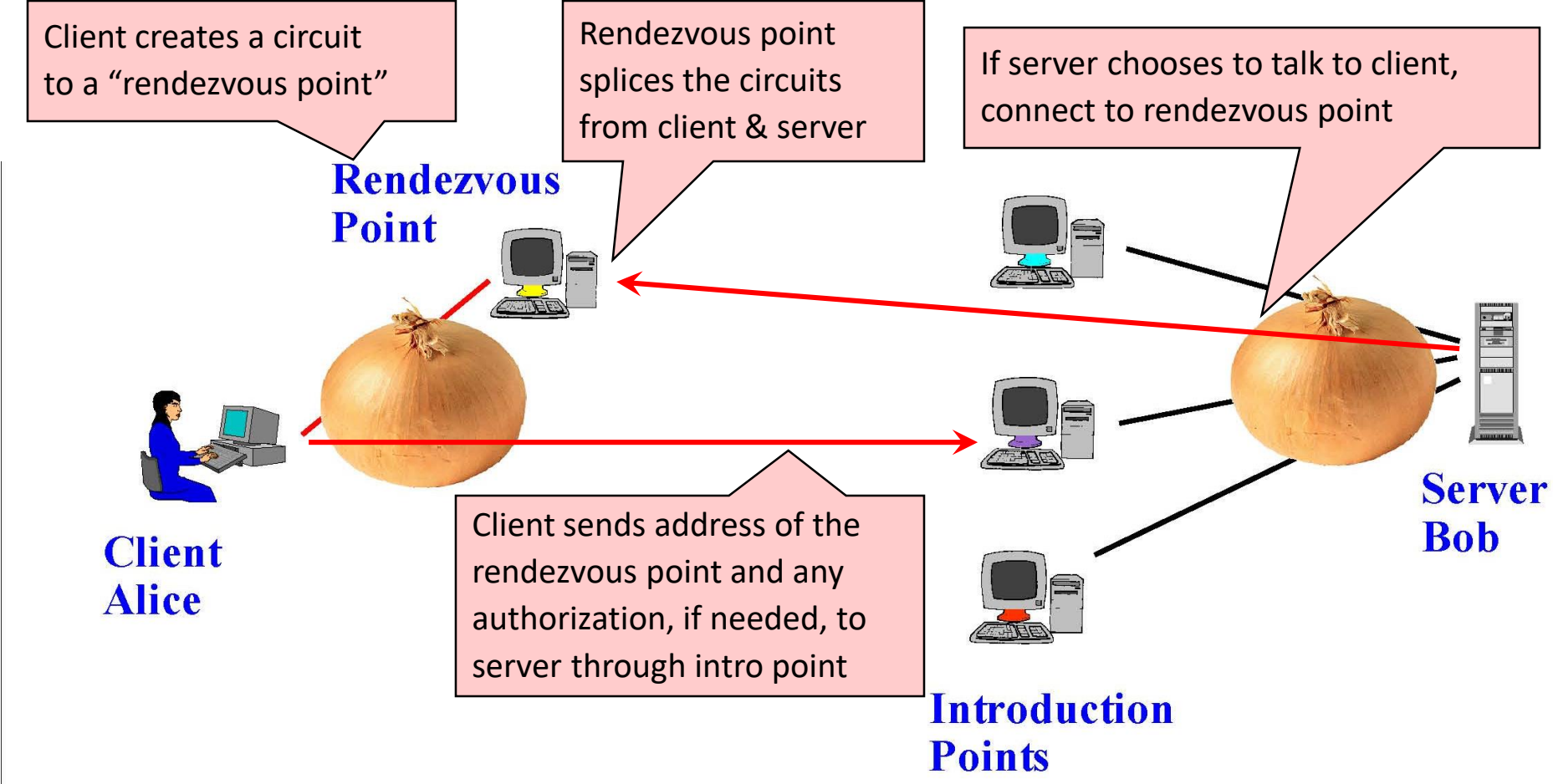
Location Hidden Service

- **Goal:** deploy a server on the Internet that anyone can connect to **without knowing where it is or who runs it**
- Accessible from anywhere
- Resistant to censorship
- Can survive a full-blown DoS attack
- Resistant to physical attack
 - Can't find the physical server!

Creating a Location Hidden Server



Using a Location Hidden Server

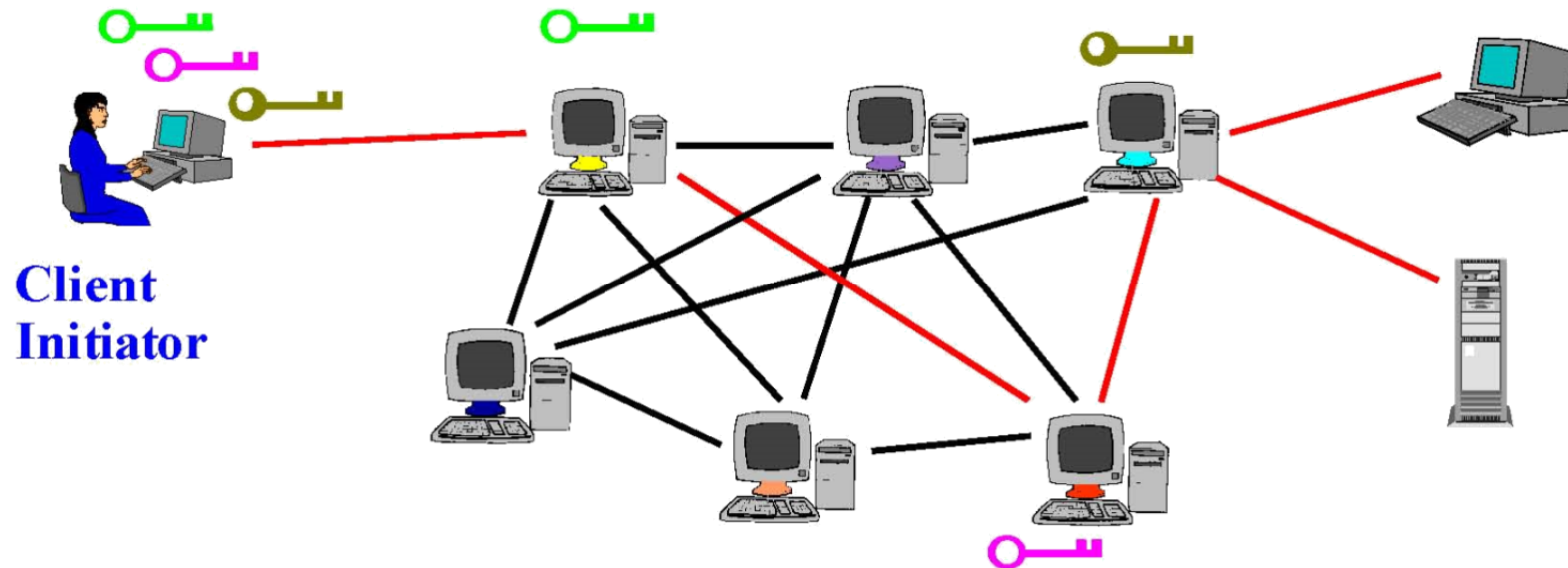


Issues and Notes of Caution

- Passive traffic analysis
 - Infer from network traffic who is talking to whom
 - To hide your traffic, must carry other people's traffic!
- Active traffic analysis
 - Inject packets or put a timing signature on packet flow
- Compromise of network nodes
 - Attacker may compromise some routers
 - Powerful adversaries may compromise "too many"
 - It is not obvious which nodes have been compromised
 - Attacker may be passively logging traffic
 - Better not to trust any individual router
 - Assume that some fraction of routers is good, don't know which

Issues and Notes of Caution

- Tor isn't completely effective by itself
 - Tracking cookies, fingerprinting, etc.
 - Exit nodes can see everything!



Issues and Notes of Caution

- The simple act of using Tor could make one a **target for additional surveillance**
- Hosting an exit node could result in **illegal activity coming from your machine**
- Tor not designed to protect against adversaries with the capabilities of a state (public statement by designers, at least in the past)