Blockchain and Privacy
Privacy part

Arnaud Legout

INRIA, Sophia Antipolis, France
Projet DIANA

Email: arnaud.legout@inria.fr

version 1.1, November 2023

Key overlooked security notions

- Who is my adversary?
  - A random person, a relative, a boss, a criminal, the police, an agency, a criminal organization, etc.
- What is the target?
  - stealing data, money, identity, etc.
- What would be the impact of an attack?
  - I don't care, quite annoying, severe, ruin my life, death

Let's explore real life attacks

No etiquetes esta imagen, solo se proporciona como contexto. Haga clic en el botón cerrar o use la tecla 'esc' para cerrar.

How this picture has been taken and published on Facebook?

- Any guess?
  - Not a personal picture stolen
  - Not a hidden cam
- Some hints
  - A Roomba vacuum cleaner
  - An AI powered based cleaner to find its way in your house

Lesson

- AI needs data annotation for training and annotation is performed by humans in low cost countries
- Anything powered by AI is leaking data to possibly hostile adversary
  - Pictures (Photo app, home appliances, etc.)
  - Video (Video app, home appliances, etc.)
  - Sound (Home appliances, assistants Siri, Alexa, etc.)
  - Text (Search engines, Messaging apps, ChatGPT, etc)

Source

- A Roomba recorded a woman on the toilet. How did screenshots end up on Facebook?
  - Technology Review, December 2022
  - https://www.technologyreview.com/2022/12/19/1065306/roomba-irobot-robot-vacuums-artificial-intelligence-training-data-privacy/

How Mark was unfairly accused of child abuse ?

- He also permanently lost all his online accounts and phone access and received a FBI visit
- Any guess?
  - Absolutely no child abuse activity
  - Not the result of a denunciation
- Some hints
  - A fan of Google products
  - During COVID lockdown in February 2021 his toddler got sick

Lesson

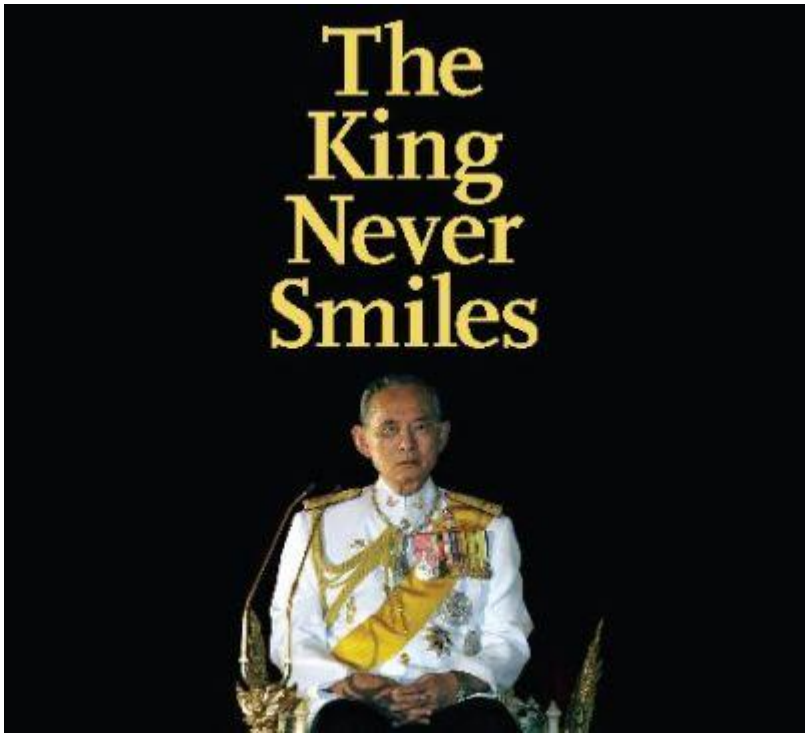
- AI again filter (here this is good), then a human cross check (good again)
  - But an issue when it is on your private data you don't know it is scrutinized
    - You might argue it is good, it is a model of society you must understand before giving a judgment
- A tech company is not a democracy
  - They have their own rules, you basically have no appeal

Source

- A Dad Took Photos of His Naked Toddler for the Doctor. Google Flagged Him as a Criminal
  - New York Time, August 2022
  - https://www.nytimes.com/2022/08/21/technology/google-surveillance-toddler-photo.html

Why Joe has been sentenced to jail ?

- Any guess?
  - American citizen
  - He has been arrested when he landed to Thailand
    - No illegal activity performed in the United states and in Thailand
- Some hints
  - 195 countries world wide, each with different laws
  - How can you be sure you did nothing illegal from the point of view of all of these countries



« Je ne suis pas Thaïlandais, je suis Américain. Je suis né en Thaïlande, mais j'ai un passeport américain. Il y a en Thaïlande des lois contre la liberté d'expression, on n'a pas de telles lois au États-Unis. »

Lesson

- Whatever you publish, it will
  - Never disappear
  - Be available worldwide instantly

- Be aware that most likely you will upset someone somewhere
  - Are you ready to face the consequences?

Source

- Thai judge gives American two years' jail for 'insulting monarchy'
  - The Guardian, December 2021
  - https://www.theguardian.com/world/2011/dec/08/thai-american-jail-insulting-monarchy
  - https://prachatai.com/english/node/2937
  - https://en.wikipedia.org/wiki/The_King_Never_Smiles

Outline

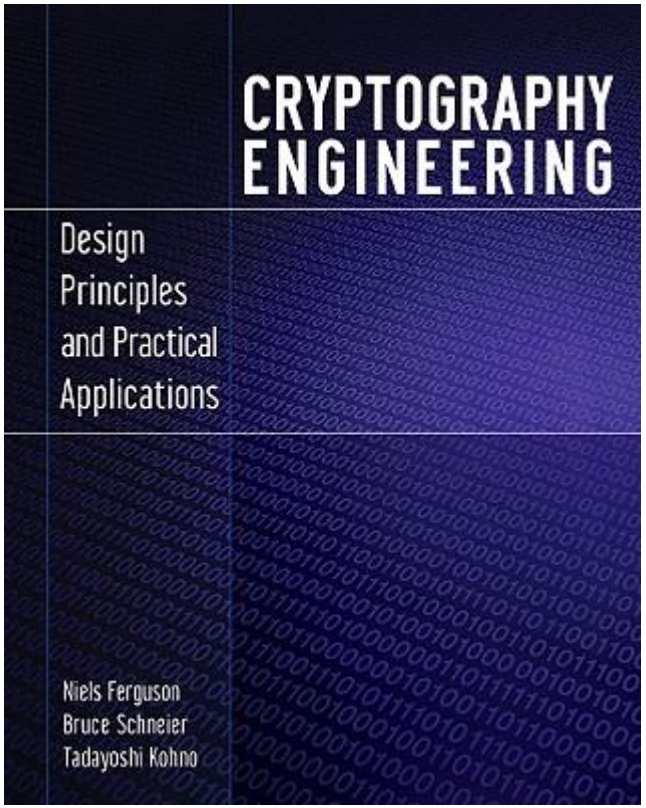
- Privacy Foundations
  - Shared Secret
  - Chaum-net
- Privacy Attacks

Foundations

- You MUST read [23] [24]
  - Small papers (2 and 4 pages only)
  - Best examples on how to use mathematics the simplest and best way
- Current work
  - Tor, Freenet, Publius, OceanStore, etc.

Foundations

- Highly recommended read [51]
  - Easy to read
  - Book to understand the difference between theory and practice



How to Share a Secret [23]

- Problem
  - 11 scientists are working on a super secret project
    - They don't trust each other
    - The project is in a digital safe
  - To open the digital safe, at least 6 out of the 11 scientists must be present
- Apply to any problem with a group of suspicious individuals with conflicting interests that must cooperate

(k,n) Threshold Scheme

- Formal definition of the previous example: (k,n) threshold scheme
- Let D be some secret data
  - Lets divide D into n pieces D_1,…, D_n such that
    - Knowledge on any k or more D_i pieces makes D easily computable
    - Knowledge of any k-1 or fewer D_i pieces leaves D completely undetermined
- Very useful when D is a decryption key

Trivial Solution

- Let's take a simple problem
  - 11 scientists are working on a secret project
  - The project is encrypted
  - To decrypt the project at least 6 scientists have to be present
- Trivial solution
  - Encrypt the content N times
  - Each encryption key is split into 6 fragments

Trivial Solution

- How many times the content must be encrypted?
  - Any set of 6 scientists is associated to a decryption key
    - Each scientist of a given set will have a fragment of a sixth of the key
  - The number of keys is the combination of 6 scientists out of 11
    - $\binom{11}{6} = \frac{11!}{6!(11-6)!} = 462$

Trivial Solution

- How many fragment each scientist must carry
  - Any set of 6 scientists must be able to reconstruct a key, that is, to decrypt the content
  - Each scientist needs a different fragment to reconstruct each key with 5 other scientists chosen among 10 (that is 11 minus himself)
    - $\binom{10}{5} = \frac{10!}{5!(10-5)!} = 252$

Trivial Solution

- Trivial solution is impractical
  - For only 11 scientists and 6 out of 11 able to decrypt the content
    - 462 keys, i.e., 462 encryptions of the content
    - 252 key fragments per scientist

Shamir's (k,n) Threshold Scheme

- Basic idea
  - A polynomial of degree k-1 is uniquely defined by k points
    - 2 points for a line, 3 points for a parabola, 4 points for a cubic curve, etc.
  - With k-1 points only there is an infinity of k polynomial that can cross those points
    - So you need at least k points to find the polynomial equation g(x) using Lagrange interpolation
    - The secret is g(0)

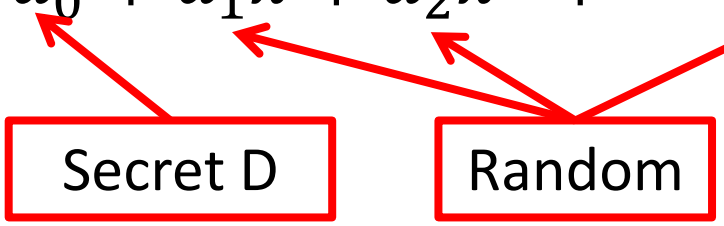
Implementation on Galois Field

- All the arithmetic is modular arithmetic on Galois field (finite field)
  - Mandatory to provide perfect secrecy, that is k-1 pieces do not give any information on the secret under a (k,n) threshold scheme
  - The set of integers modulo a prime number p forms a field in which interpolation is possible
  - Addition subtraction and multiplication is the same as for integers, but not divisions

Implementation (k,n) Threshold Scheme

- Create the n fragments
  - Let D be your secret (D is an integer without loss of generality)
  - Choose a prime p > max(D,n)
  - g(x) is a random polynomial of degree k-1 so that $g(x)=\sum_{i=0}^{k-1} a_i x^i$
    - a_0=D, and a_i, i∈{1,…,k-1} are chosen with a uniform distribution on [0,p[

$$g(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_{k-1} x^{k-1}$$



Implementation (k,n) Threshold Scheme

- Create the n fragments
  - Compute
    - $D_1=g(1) \bmod p$, $D_2=g(2) \bmod p$, …, $D_n=g(n) \bmod p$
  - Distribute the tuples (i,D_i)

Implementation (k,n) Threshold Scheme

- Retrieve D based on k fragments (x_i, D_i)
  - Use Lagrange polynomial interpolation to reconstruct the polynomial
    - g(0) = D is the secret

$$g(0) = \sum_{i=1}^{k} D_i \left(\prod_{j=1, j \neq i}^{k} \frac{-x_j}{x_i - x_j} \right)$$

Example for (k=3,n=5)

- The secret is D=148
- Let's take
  - p=997 (prime), a_1=59 (random), a_2=340(random)
  - $g(x)=148 + 59x + 340x^2$
- We compute 5 fragments
  - $D_1 = g(1) \bmod 997 = 547$
  - $D_2 = g(2) \bmod 997 = 1626 \bmod 997 = 629$
  - $D_3 = g(3) \bmod 997 = 3385 \bmod 997 = 394$
  - $D_4 = g(4) \bmod 997 = 5824 \bmod 997 = 839$
  - $D_5 = g(5) \bmod 997 = 8943 \bmod 997 = 967$

Example for (k=3,n=5)

- We give to each user a fragment among
  - (1,547), (2,629), (3,394), (4,839), (5,967)
- Assume users with fragments 1,3,4 want to reconstruct the secret
  - They compute g(0)

$$g(0) = 547\left(\frac{-3}{1-3}\frac{-4}{1-4}\right) + 394\left(\frac{-1}{3-1}\frac{-4}{3-4}\right) + 839\left(\frac{-1}{4-1}\frac{-3}{4-3}\right)$$

$$g(0) = 547 * 2 - 394 * 2 + 839 = 1145$$

$$g(0)\ mod\ 997 = 148$$

Properties of Shamir's Scheme

- The size of each fragment does not exceed the size of the secret (if p is the same size order as the secret)
- New fragments can be generated at any time without affecting existing ones
- All fragments can be changed without changing the secret by generating a new polynomial

Properties of Shamir's Scheme

- Possibility of hierarchical schemes by giving a different number of fragments depending on roles (e.g., president 3 fragments, executives 1 fragment)
- No unproven assumptions (unlike cryptographic or hash protocols)

Outline

- Privacy Foundations
  - Shared Secret
  - Chaum-net
- Privacy Attacks

Chaum-net [24]

- Chaum-net = mix-net
  - Basis for onion routing
- Problem
  - Alice wants to send a message M to Bob
    - Assume an unsecure communication network
    - Nobody knows who is the sender (even Bob)
    - Nobody knows who is the receiver (except Alice)
    - Nobody, except Bob, is able to get M

Notations

- Assume a public key cryptosystem (e.g., RSA)
  - M is a message
  - K is a public key, K^{-1} is the corresponding private key
  - $K(K^{-1}(M)) = K^{-1}(K(M)) = M$

Sealed Message

- A message M is sealed with public key K if only the holder of K^{-1} can retrieve M
- K(M) is not sealed because anyone can verify the guess K(N) = K(M)
  - Keep in mind that M might be easy to guess due to its semantic
  - The attacker might know M and just want to find who is sending it

Sealed Message

- Solution
  - Create a large random string R (e.g., 256 bits large)
  - Append R to the message M: R,M
  - The sealed message is K(R,M)
    - As R is a large random string, not practical to guess R,M
  - Once Bob get K(R,M)
    - Compute K^{-1} (K(R,M)) = R,M
    - Remove R (easy if R is fixed length)

Mix

- A mix is a machine
  - Might be a dedicated machine, a router, an end-user in an overlay
- Mix purpose
  - Hide correspondences between incoming and outgoing messages
    - Not possible to map a source and an outgoing message (apart for the mix)
    - No possible to map a receiver and an incoming message (apart for the mix)

Trust in Mix

- But, the mix can make the correspondence between incoming and outgoing messages
  - If the mix compromised
    - Possible to know the sender and receiver for each message
    - But, impossible to find what is the message

Trust in Mix

- Use a cascade of mixes
  - A single mix in the cascade is enough to hide correspondences between incoming and outgoing messages
  - Work with a partially trusted set of mixes
    - As long as one mix in the cascade can be trusted
    - Or, as long as all untrusted mixes in the cascade do not cooperate

Cascade of Mixes

- No guarantee that it works
  - Increasing the number of mixes in the cascade
    - Increases the confidence
    - But, increases the end-to-end delay
- Tor uses at least 3 mixes selected at random (see [50] for details)
  - Called a Circuit
    - Periodically select new random mixes to form a new circuit

Goal of Chaum-net

Send sealed messages from Alice to Bob through a cascade of mixes

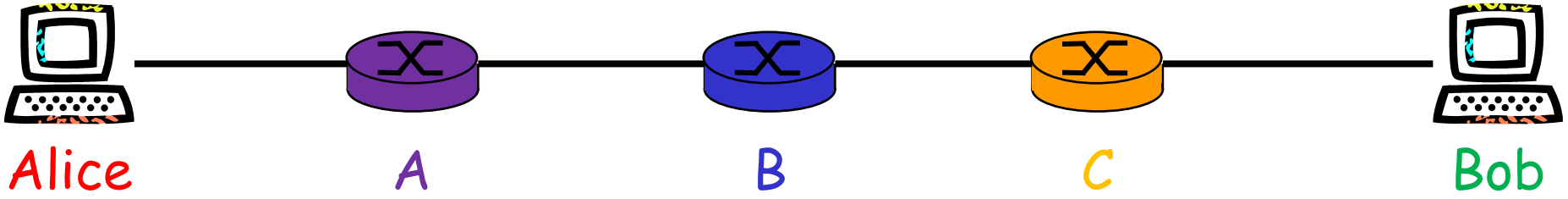
How It Works?

- Assume an overlay of end-users
  - Each end-user has a couple of private and public keys
    - We note the public key K_A for end-user A
  - The public keys and the address of owners are publicly available
    - (K_A, IP_A) for each end-user A
    - In a central repository, using a distributed storage, etc.

How It Works?

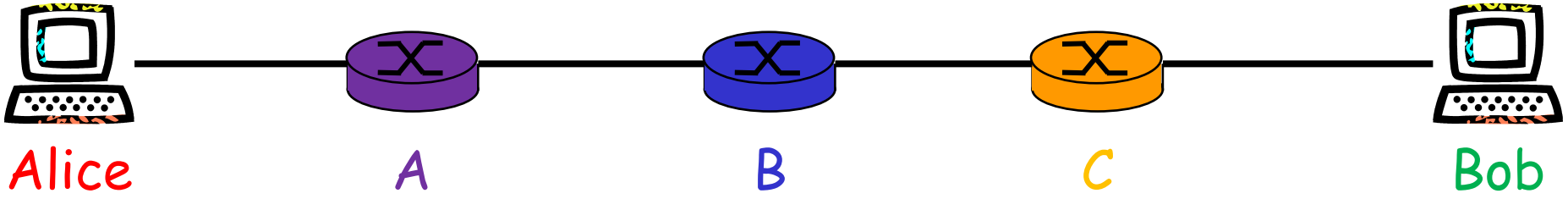
- Alice wants to send the message M to Bob
  - Any other end-user may act as a mix
  - Alice selects at random a few end-users
    - Get their public key and address
    - Typically select 3 mixes

How to Send the Message?



- The path is A -> B -> C -> Bob
  - Create layered (onion) sealed messages from Bob to A

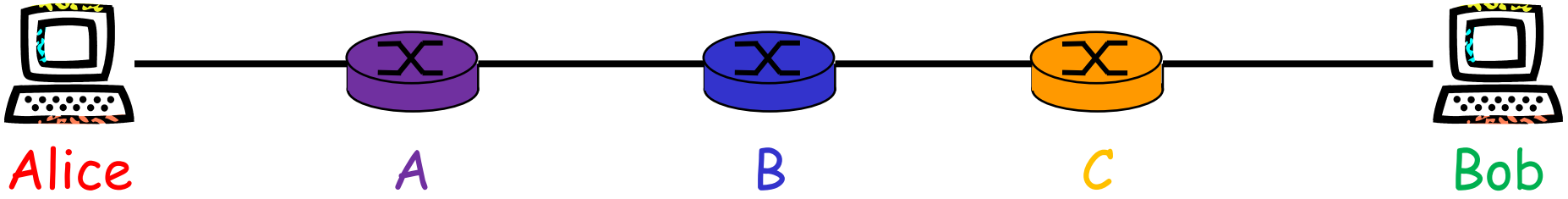
How to Send the Message?

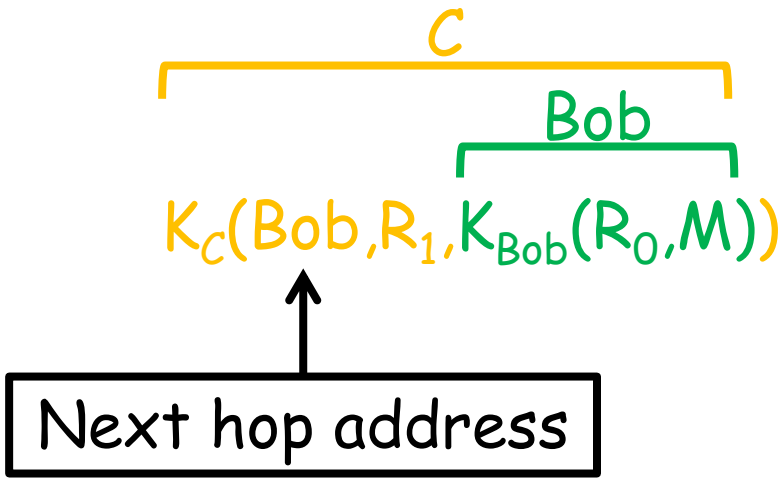


Bob

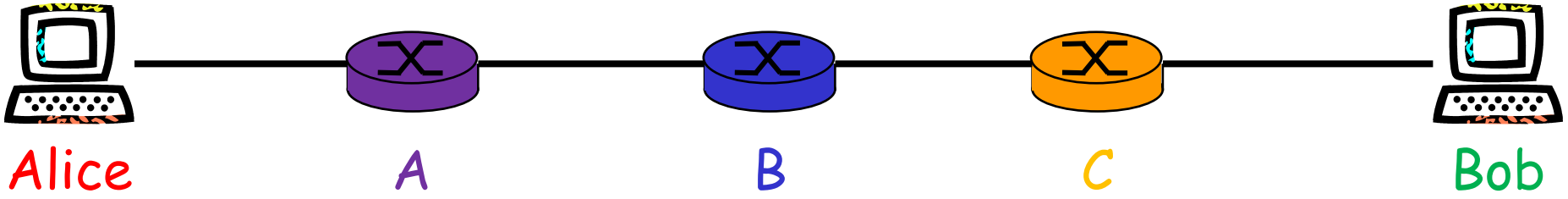
$K_{Bob}(R_0, M)$

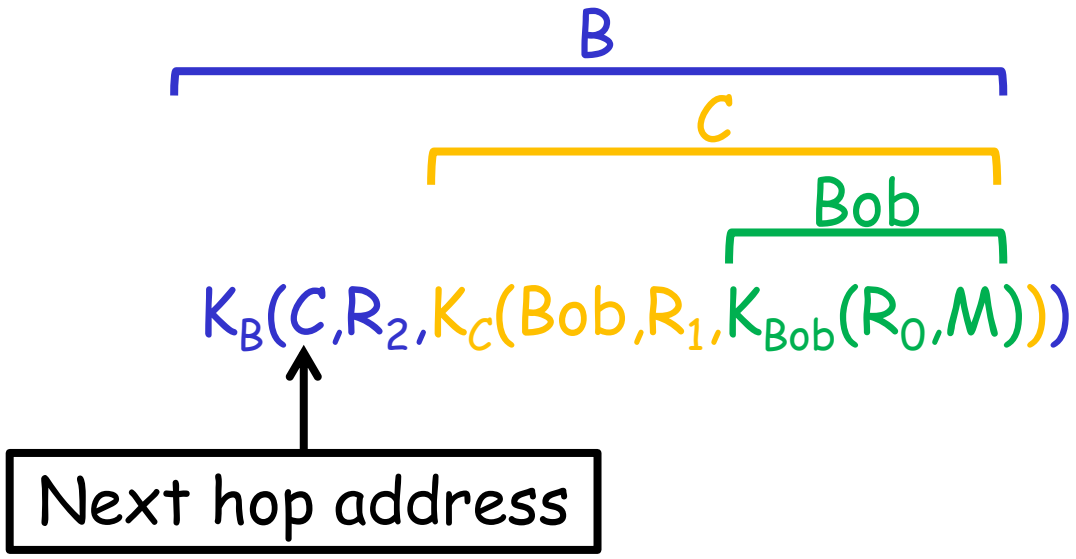
How to Send the Message?



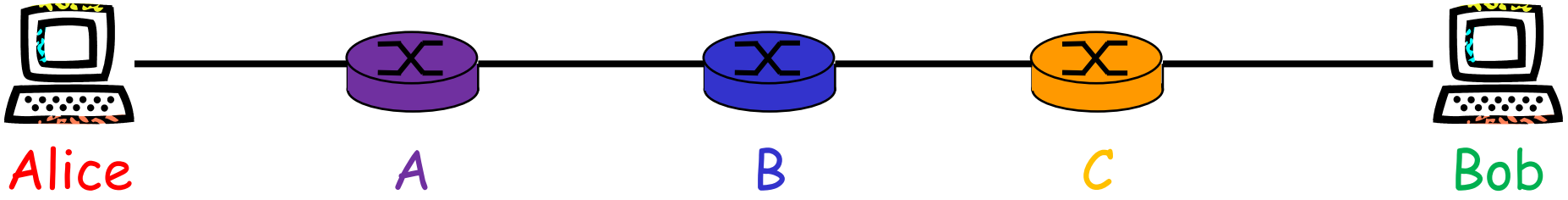


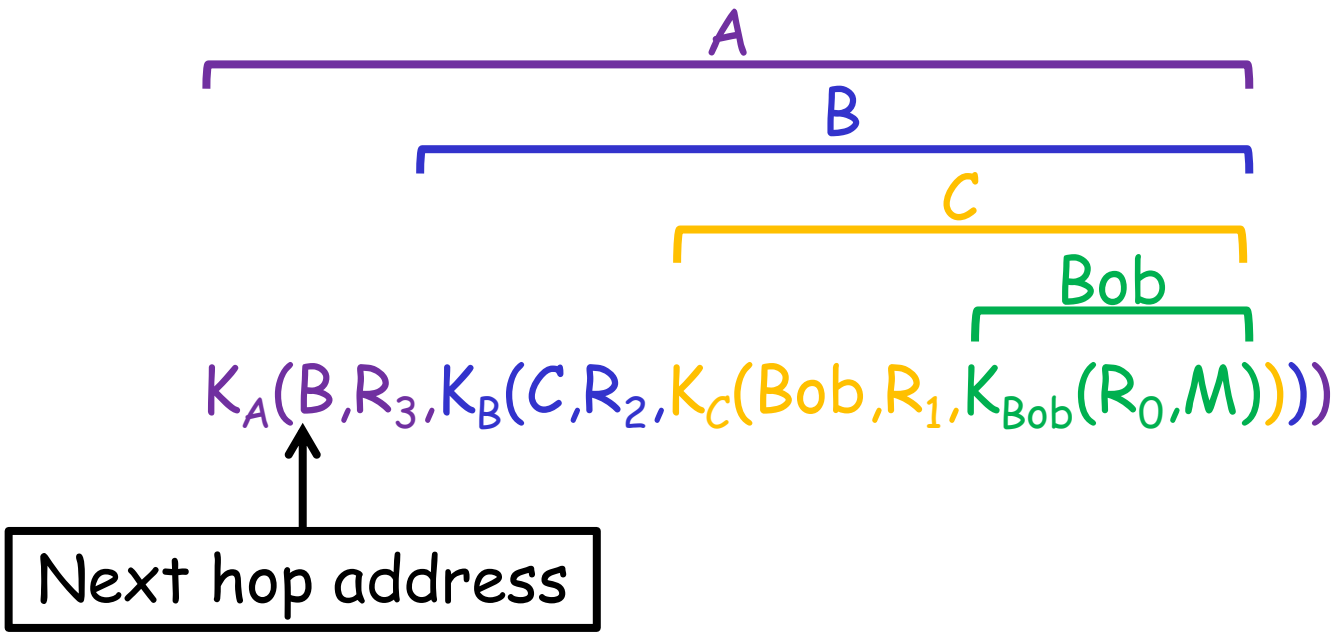
How to Send the Message?



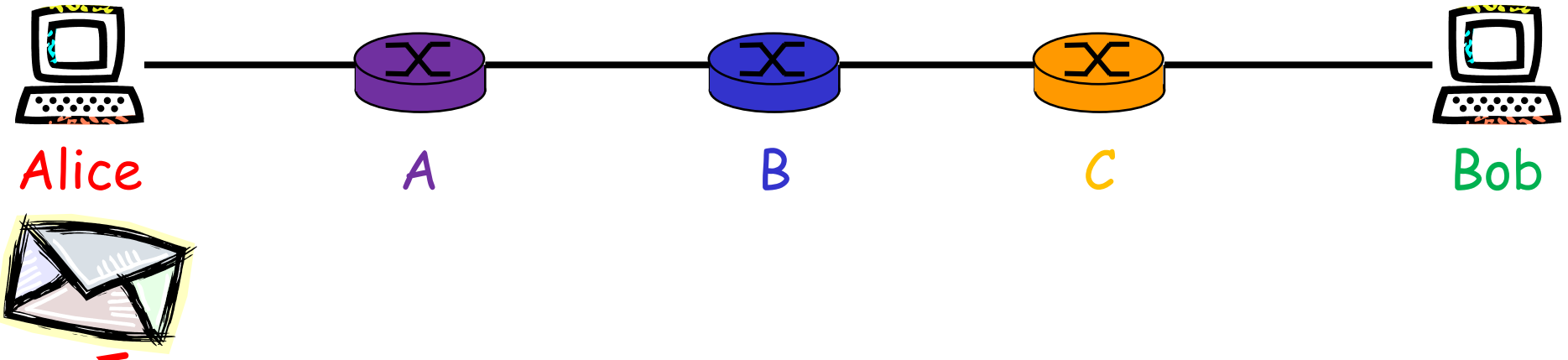


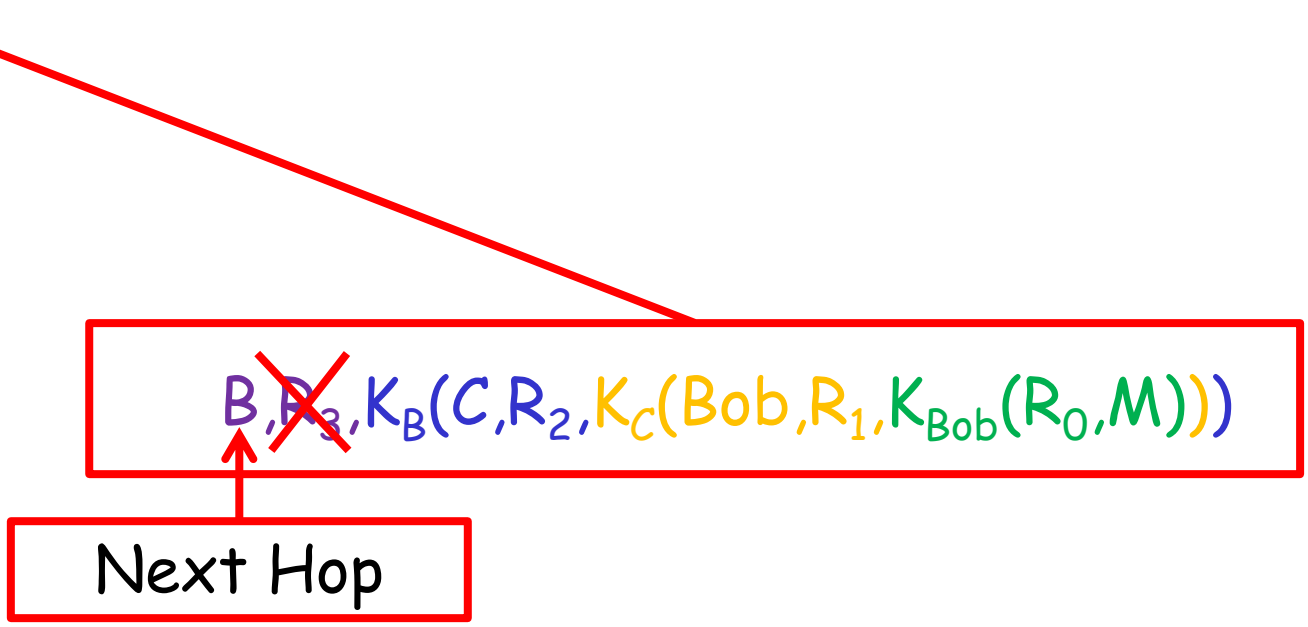
How to Send the Message?



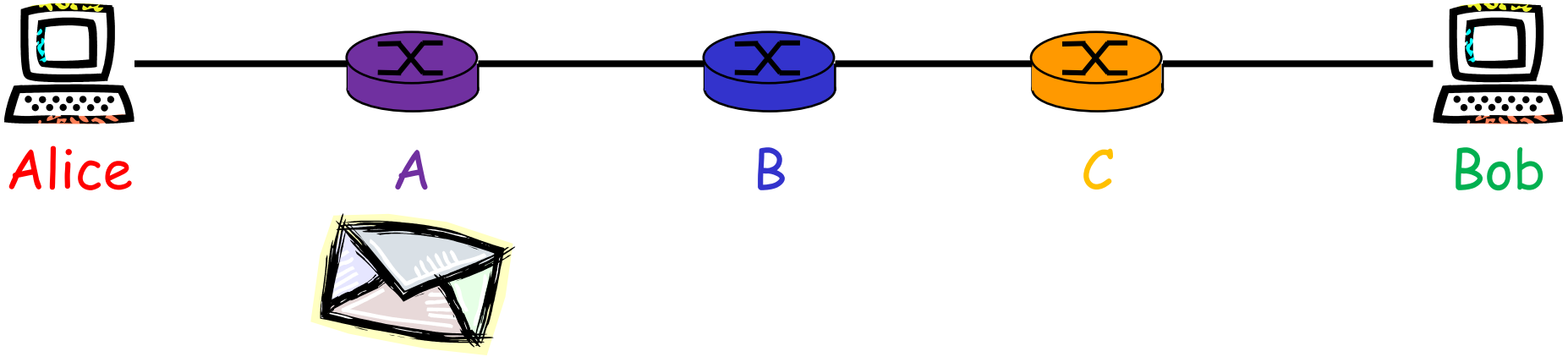


How to Relay the Message?





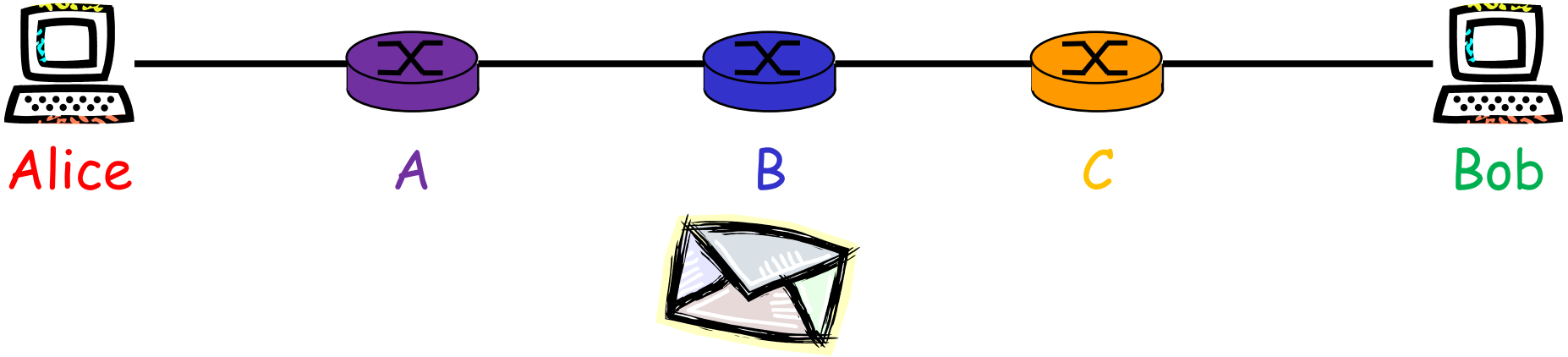
How to Relay the Message?

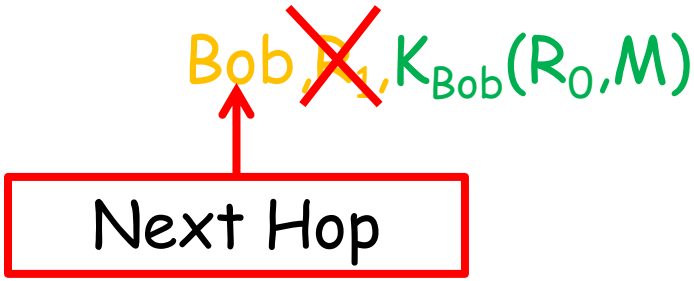


$C, \cancel{R_2}, K_C(Bob, R_1, K_{Bob}(R_0, M))$

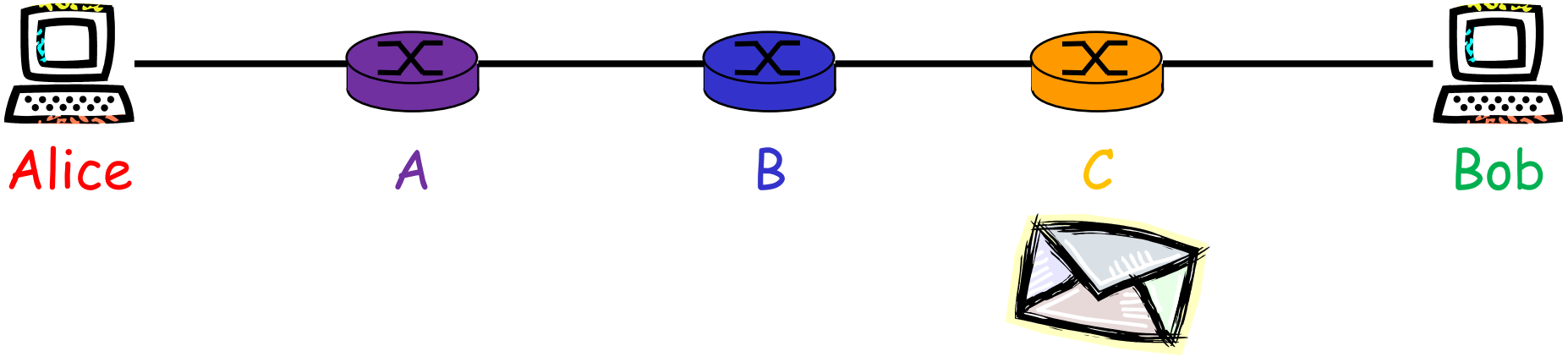
Next Hop

How to Relay the Message?





How to Relay the Message?



M

From Theory To Practice

- Attacks still possible
  - Timing
    - Correlate when packets are received and sent at each mix
    - Can be solved using batches
  - Active end-users
    - Possible to know who is sending and who is receiving
    - Can be solved with padding, but highly costly
  - Read [24] and [50]

Schneier on Security

Someone Is Running Lots of Tor Relays

Since 2017, someone is running about a thousand — 10% of the total — Tor servers in an attempt to deanonymize the network:

> Grouping these servers under the KAX17 umbrella, Nusenu says this threat actor has constantly added servers with no contact details to the Tor network in industrial quantities, operating servers in the realm of hundreds at any given point.
>
> The actor's servers are typically located in data centers spread all over the world and are typically configured as entry and middle points primarily, although KAX17 also operates a small number of exit points.
>
> Nusenu said this is strange as most threat actors operating malicious Tor relays tend to focus on running exit points, which allows them to modify the user's traffic. For example, a threat actor that Nusenu has been tracking as **BTCMITM20** ran thousands of malicious Tor exit nodes in order to replace Bitcoin wallet addresses inside web traffic and hijack user payments.
>
> KAX17's focus on Tor entry and middle relays led Nusenu to believe that the group, which he described as "non-amateur level and persistent," is trying to collect information on users connecting to the Tor network and attempting to map their routes inside it.
>
> In research published this week and shared with *The Record*, Nusenu said that at one point, there was a 16% chance that a Tor user would connect to the Tor network through one of KAX17's servers, a 35% chance they would pass through one of its middle relays, and up to 5% chance to exit through one.

Slashdot thread.

Tags: de-anonymization, privacy, Tor

Posted on December 7, 2021 at 6:25 AM • 46 Comments

https://www.schneier.com/blog/archives/2021/12/someone-is-running-lots-of-tor-relays.html

More Tor attacks

- https://en.wikipedia.org/wiki/Tor_(network)
- Stevens Le Blond, Pere Manils, Abdelberi Chaabane, Mohamed Ali Kaafar, Claude Castelluccia, Arnaud Legout, Walid Dabbous. **One Bad Apple Spoils the Bunch: Exploiting P2P Applications to Trace and Profile Tor Users**. In *Proc. of LEET'11*, March 29, 2011, Boston, MA, USA.
  - De-anonymized 10k IP addresses (23 days)
  - Traced 9% of all streams crossing our exit node!

Outline

- Privacy Foundations
- Privacy Attacks
  - Introduction to Privacy
  - Practical Attacks

Who (Really) Cares About Privacy?

- Privacy has an ambivalent status
  - Law/states fails to protect privacy of users in the internet
    - It is on a per-country basis, but the internet is worldwide
  - Users spread personal information all over the internet
    - Anonymizing IP addresses is ineffective
    - Extremely complex to preserve privacy

Why Should I Care?

- "I do nothing illegal"
  - Where?
  - No problem to publish all your browsing history of the past 2 years?

Why Should I Care?

- "I do not leave any personal information"
  - Are you browsing the Web?

Why Should I Care?

- "I don't have any immoral activities"
  - Morality is vastly different from countries to countries
    - Facebook breast-feeding vs. racism

Pornography

facebook

Freedom of speech

Why Should I Care?

- "I know that the Internet does not protect my privacy and I don't care"
  - No problem to see exposed you past 2 years BitTorrent downloads, mobility, and social interactions?
    - I explain how it is possible in the following

Why Is Privacy So Complex?

- Privacy is no more a protocol or system issue only
- Protocols and systems interact in many complex ways
  - Might be closed systems (facebook, google, skype, etc.)
  - Might have many implementation flavors (BitTorrent, HTML, Javascript, etc.)

Definition of Privacy Threat

- There is a privacy threat (or attack) when

activity is linked to identity

- Anonymization is breaking this link

Definition of Activity

- Your activity at any network layer
  - Web history
  - Localization (based on IP, MAC, etc.)
  - BitTorrent download history
  - VoIP conversation (who and what)
  - Etc.

Definition of Identity

- Three different notion of identities
  - Used in different ways
- Network identity
  - An IP address in IP networks
  - Enables network profiling
- Application identity
  - An application specific unique identity
    - Temporary (cookie) or permanent (skype ID)
  - Enables attacks targeted to applications
    - Not always possible to exploit the network identity

Definition of Identity

- Social identity
  - Everything that identify a user in real life
    - Name, postal address, email, etc.
  - Enables sophisticated and harmful attacks
    - Blackmail, targeted phishing attacks, etc.

The more identities an attacker knows, the more severe the attack

IP Address and Social Identity

- The IP address is a signature of your network activity
  - Only your ISP knows the mapping (IP address, social identity)
  - Some big companies might know
- Strong privacy issue if an individual can link IP address and social identity
  - More sophisticated attacks

Who Can Perform a Privacy Attacks

- Big companies or ISPs
  - Google, Facebook, Dropbox, Amazon, etc.
  - Anybody that can ask data to those companies
    - National security agencies, judge, etc.
- Individuals
  - No dedicated infrastructure
  - No access to privileged information
  - Can be an employer, a relative, a criminal, etc.

What Can Big Companies Do?

« If there are stuff you want to be hidden to the public, you are better not doing them in the first place. »

Eric Schmidt, CEO Google, 2009

Big Internet Companies

Which of Your Data Are Available to Big Companies?

- Data published with user consent
  - Facebook, twitter, Google apps, Dropbox, blogs, Web site, etc.
- Data published without user consent
  - Web history, localization, identifiers (IP, port, unique user ID, etc.)
  - Data published with consent on which you lost control
- Correlate data
  - User profiling

Data Published With User Consent

- Huge amount of data published
  - Facebook, twitter, gmail, etc.
- Users trust big companies
  - They gave their consent by accepting the “Terms of use” and “privacy policy”
    - But, they don’t read them
    - Contradictory and never preserve privacy

We will share personal information with companies, organizations or individuals outside of Google when we have your consent to do so. We

We will share personal information with companies, organizations or individuals outside of Google if we have a good-faith belief that access, use, preservation or disclosure of the information is reasonably necessary to: …

Data Published With User Consent

- Informed consent impossible
  - People don't know where their data are
  - Data covered by the law that apply to
    - The data location or nationality of the company
    - Choice always made at the detriment of privacy

Data Published Without User Consent

- Every site that embeds javascript from other sites might leak that you are visiting it
  - Facebook 

  - Twitter 

  - Google analytics (you don't even see a button)
    - xvideos.com (48), pornhub.com (67), youporn.com (80)
    - isohunt.com
    - 4chan.org (can you believe it?)

Why Private Data Are Collected

- Companies use all collected data to
  - Offer a better service
  - Run their business of targeted advertisements
  - Comply with law enforcement requests

Don’t be naïve: Google’s business is ads!

Note 2. Revenues

Revenue Recognition

The following table presents our revenues disaggregated by type (in millions).

	Three Months Ended June 30,		Six Months Ended June 30,	
	2020	2021	2020	2021
Google Search & other	$ 21,319	$ 35,845	$ 45,821	$ 67,724
YouTube ads	3,812	7,002	7,850	13,007
Google Network	4,736	7,597	9,959	14,397
Google advertising	29,867	50,444	63,630	95,128
Google other	5,124	6,623	9,559	13,117
Google Services total	34,991	57,067	73,189	108,245
Google Cloud	3,007	4,628	5,784	8,675
Other Bets	148	192	283	390
Hedging gains (losses)	151	(7)	200	(116)
Total revenues	$ 38,297	$ 61,880	$ 79,456	$ 117,194

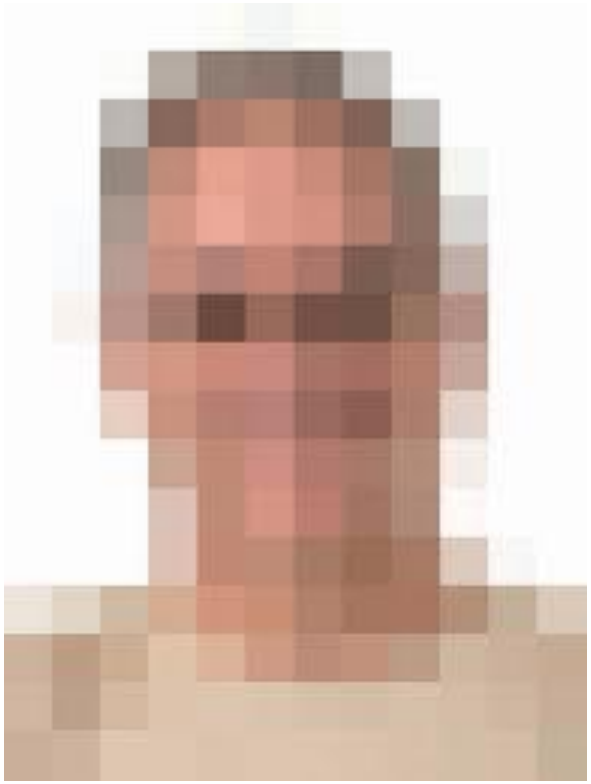
https://abc.xyz/investor/static/pdf/20210728_alphabet_10Q.pdf

Why Private Data Are Collected

- Risk assessment
  - Risk
    - Huge amount of collected information
    - Know your social and network identities
  - Mitigate risk
    - Core business based on user satisfaction. No user, no business
    - Under scrutiny
    - Subject to laws

What Can An Individual Do?

Definition of Individuals



- No dedicated infrastructure
- No access to privileged information
- Can be an employer, a relative, a criminal, etc.

Which of Your Data Are Available to Individuals?

- Overlooked issue
- Risk assessment
  - Risk
    - No control, hard to identify
    - Motivation to infringe privacy: blackmail, phishing…
  - Mitigate risk
    - Believed to be hard to perform at scale

Focus of Bluebear

- We explore whether it is possible for an individual to access all your legal activities that you don't want to disclose for any reason (it is your privacy)
  - Can be used for blackmail, phishing, social attacks, etc.
- We show that anybody (not a government, not a big company) can severely infringe users privacy

Bluebear Publications [52]

- *Spying the World from your Laptop - Identifying and Profiling Content Providers and Big Downloaders in BitTorrent*
  - Map IP addresses to all their BitTorrent downloads
    - 148M IP addresses in 1.2M torrent during 103 days
  - Find IP addresses of BitTorrent sources
    - 70% of all BitTorrent sources
- TR: Angling for Big Fish in BitTorrent

Bluebear Publications [55]

- *One Bad Apple Spoils the Bunch: Exploiting P2P Applications to Trace and Profile Tor Users*
  - Identify the public IP addresses of BitTorrent users on Tor
  - Bad apple attack
    - Intra-circuit and inter-circuit
    - Map non BitTorrent streams to the IP addresses

Bluebear Publications [54]

- *I Know Where You are and What You are Sharing: Exploiting P2P Communications to Invade Users' Privacy*
  - Exploit Skype to map an identity to an IP address
  - Use this exploit to track mobility and associate an identity to a list of BitTorrent downloads

Outline

- Privacy Foundations
- Privacy Attacks
  - Introduction to Privacy
  - Practical Attacks
    - Distributed systems
    - Web

Spying the World From Your Laptop [52]

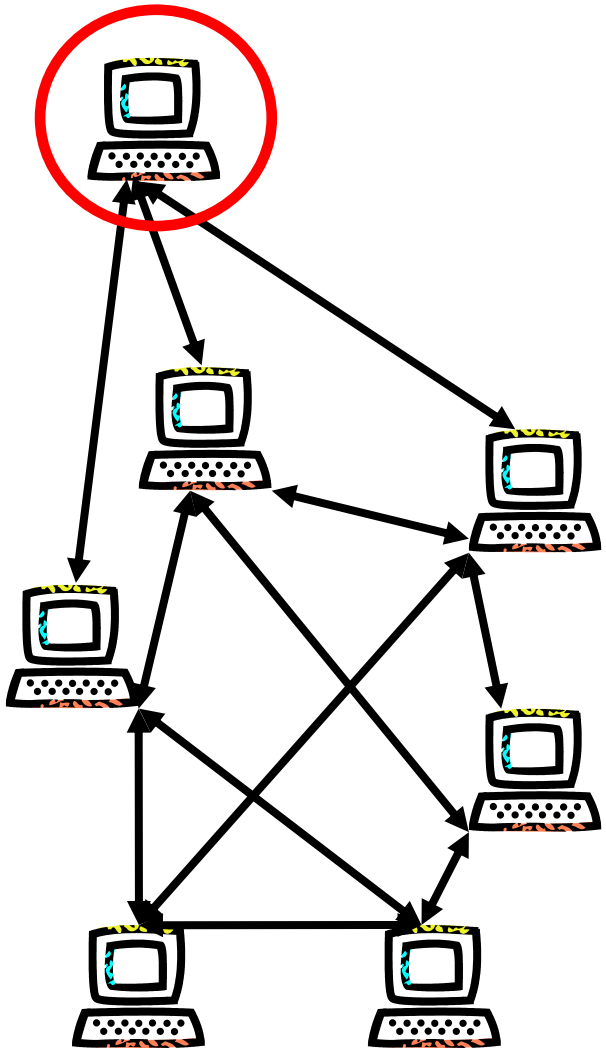
Why BitTorrent?

- BitTorrent widely popular
  - Several 10M of users at any moment in time
  - Several 100M of users cumulated over months
- BitTorrent most efficient P2P protocol
  - The only one candidate for legal P2P delivery

What is the privacy implication of BitTorrent usage?

BitTorrent Overview

BitTorrent Overview



- Who inserts contents?
- Who is downloading what?

Why Is It Hard?

- No way to get this information directly
  - Very good engineering of the implementations
  - Many blacklisting policies
- Need to correlate many different sources of information
  - Deep understanding of protocols and implementations
    - Experiments and measurements

Why Is It Hard?

- Design goal
  - Data collection without dedicated infrastructure and without being blacklisted
- High volume of data
  - 148M IP addresses * 1.2M contents
    - 2000M downloads
    - 3TB of storage on a NAS

Challenge in collecting and analyzing the data

Who inserts contents?

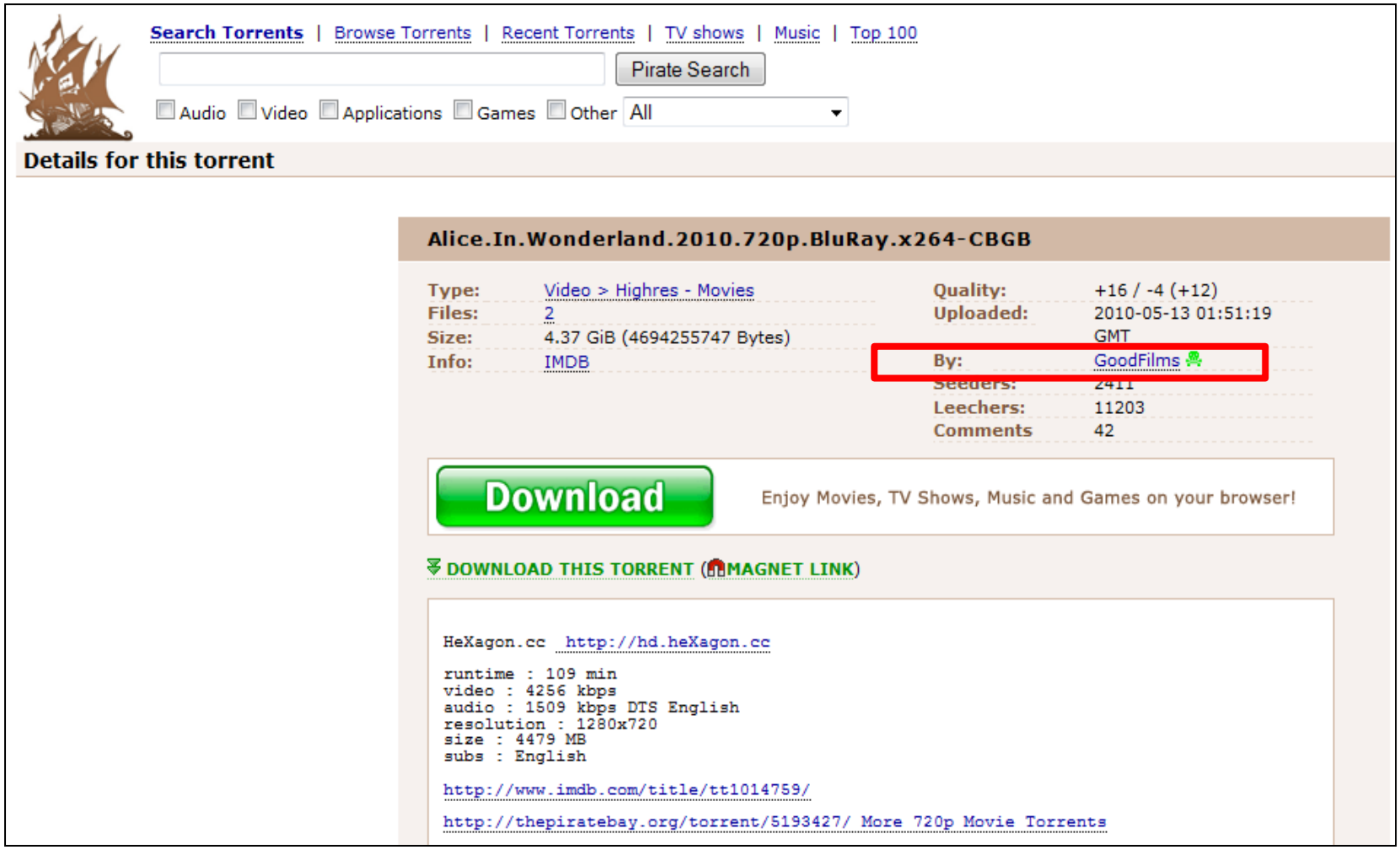
State of the Art

- BitTorrent
  - Introduction in 2000
  - Half of the Internet traffic in 2004
- Nobody ever looked at content providers
  - Believed to be impossible
    - Initial private phase for torrents
    - Content providers lies on their status

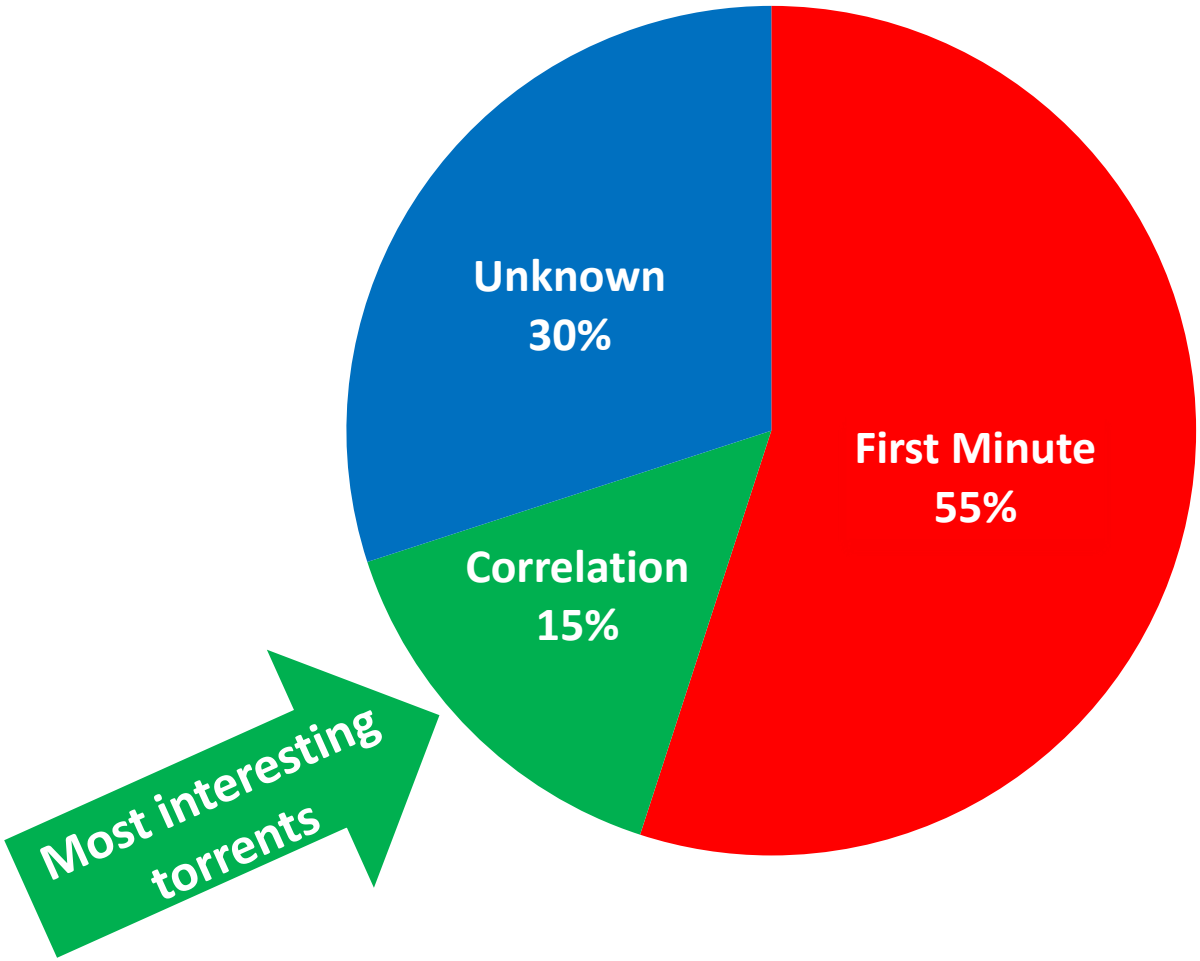
Method: First Minute

- Join torrents within its first minute
  - After announcement on TPB web site
  - If we are alone with another peer
    - It is the initial seed
- Fails for most interesting torrents

Method: Correlation



Success of the Method



Who is downloading what?

Method

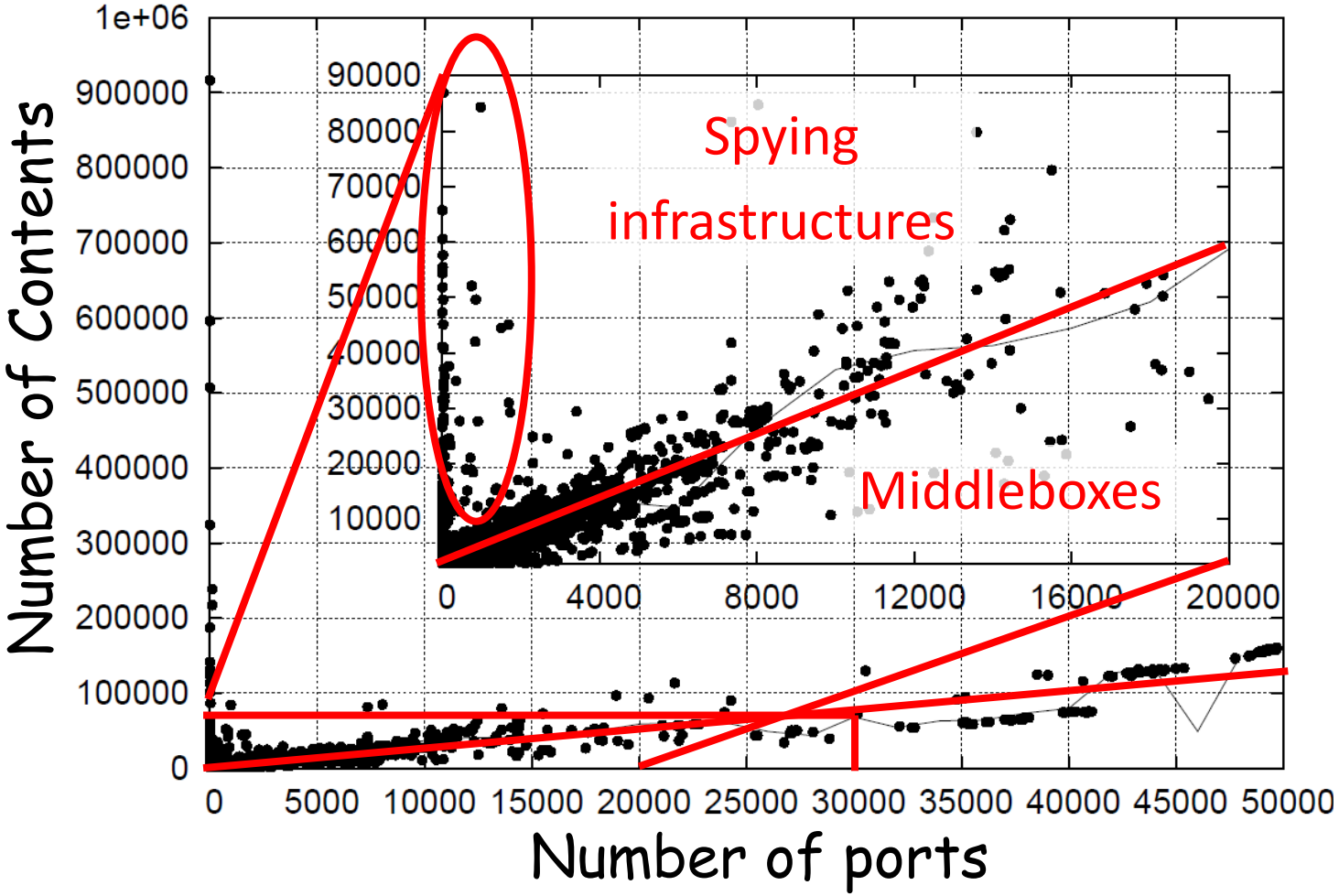
- Retrieved continuously IP addresses of most BitTorrent peers
  - For 103 days, every two hours
    - 700 000 torrents per snapshot
    - 5M to 10M IP addresses
  - ➢ 148M IP addresses in 1.2M torrents downloading 2000M of contents
- Analysis of such a large amount of data is complex

How To Identify Heavy Downloaders?

- Lets take top 10,000 IP addresses
  - Subscribed to at least 1636 contents

Do we find the heavy downloaders?

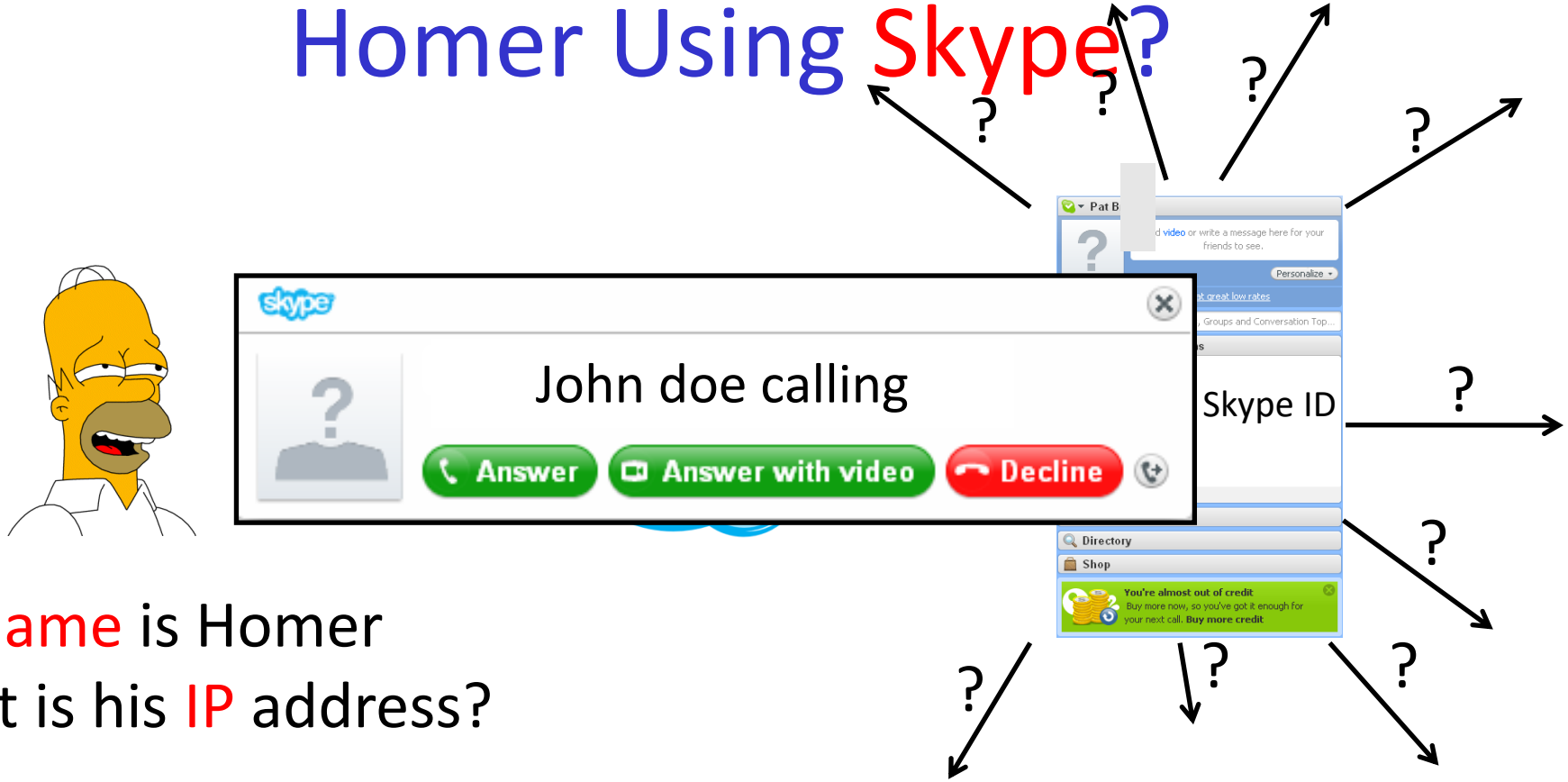
Who Are These Peers?



I Know Where You are and What You are Sharing [54]

Link an IP address to a social identity

Can We Find the IP Address of Homer Using Skype?



His name is Homer

What is his IP address?

- ❑ Step 1: Can we find Homer's Skype ID?
- ❑ Step 2: Can we find Homer's IP address from Skype?
- ❑ Step 3: Can we find Homer's IP address silently?

Step 1: Can We Find Homer's Skype ID?

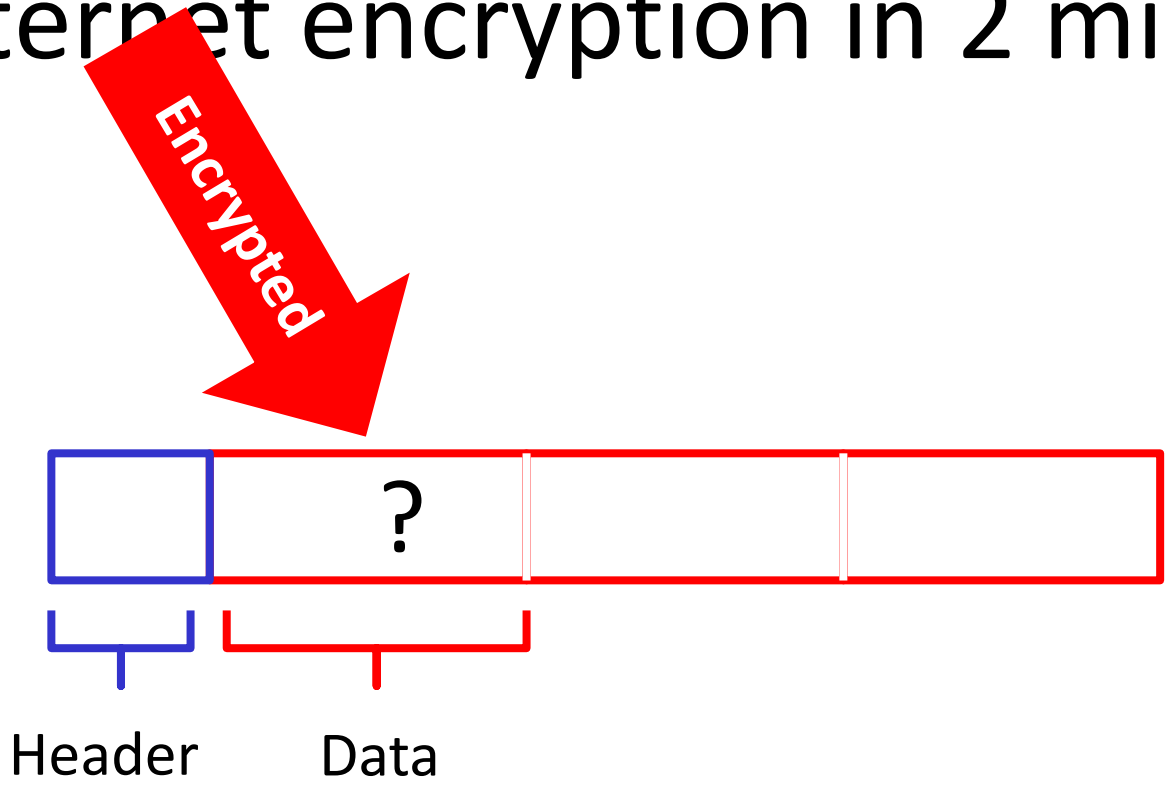
- 560M Skype users
  - 88% provides their birth name
  - 82% provides either age, country, URL, etc.
- Search for Homer on Skype directory
  - Remove duplicates based on country and city
- Call Candidates Homer
  - Disambiguate using the found IP address based on known locations
    - Enterprise, University, Internet coffee

Step 2: What Is Homer's IP Address From Skype?

- All communications encrypted
  - Not possible to get Homer IP address from the message payload
- Caller communicates with tens of peers
  - Who is the Homer out of 100 peers?
- We perform a VoIP call to Homer

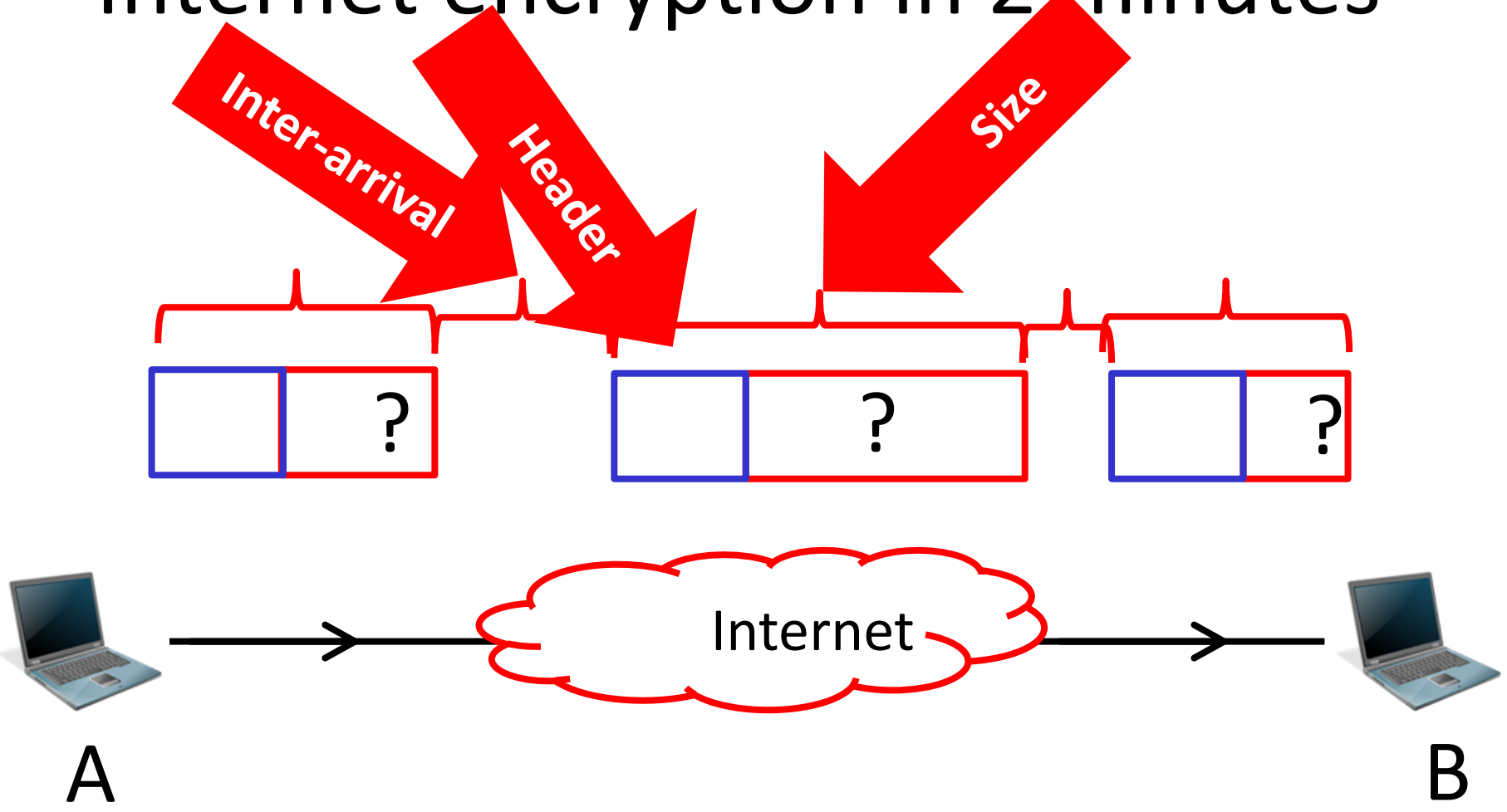
We identified specific traffic patterns with the callee

Internet encryption in 2 minutes

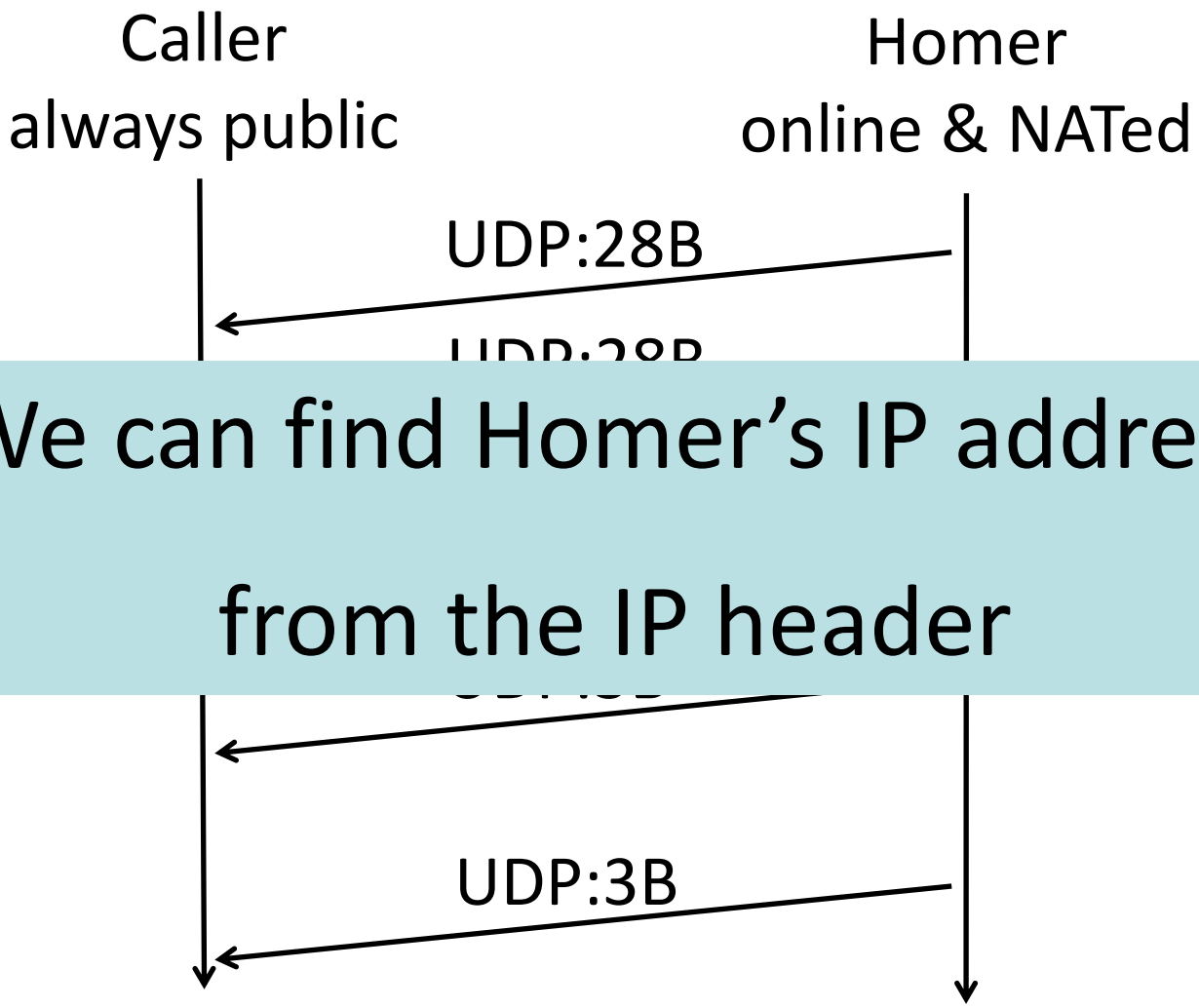


If it is encrypted, it is secure? Not really!

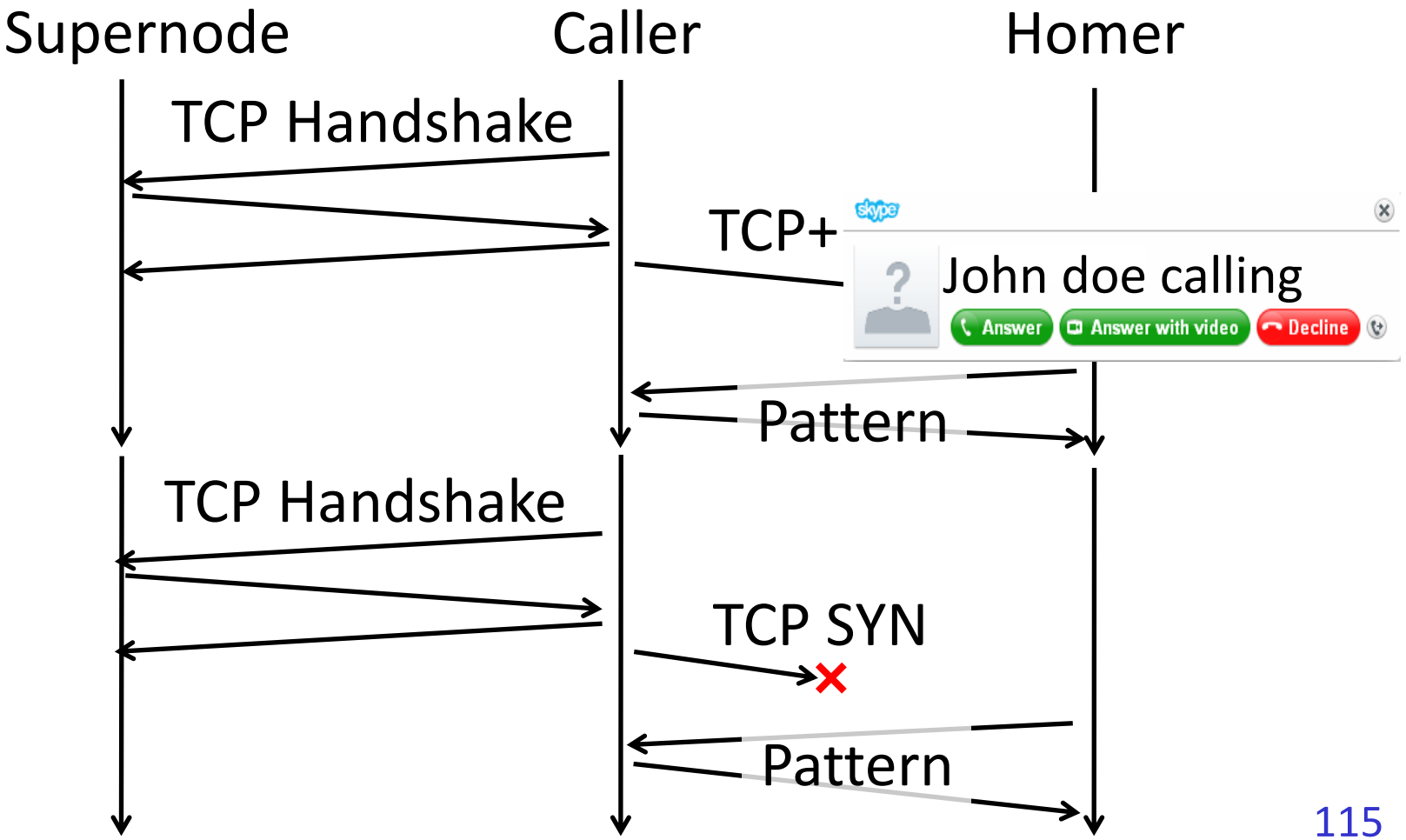
Internet encryption in 2 minutes



One Example of Pattern



Step 3: Can We Find Homer's IP Address Silently?



We Can Link an IP Address to an Social Identity

- Works for all 560M Skype users
- Undetectable and unstoppable
- No dedicated infrastructure

Mobility of Skype Users

What Is the Problem to Track My Mobility?

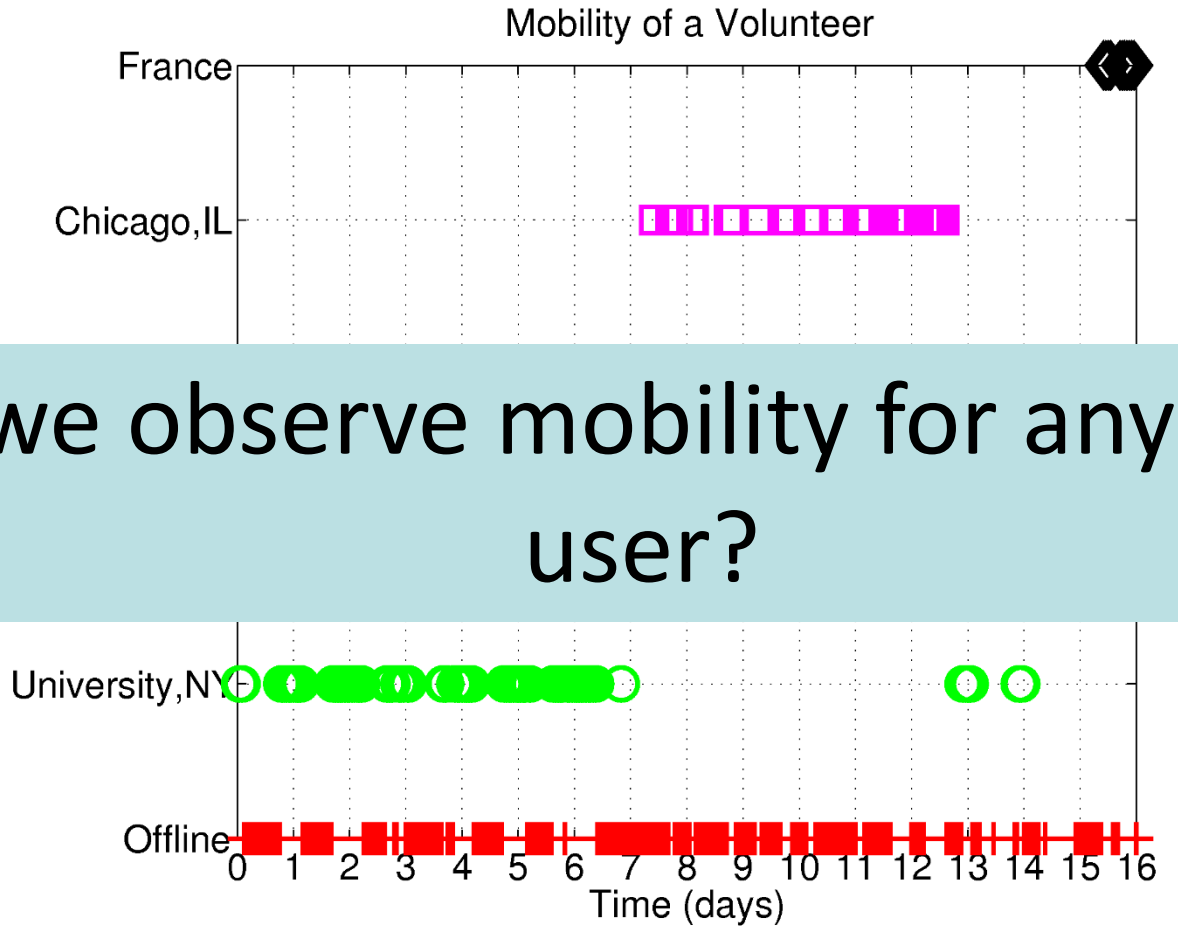
- Tracking mobility means
  - Knowing where you are
  - Knowing who you meet and where

- Tracking social interactions is a huge privacy concern

Real Use-Case: What Can We Get?

Name
Gender
Birthday
Language
City of residence
Job
Photos
Friends

Real Use-Case: What Can We Get?

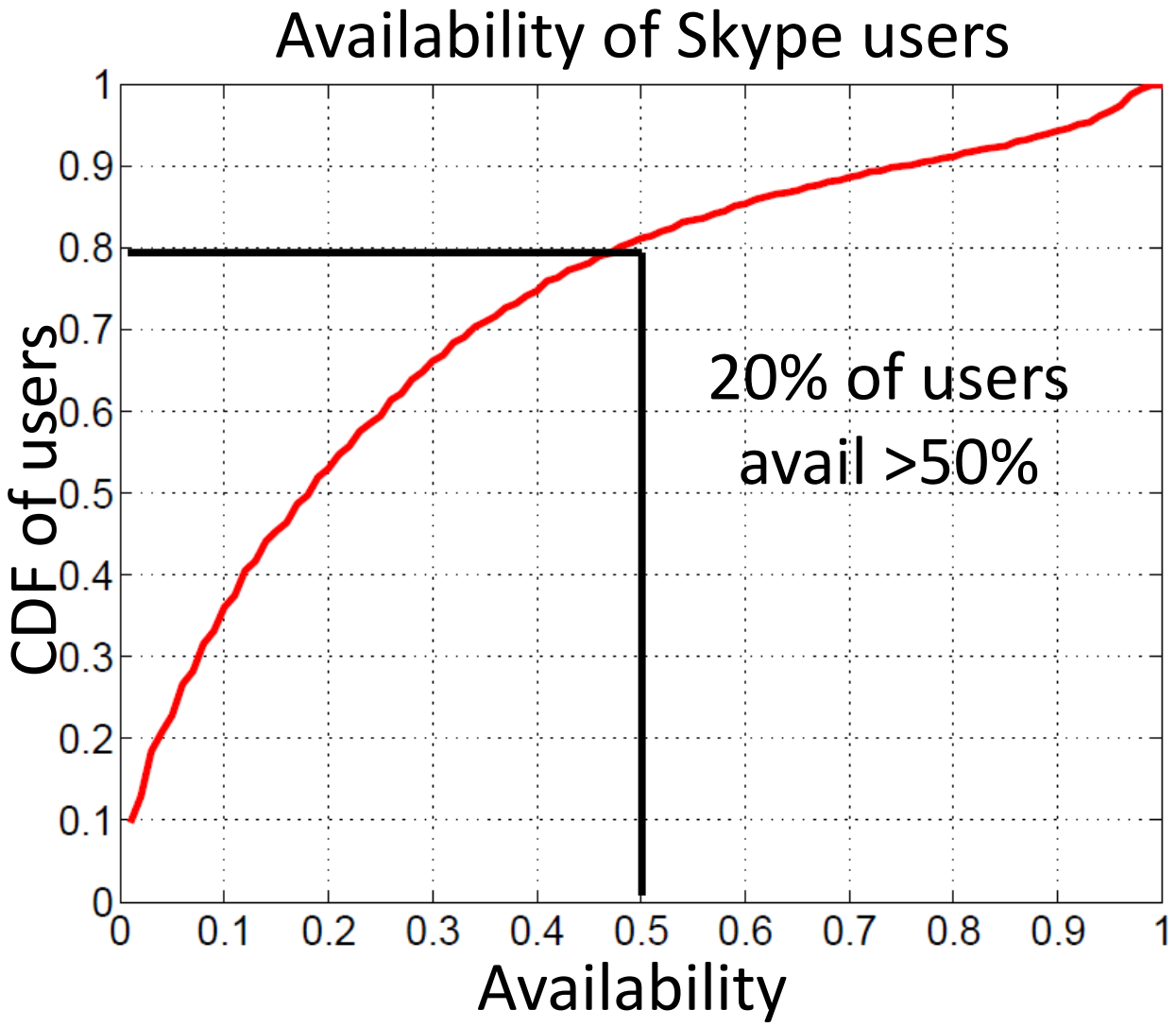


Do we observe mobility for any Skype user?

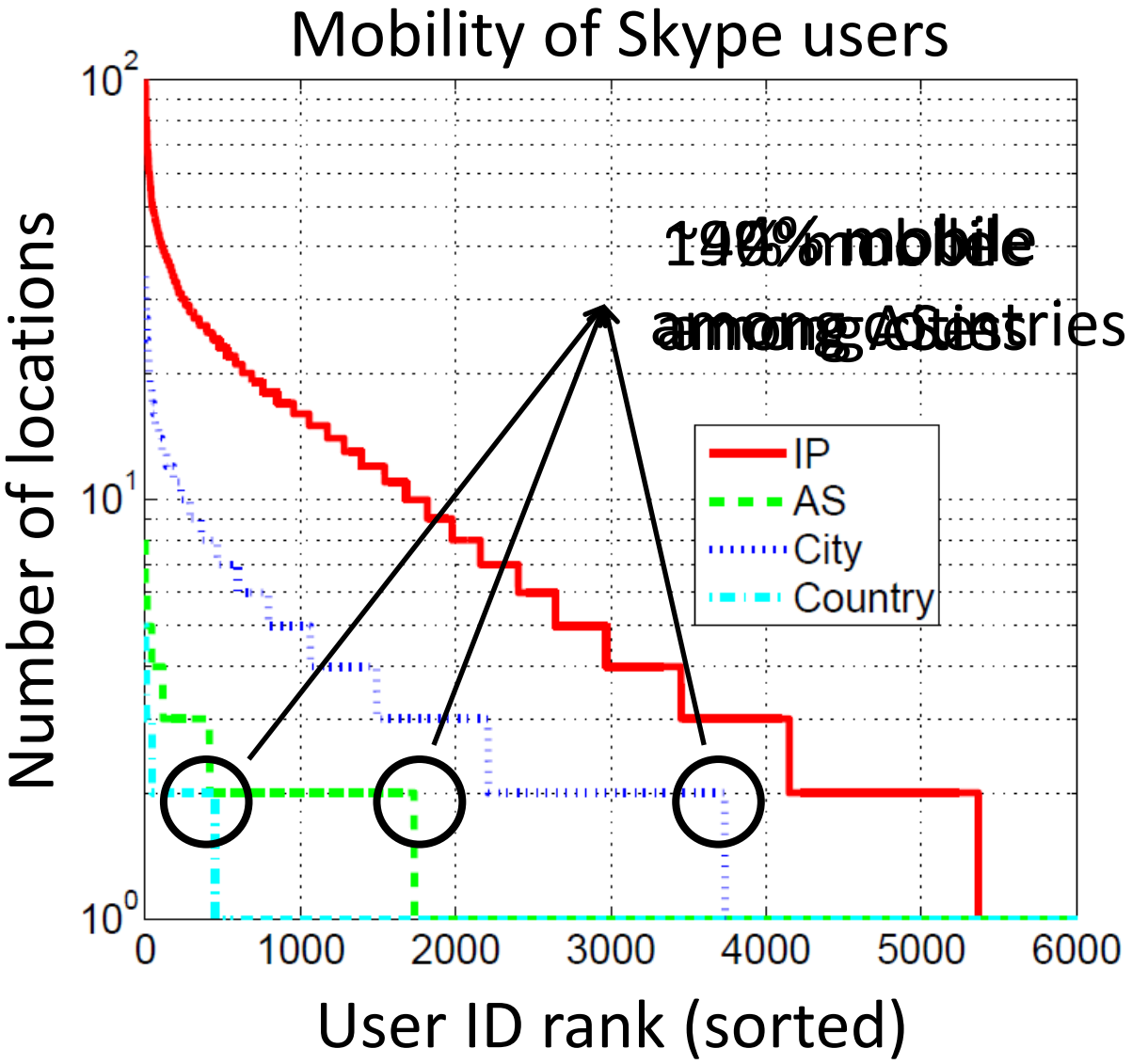
Can We Call Many Users Hourly?

- With a simple optimization
  - We can track from a single machine 340 Skype users per hour
- With 30 machines we can track 10,200 Skype users hourly
  - Cost $500 per week on Amazon EC2
  - Using virtualization we can significantly reduce cost (down to $50 per week)

Do We Observe Availability?



Do We Observe Mobility?



We Can Massively Track Mobility

- Works for all 560M Skype users
- Undetectable and unstoppable
- No costly infrastructure
- We observe real mobility

File-Sharing usage of Skype Users

Can We Associate Identified Users to Their Downloads?

NAT

"What am I downloading?"

BitTorrent

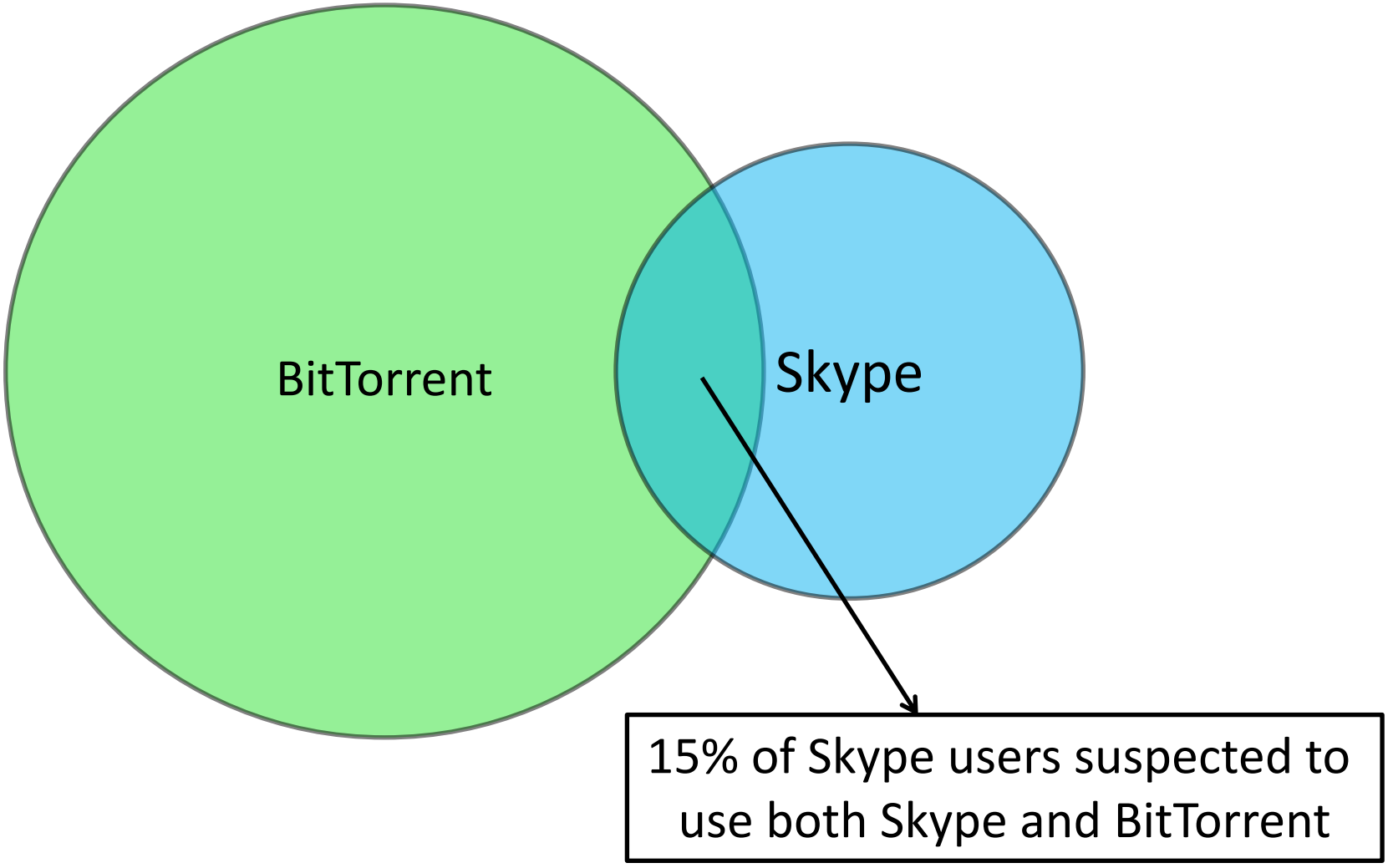
- Do we find IP addresses using both Skype and BitTorrent?
- Do NATs introduce false positives?
- Can we identify users despite NATs?

Methodology

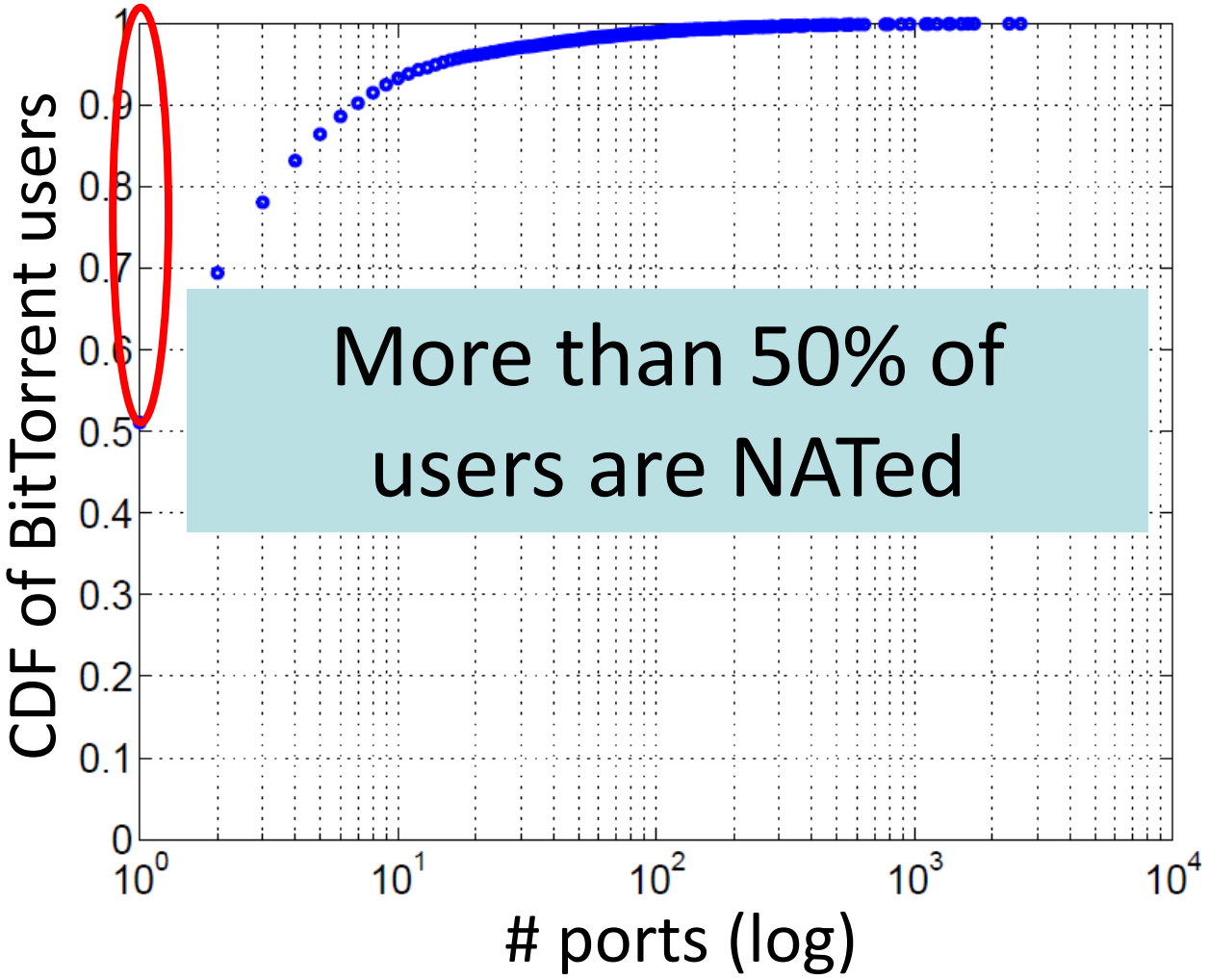
- Daily crawl 100,000 active Skype users
  - Chosen out of 1M random Skype users
- Crawl every hours 50,000 most popular torrents

- If an IP address appears in both Skype and BitTorrent it is a suspected user

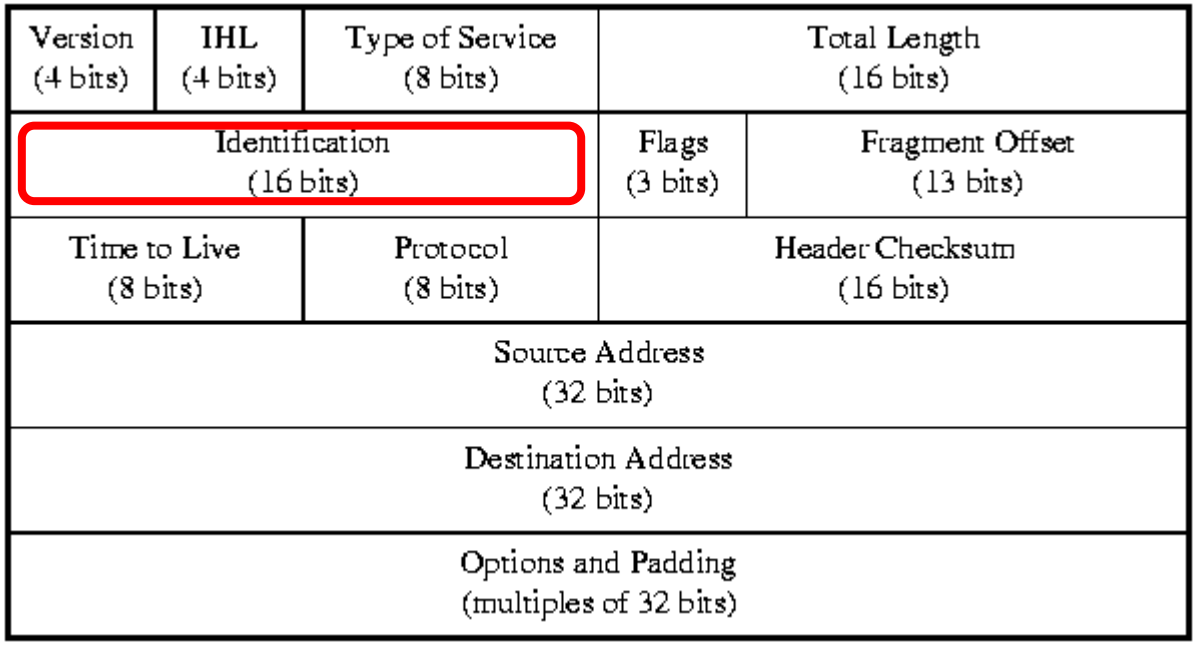
Do We Find IP Addresses Using Both Skype and BitTorrent?



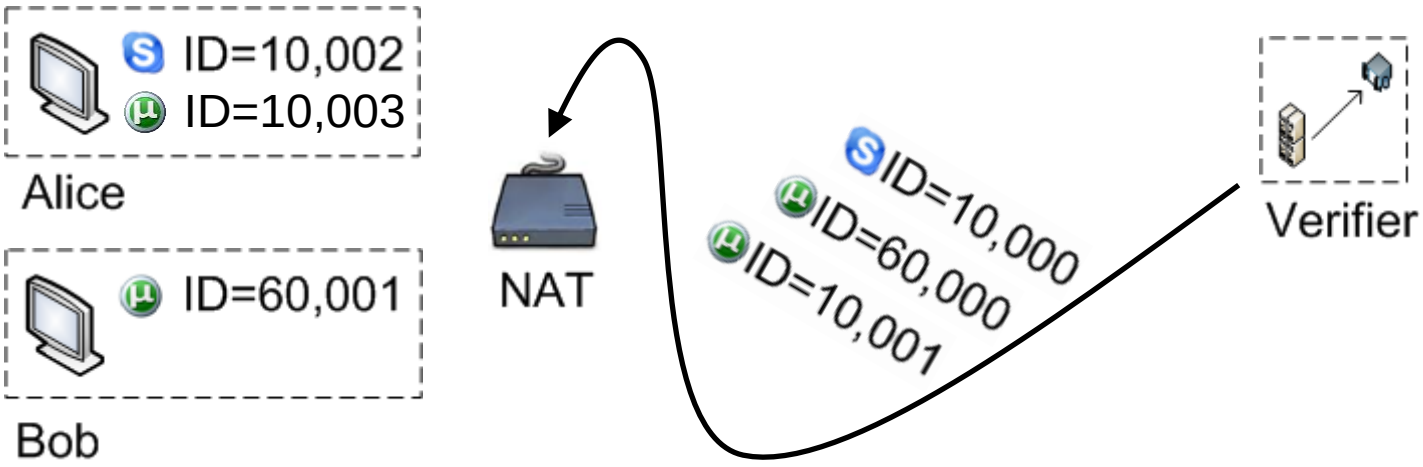
Do NATs Introduce False Positives?



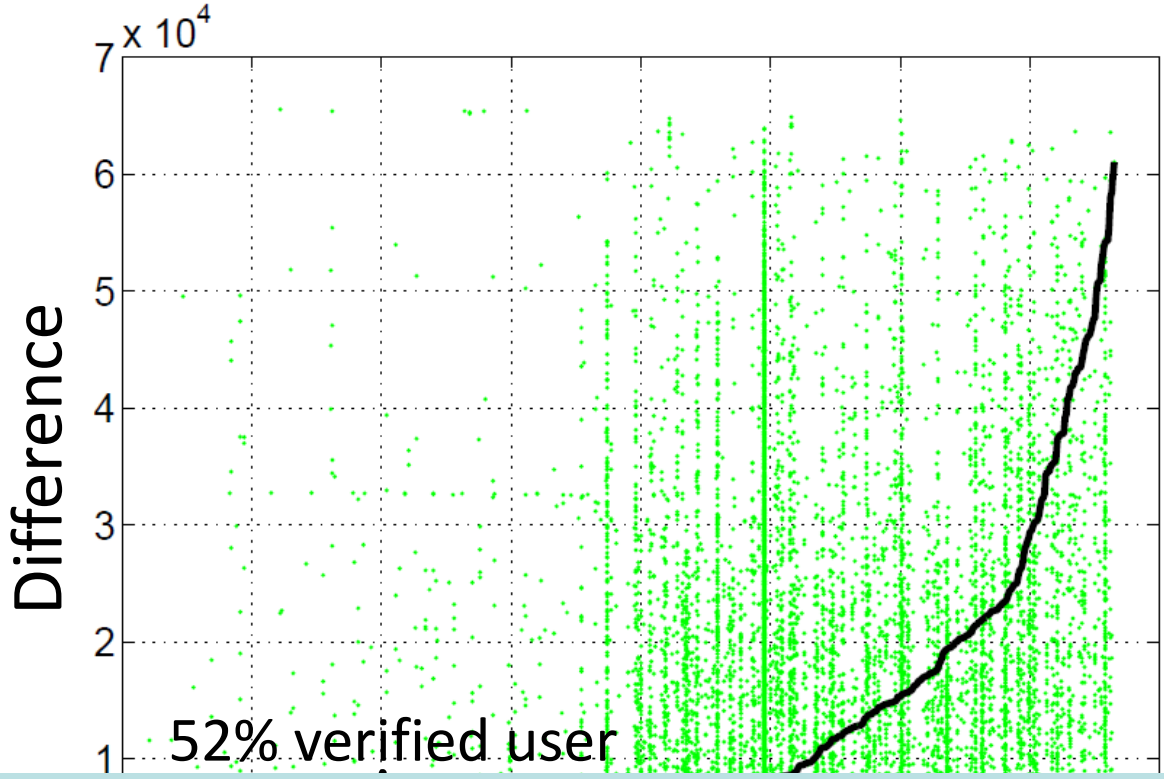
Can We Identify Users Despite NATs?



Can We Identify Users Despite NATs?



Verifying Downloaders



Verification reveals
~48% of false positives

Personal Info for the Top10 Verified Downloaders

Rank	# Files	First name	Last name	City	Country
1	23	✓	✓	✓	✓
2	18	✓	✓	✓	✓
3	12	✓	✓	×	✓
4	11	✓	✓	✓	✓
5	11	✓	✓	✓	✓
6	11	✓	✓	✓	✓
7	9	×	✓	✓	✓
8	8	×	✓	✓	✓
9	7	✓	✓	✓	✓
10	6	✓	✓	✓	✓

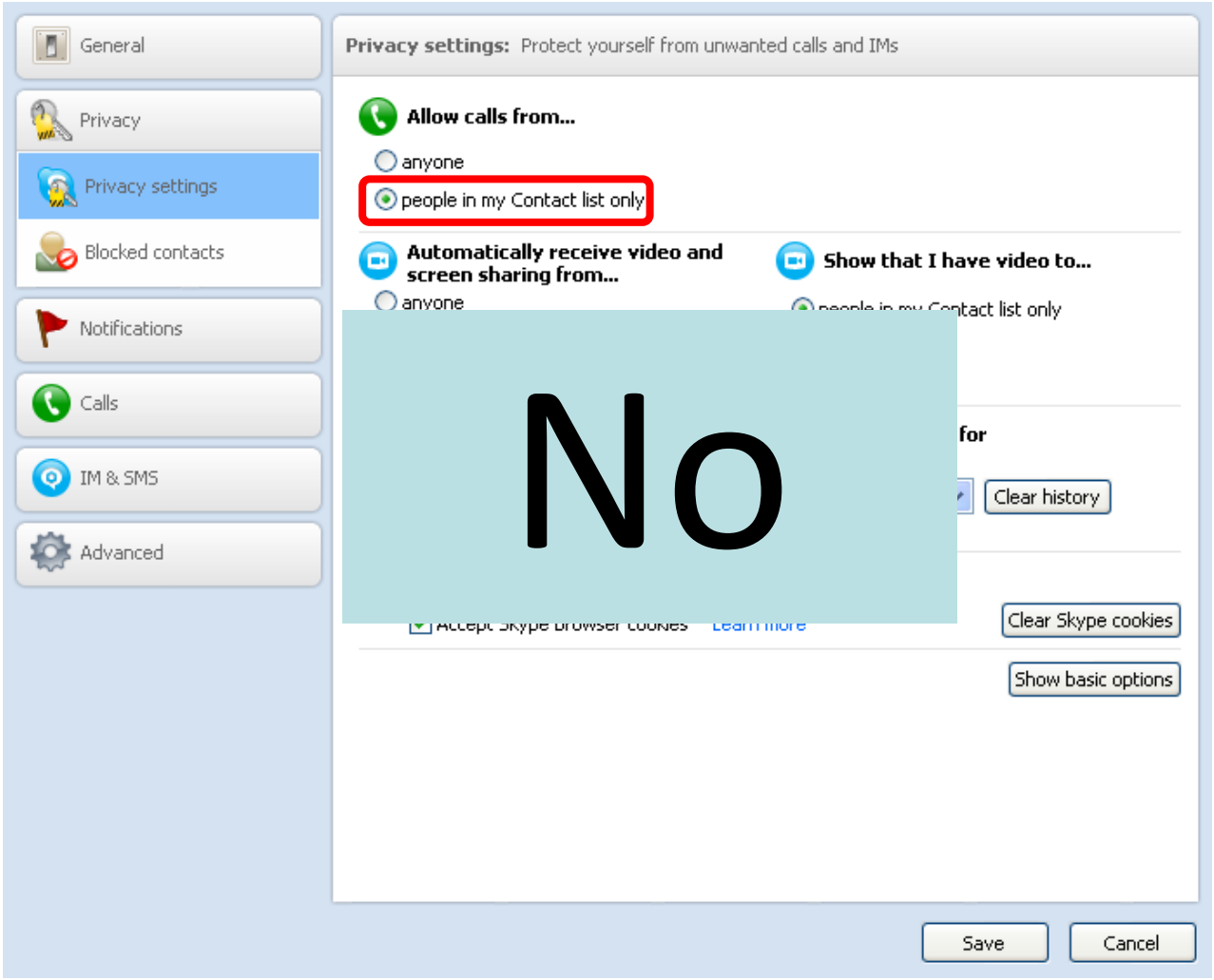
We find BitTorrent users identity without ISP support

Conclusion

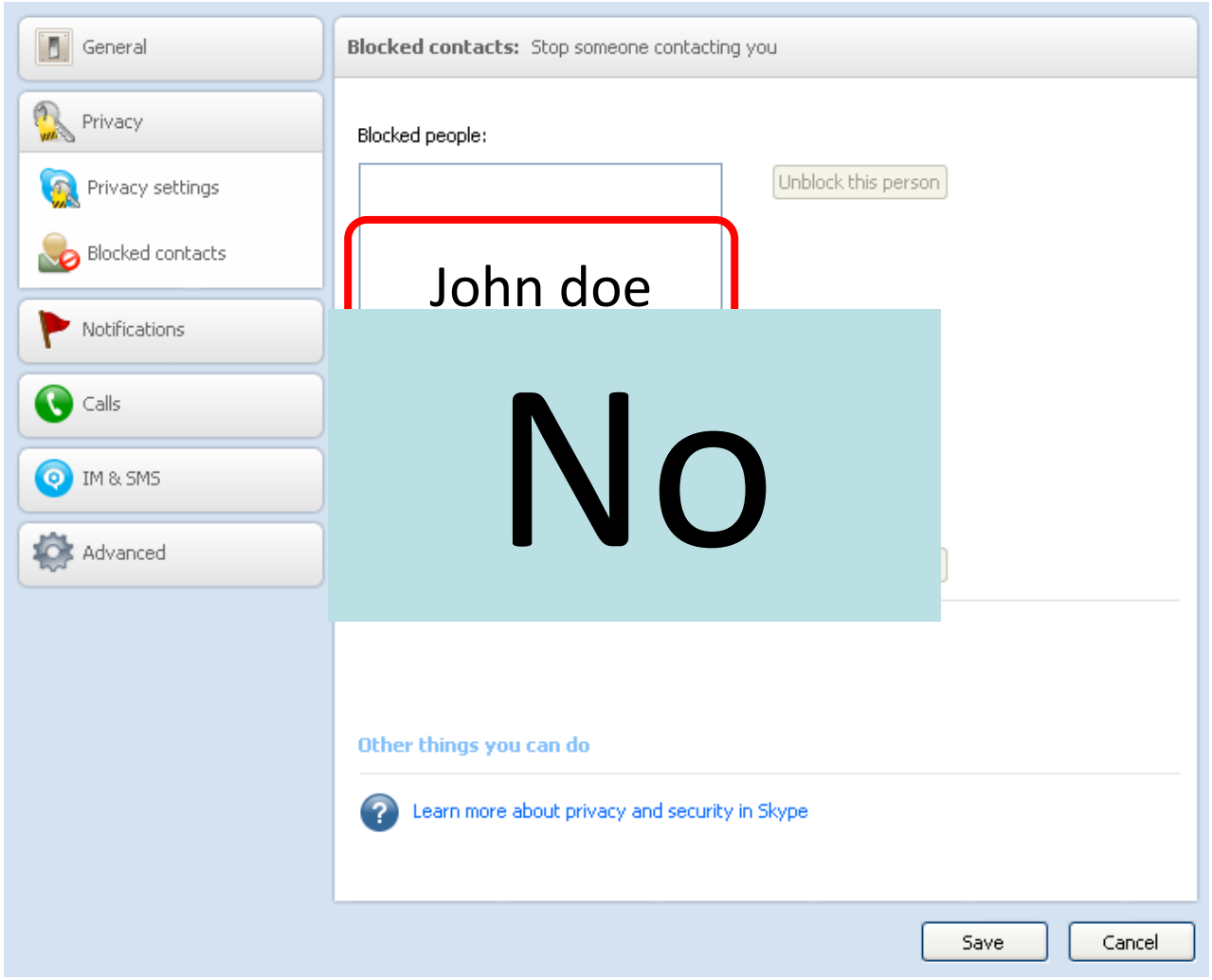
- We can with no dedicated infrastructure or ISP collaboration
  - Link an IP address to a social identity
  - Massively track mobility
  - Link BitTorrent downloads to a social identity
- Works for all Skype Users (560M)

Very hard to prevent such privacy attacks

Does Rejecting Calls From Strangers Protect Against Tracking?



Does Blocking Attacker's Account Protect Against Tracking?



Can We Protect Against These Attacks?

- ❑ Get rid of packet patterns
- ❑ Relay calls from strangers
- ❑ Relay calls from certain friends

We can only make tracking harder

Outline

- Privacy Foundations
- Privacy Attacks
  - Introduction to Privacy
  - Practical Attacks
    - Distributed systems
    - Web

UNIVERSITÉ CÔTE D'AZUR

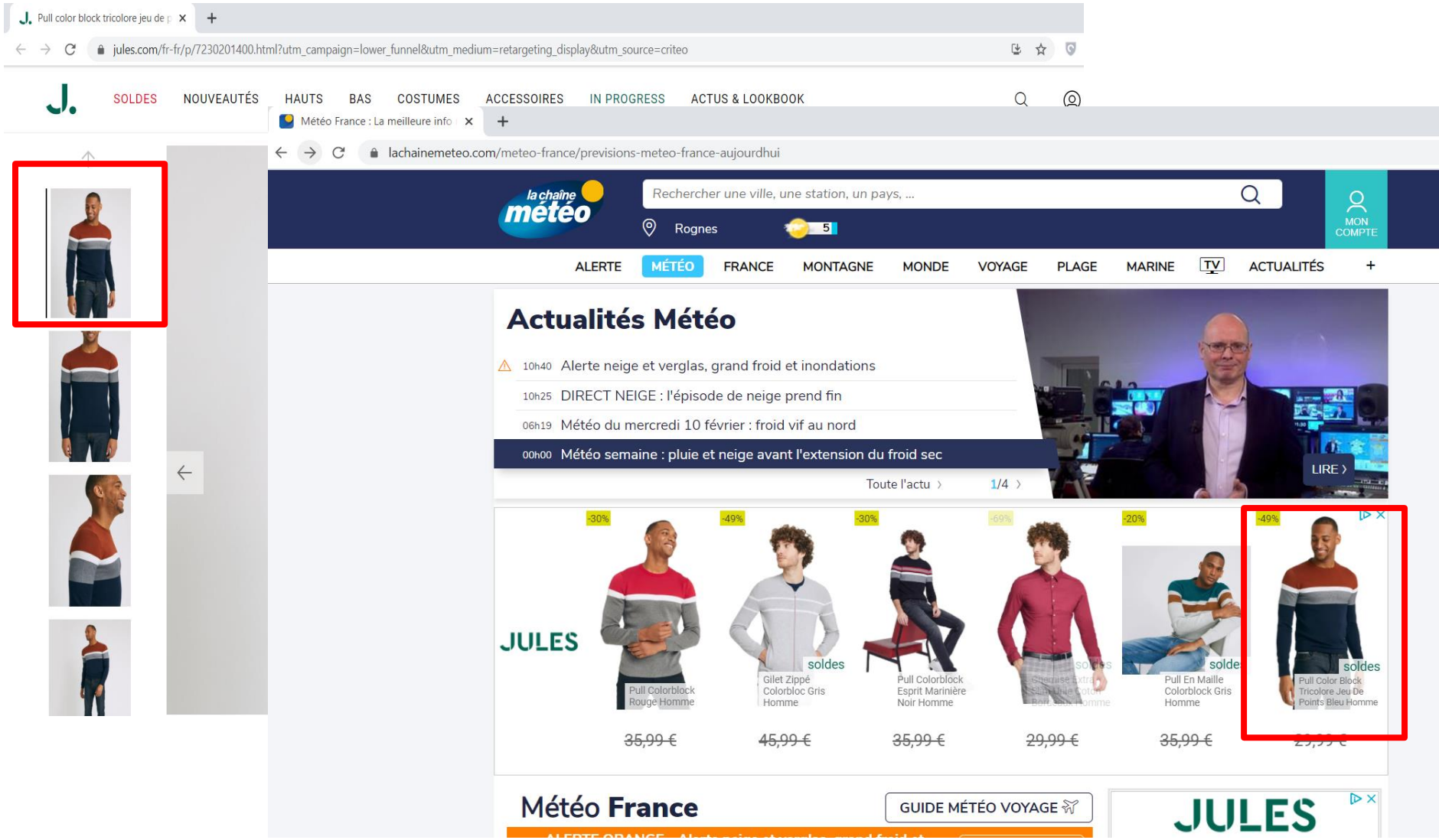
Inria
INVENTEURS DU MONDE NUMÉRIQUE

DESPERATE QUEST FOR PRIVACY!

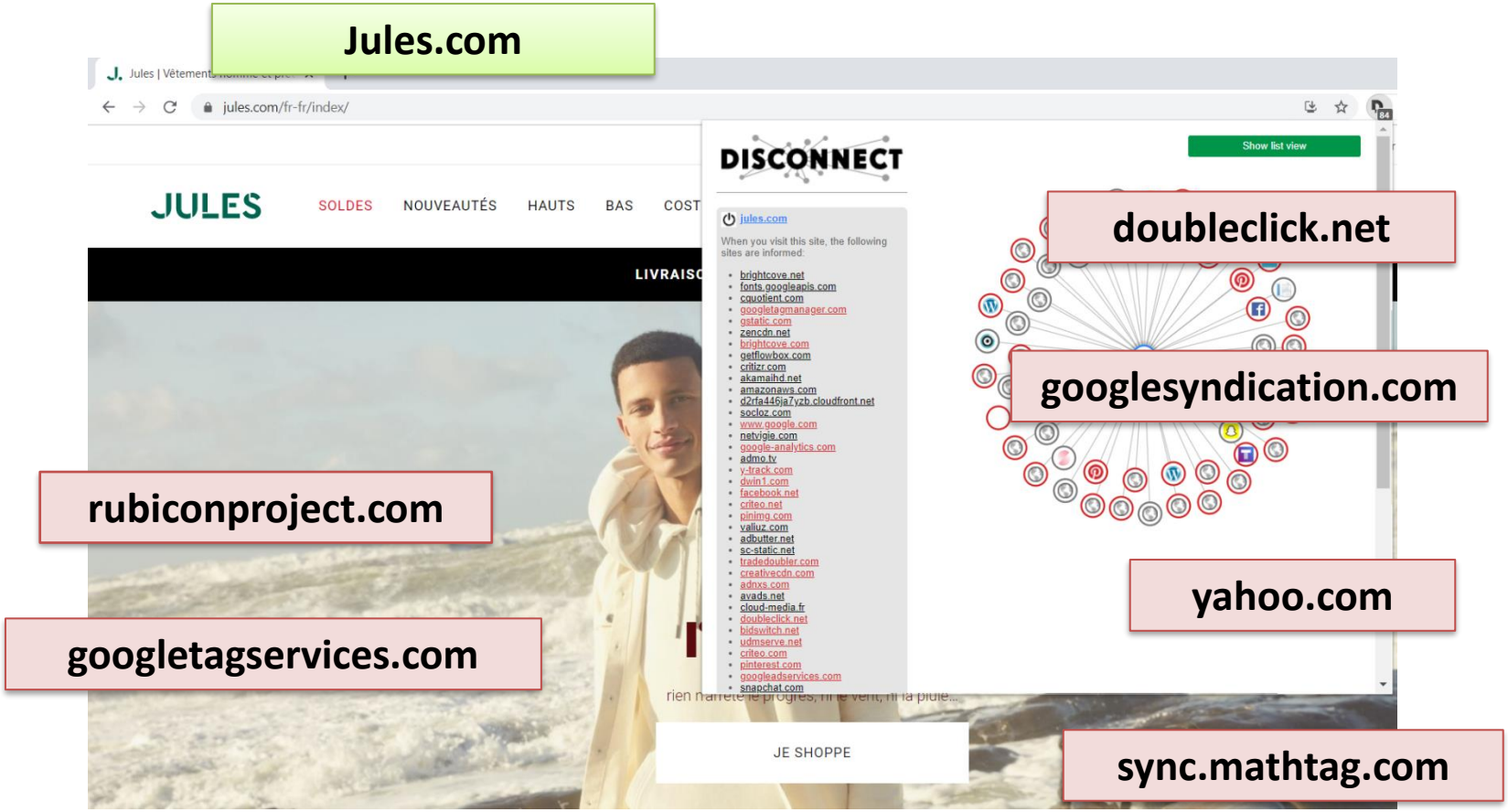
Slides by Imane Fouad

joint work with Nataliia Bielova and Arnaud Legout

Once upon a time…



HOW DOES WEB TRACKING WORK?



84 domains will know that you visited this website!


Third-Party Trackers
LIGHTYEAR
Third-Party Trackers Everywhere

Why Web Tracking is important?

- Collection of our data without our knowledge
  - on sensitive websites
  - collection of our browsing patterns, preferences, tastes, even mood…

- **Usage of our data!**
  - targeted advertisement
  - **manipulation**

Cambridge Analytica

"We exploited Facebook to harvest millions of people's profiles. And built models to exploit what we knew about them and target their inner demons. That was the basis the entire company was built on."

Christopher Wylie

18 March 2018

Facebook

Admiral to price car insurance based on Facebook posts

Insurer's algorithm analyses social media usage to identify safe drivers in unprecedented use of customer data

Graham Ruddick

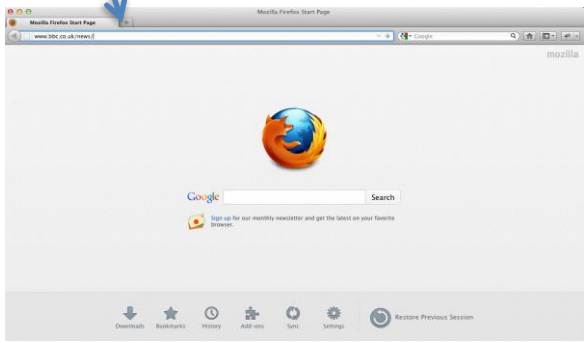
Wednesday 2 November 2016
00.01 GMT

Admiral says its firstcarquote initiative is aimed at first-time drivers or car owners. Photograph: Image Source/Rex Features

Cookies in HTTP header

www.bbc.co.uk/news/

Web browser



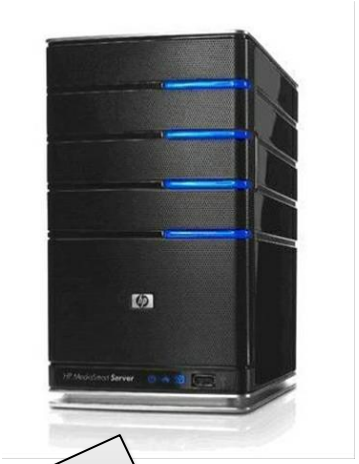
URL path: bbc.co.uk/news
Parameters
Method: GET
…

HTTP request

HTTP response

Status: 200 OK
Content: HTML page
Set-cookies: session-id=**2082787201l** & …

Web server



- Rendering pages
- Executing scripts/plugins (JavaScript)
- Launching new HTTP requests

- Mapping requests to apps
- Contacting DBs
- Constructing responses

Cookies in HTTP header

Web browser

Cookie Database

bbc.co.uk/news:
session-id=**2082787201l**

URL path: bbc.co.uk/news
Parameters
Method: GET
…

HTTP request

HTTP response

Status: 200 OK
Content: HTML page
Set-cookies: session-id=**2082787201l** & …

…

Web server

Cookies in HTTP header

Web browser

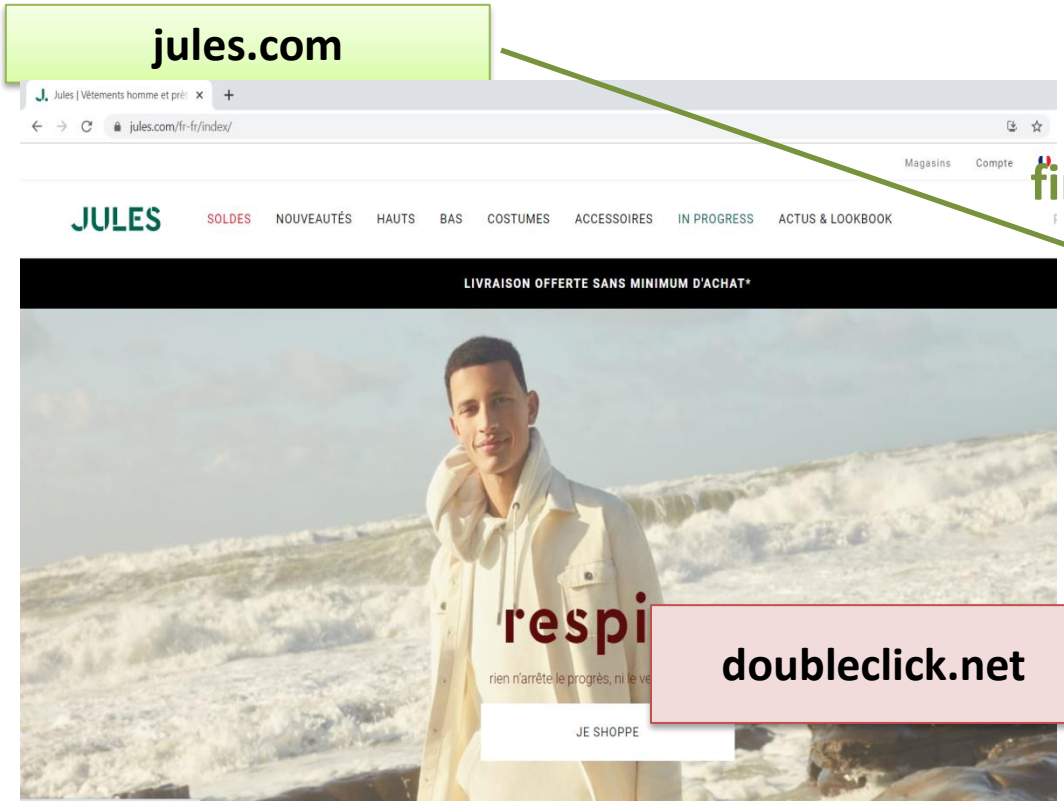
URL path: bbc.co.uk/news...
Method: GET
Cookies: session-id=**2082787201l** & …
…

HTTP request

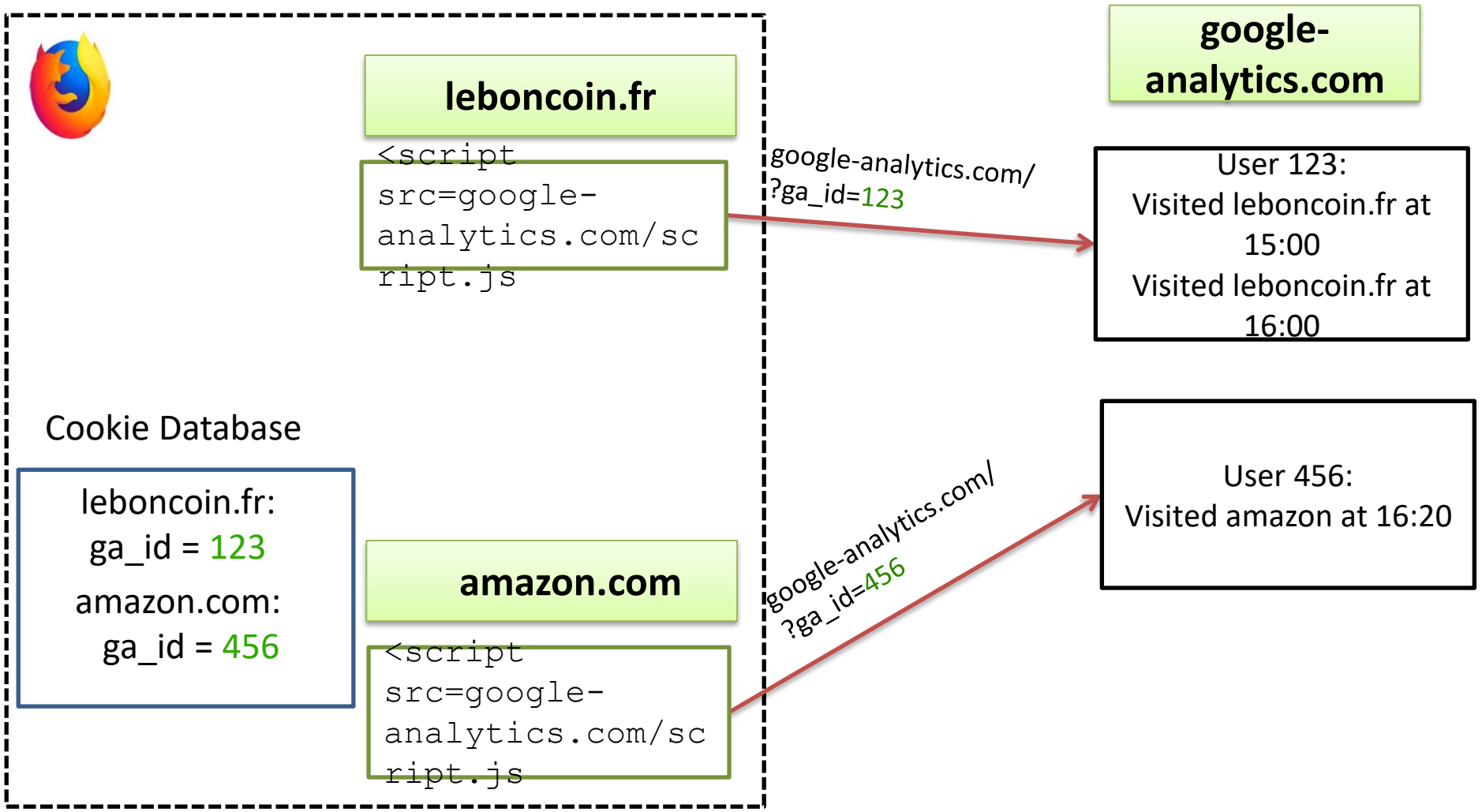
Web server

Cookie Database

bbc.co.uk/news:
session-
id=**2082787201l**


jules.com
Jules | Vêtements homme et pré
jules.com/fr-fr/index/
Magasins
Compte
JULES
SOLDES
NOUVEAUTÉS
HAUTS
BAS
COSTUMES
ACCESSOIRES
IN PROGRESS
ACTUS & LOOKBOOK
LIVRAISON OFFERTE SANS MINIMUM D'ACHAT*
respi
rien n'arrête le progrès, ni le vé
JE SHOPPE
doubleclick.net
first-party cookie
Cookie Database
Jules.com:
id = abcd
doubleclick.net:
id = 1234
third-party cookie

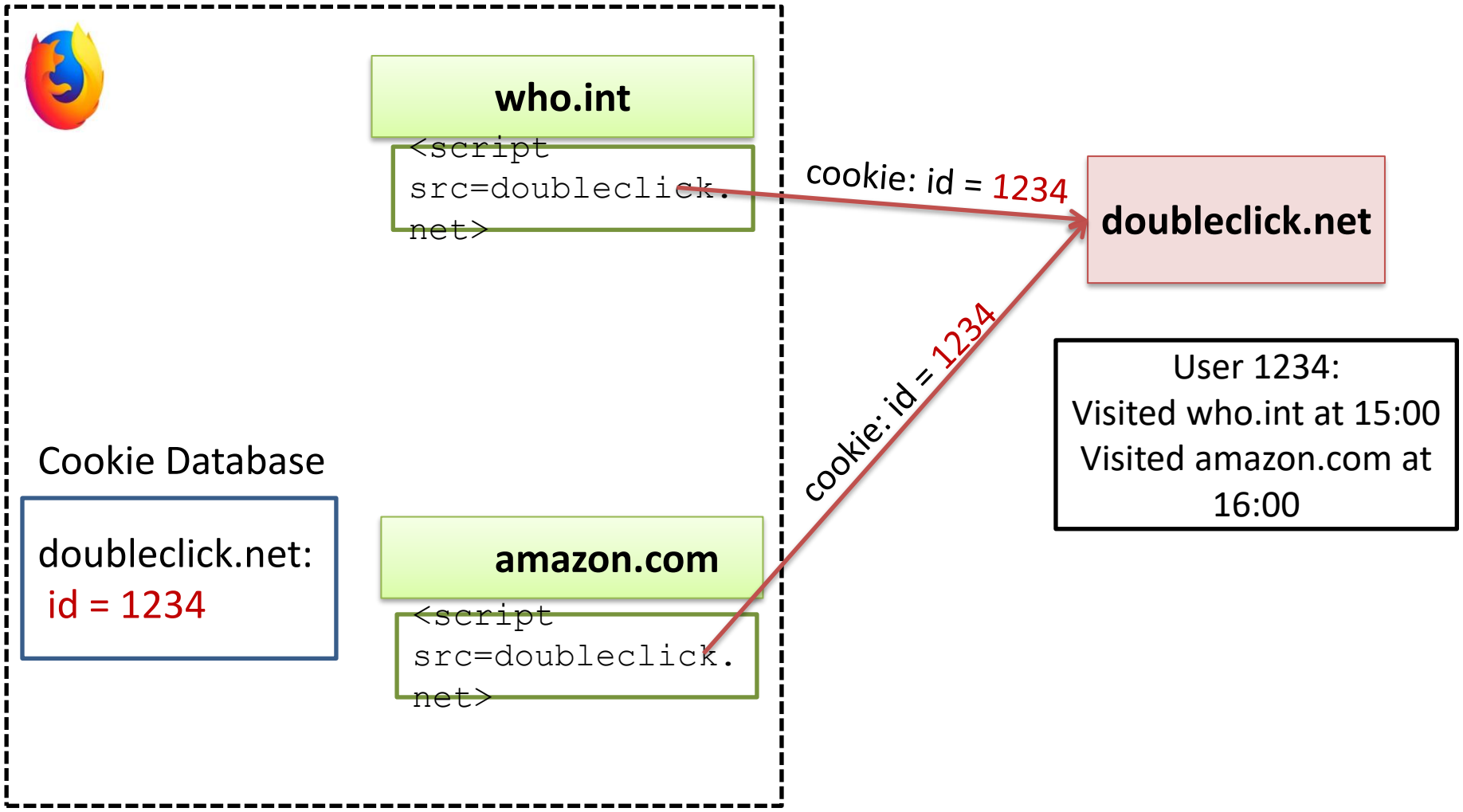
Analytics (Within-Site Tracking)



First-party cookies have more benefits

- Website owners can evaluate
  - website statistics
  - popularity of certain pages

Cross-site tracking



In summary

- Cross-site tracking
  - Based on tird party cookies
  - Used to track user cross websites

- Analytics (Within-Site Tracking)
  - Based on first party cookies
  - Used to track repeat visits to a site.

sciendo

Proceedings on Privacy Enhancing Technologies ..; .. (..):1–20

Imane fouad*, Nataliia Bielova, Arnaud Legout, and Natasa Sarafijanovic-Djukic

Missed by Filter Lists: Detecting Unknown Third-Party Trackers with Invisible Pixels

Abstract: Web tracking has been extensively studied over the last decade. To detect tracking, previous studies and user tools rely on filter lists. However, it has been shown that filter lists miss trackers. In this paper, we propose an alternative method to detect trackers inspired by analyzing behavior of invisible pixels. By crawling 84,658 webpages from 8,744 domains, we detect that third-party invisible pixels are widely deployed: they are present on more than 94.51% of domains and constitute 35.66% of all third-party images. We propose a fine-grained behavioral classification of tracking based on the analysis of invisible pixels. We use this classification to detect new categories of tracking and uncover new collaborations between domains on the full dataset of 4,216,454 third-party requests. We demonstrate that two popular methods to detect tracking, based on EasyList&EasyPrivacy and on Disconnect lists respectively miss 25.22% and 30.34% of the trackers that we detect. Moreover, we find that if we combine all three lists, 379,245 requests originated from 8,744 domains still track users on 68.70% of websites.

Keywords: online tracking; ad-blocker; cookie synching; invisible pixels

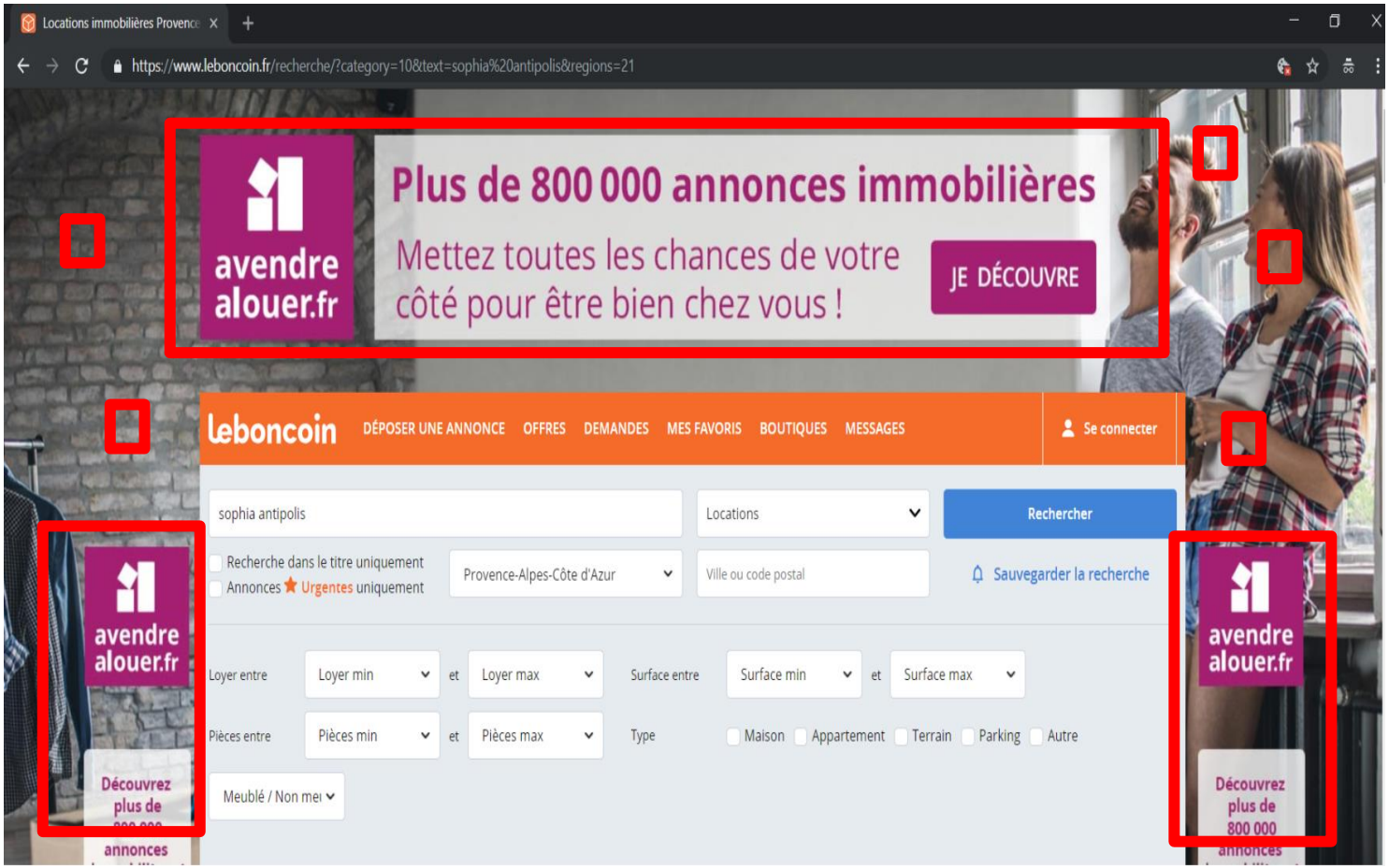
DOI Editor to enter DOI

In the last decade, numerous studies measured prevalence of third-party trackers on the Web [2, 11, 12, 24, 31–33, 37, 43, 49]. Web Tracking is often considered in the context of targeted behavioral advertising, but it's not limited to ads. Third-party tracking has become deeply integrated into the Web contents that owners include in their websites.

But what makes a tracker? How to recognize that a third-party request is performing tracking? To detect trackers, the research community applied a variety of methodologies. The most known Web tracking technique is based on *cookies*, but only some cookies contain unique identifiers and hence are capable of tracking the users. Some studies detect trackers by analysing cookie storage, and third-party requests and responses that set or send cookies [31, 43], while other works measured the mere presence of third-party cookies [32, 33]. To measure *cookie syncing*, researchers applied various heuristics to filter cookies with unique identifiers [1, 24, 25]. However, this approach has never been applied to detect tracking at large scale. Overall, previous works provide different methods to identify third-party requests that are responsible for tracking [43, 49]. Detection of identifier cookies and analysing behaviors of third-party domains is a complex task. Therefore,

Imane Fouad, Nataliia Bielova, Arnaud Legout, Natasa Sarafijanovic-Djukic. **Missed by Filter Lists: Detecting Unknown Third-Party Trackers with Invisible Pixels.** In *Proc. of Privacy Enhancing Technologies Symposium (PETS 2020)*, July 14--18, 2020, Montréal, Canada.

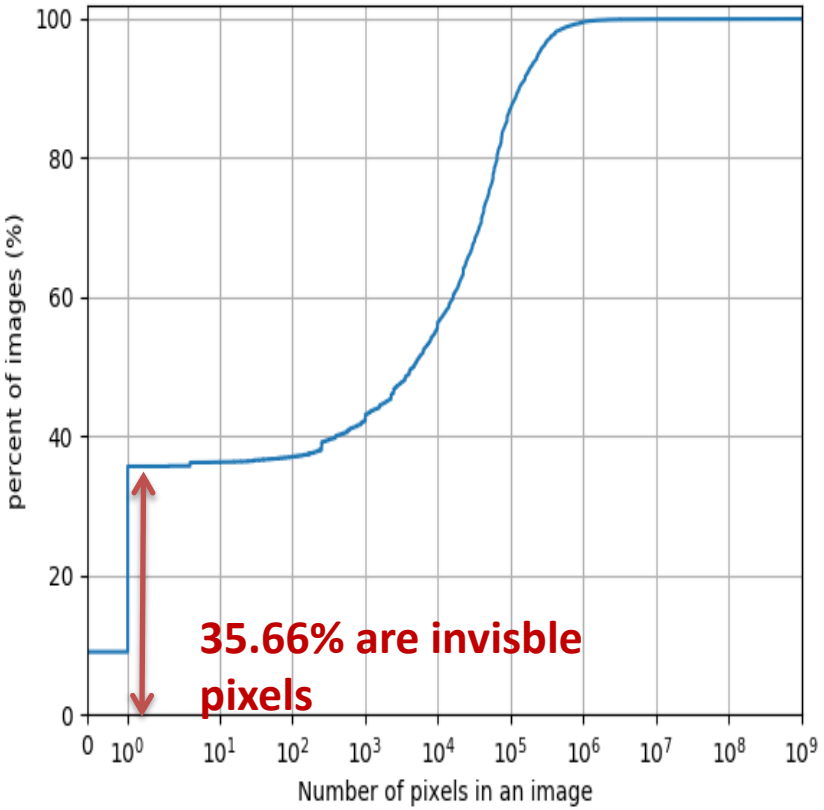
Invisible pixels



Invisible pixels are perfect suspects for tracking

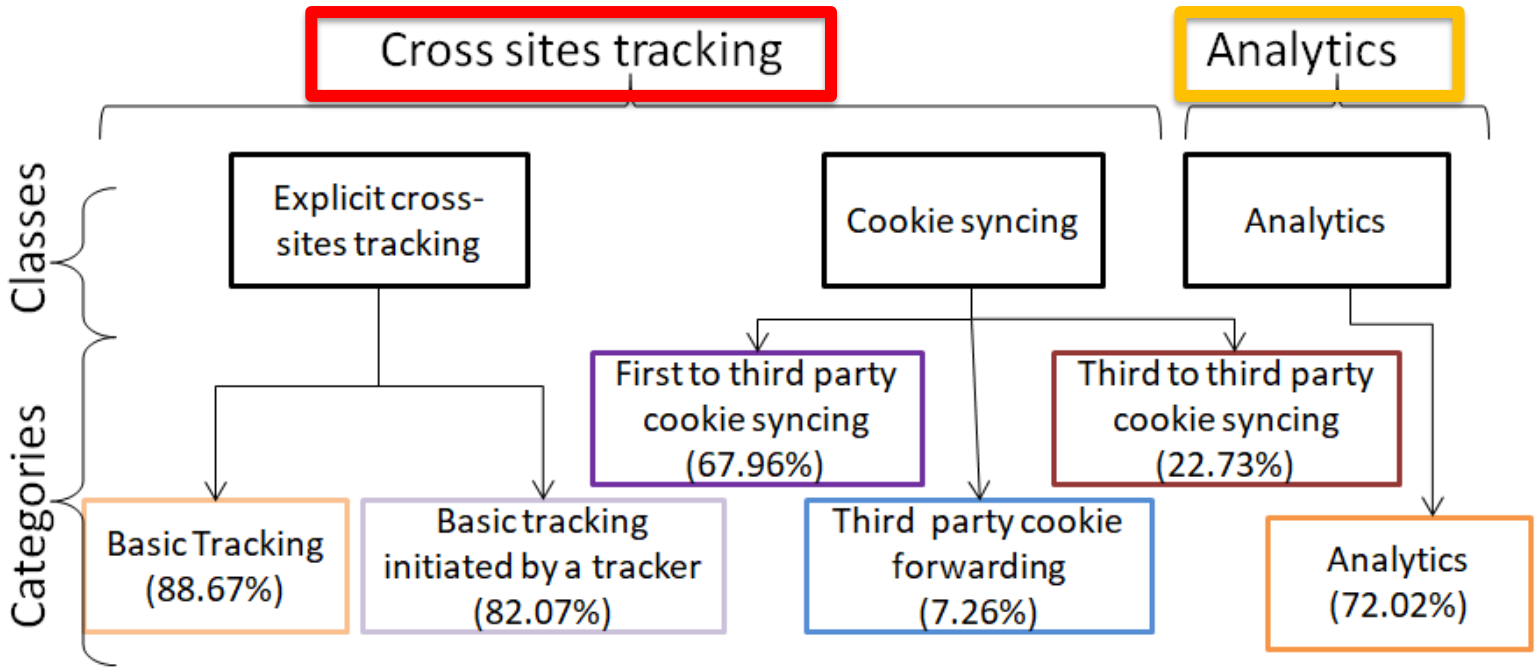
Data collection

- Top 8,744 domains
- 84,658 pages
- For each domain:
  - Homepage
  - 10 first links
- February 2019
- 2,297,716 images



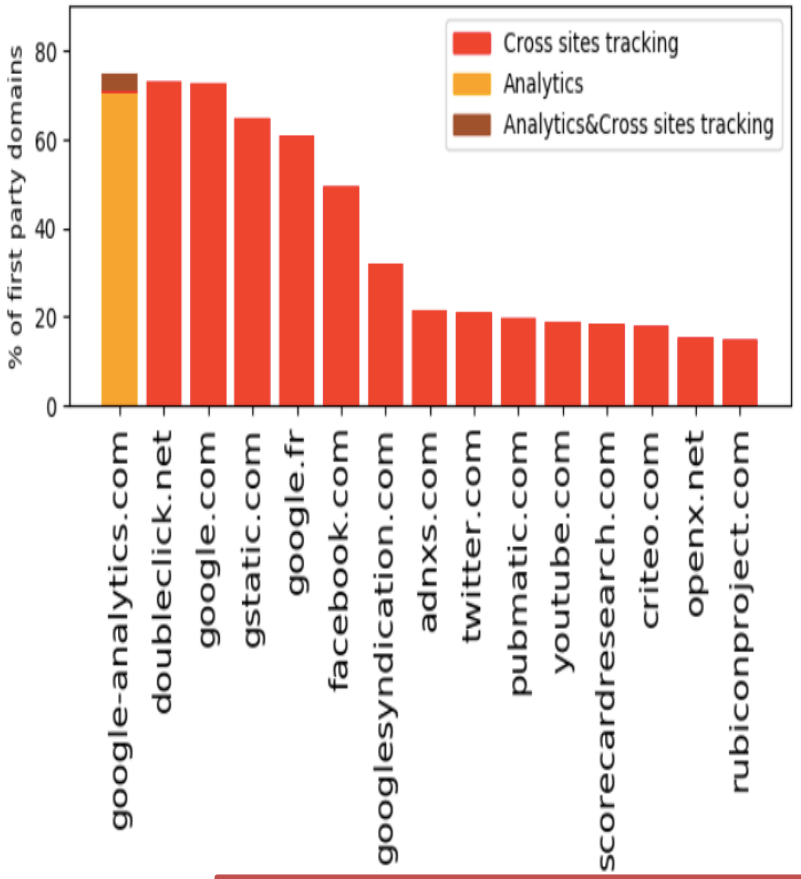
Invisible pixels are widely present on the web.

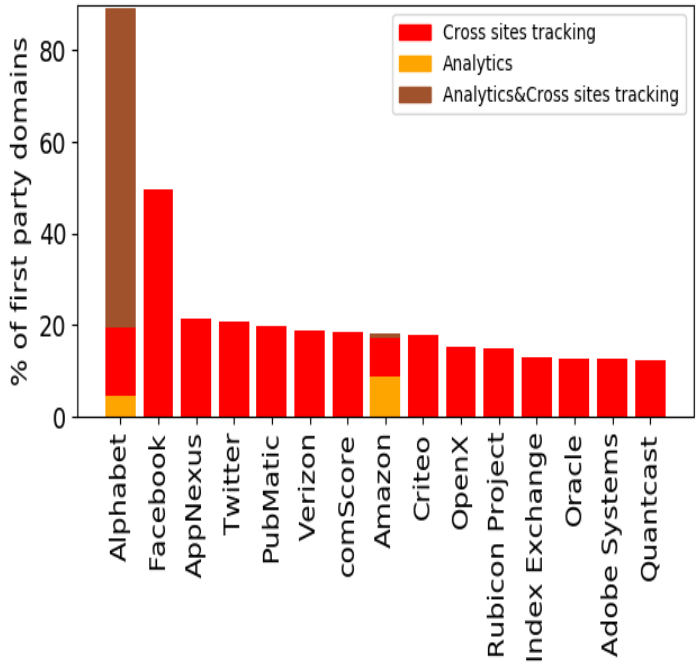
Classification: six cookie-based tracking behaviors



At least one type of tracking found on 92% of domains!

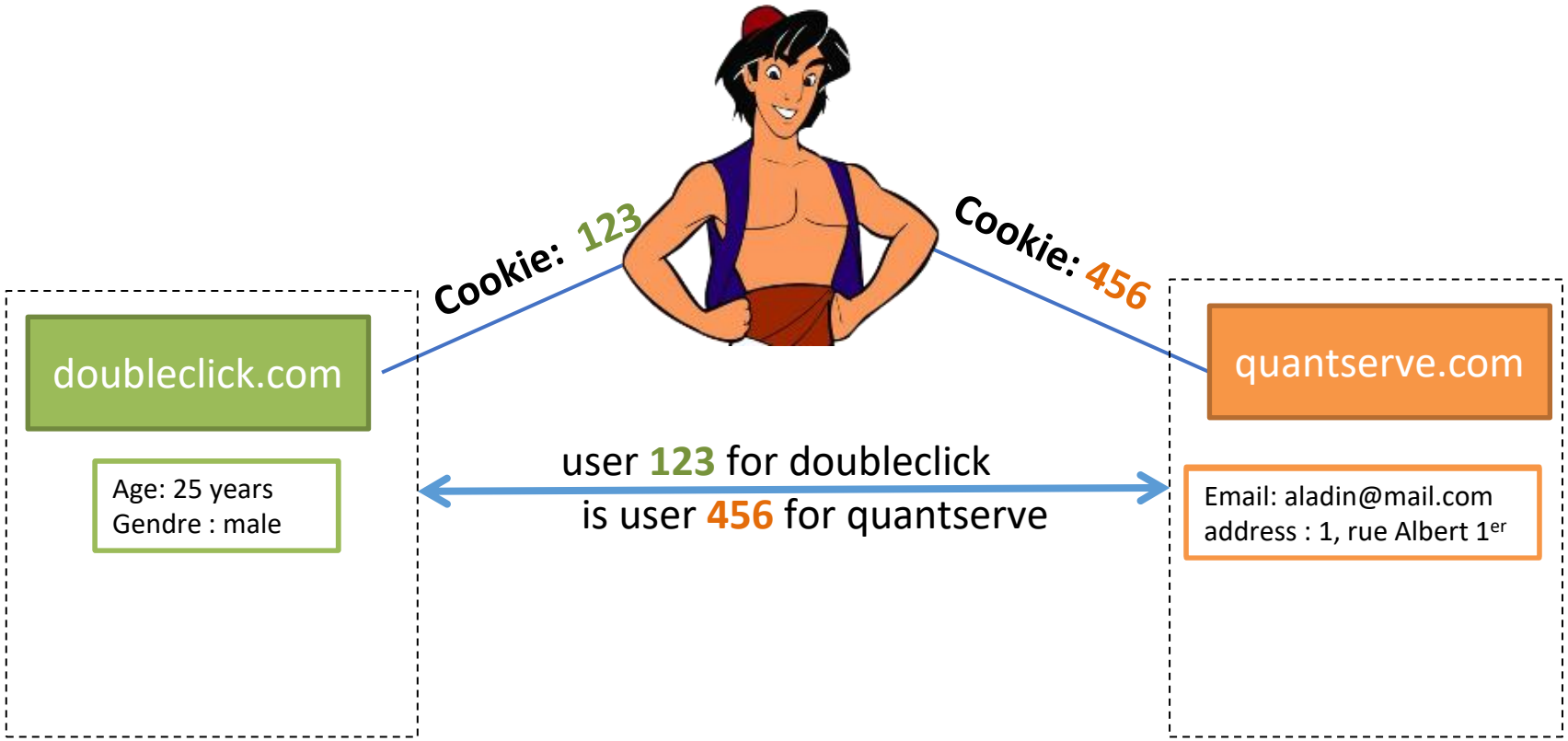
Domains behavior



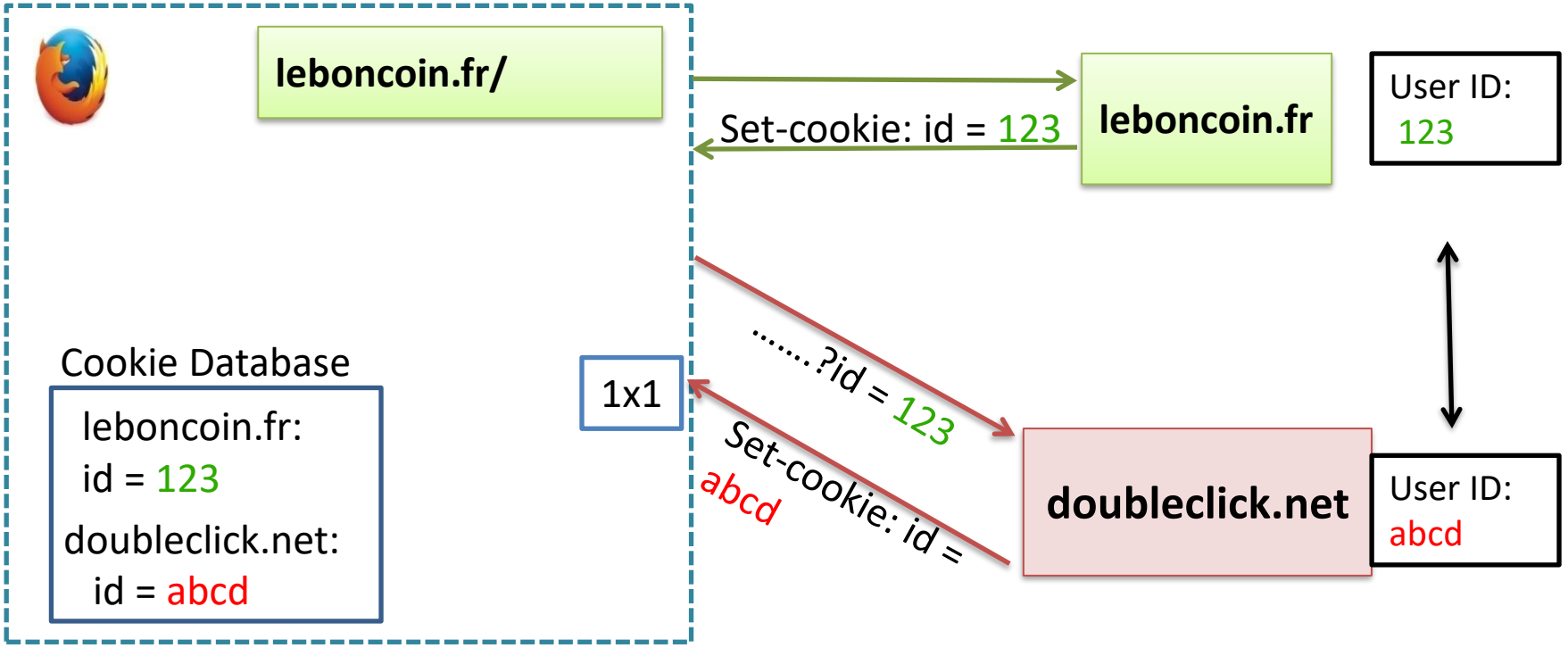


Some third parties act both as trackers and analytics services.

Cookie Syncing



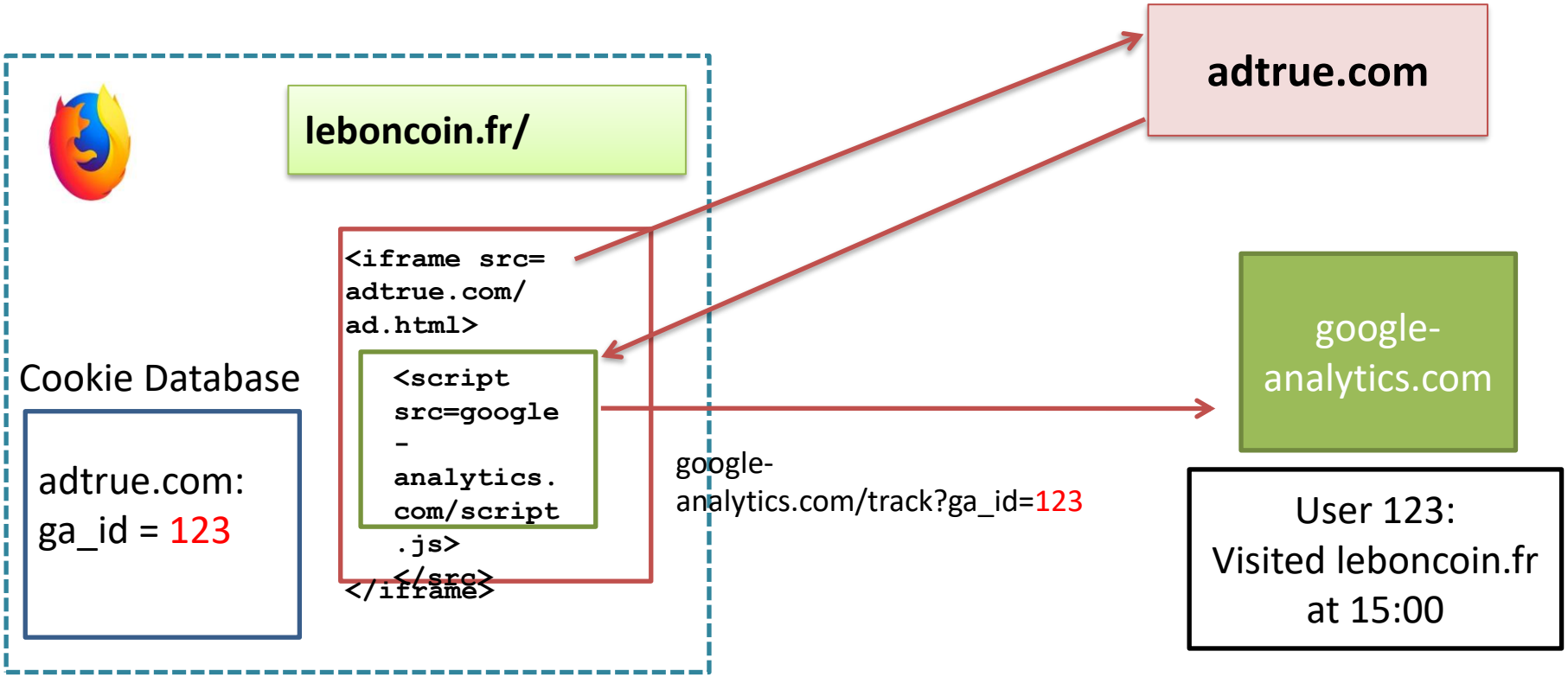
First to third party cookie syncing (67.96%)



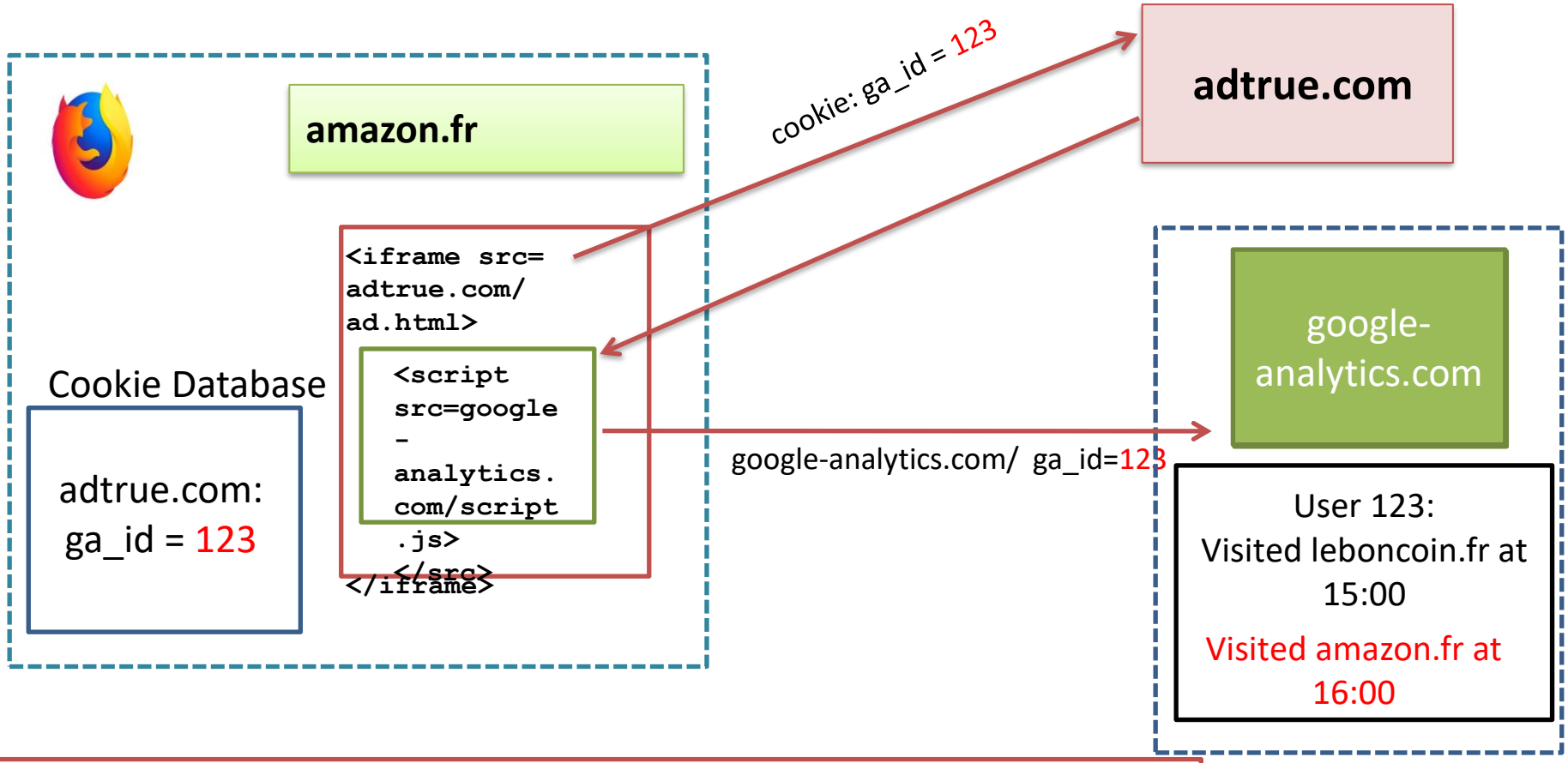
Within-site profile is shared with a third party

Analytics services enable third-party tracking because of inclusion

Third party cookie forwarding (7.26 %)



Third party cookie forwarding (7.26 %)



Analytics service becomes cross-site tracker

Are Filter Lists Effective at Detecting Trackers?

EasyList & EasyPrivacy

- Research community: **16 papers**
- Adblock: **10 000 000+ users**
- Adblock Plus: **10 000 000+ users**
- uBlock Origin: **10 000 000+ users**

Disconnect

- Firefox tracking protection
- Disconnect: **600 000+ users**

Do browser extensions block all trackers?

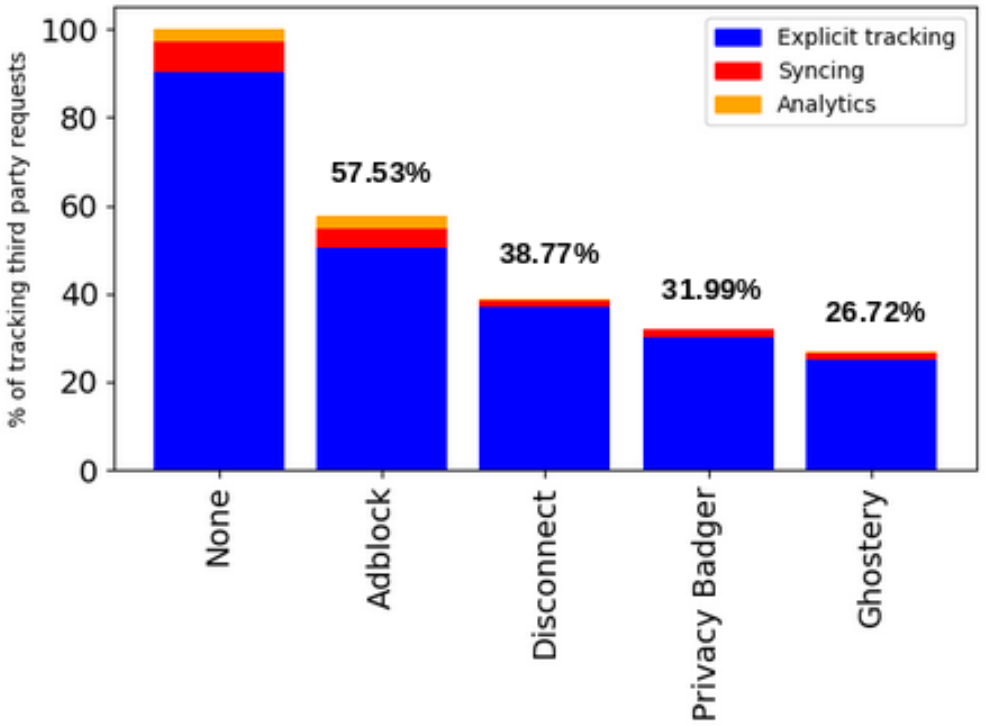
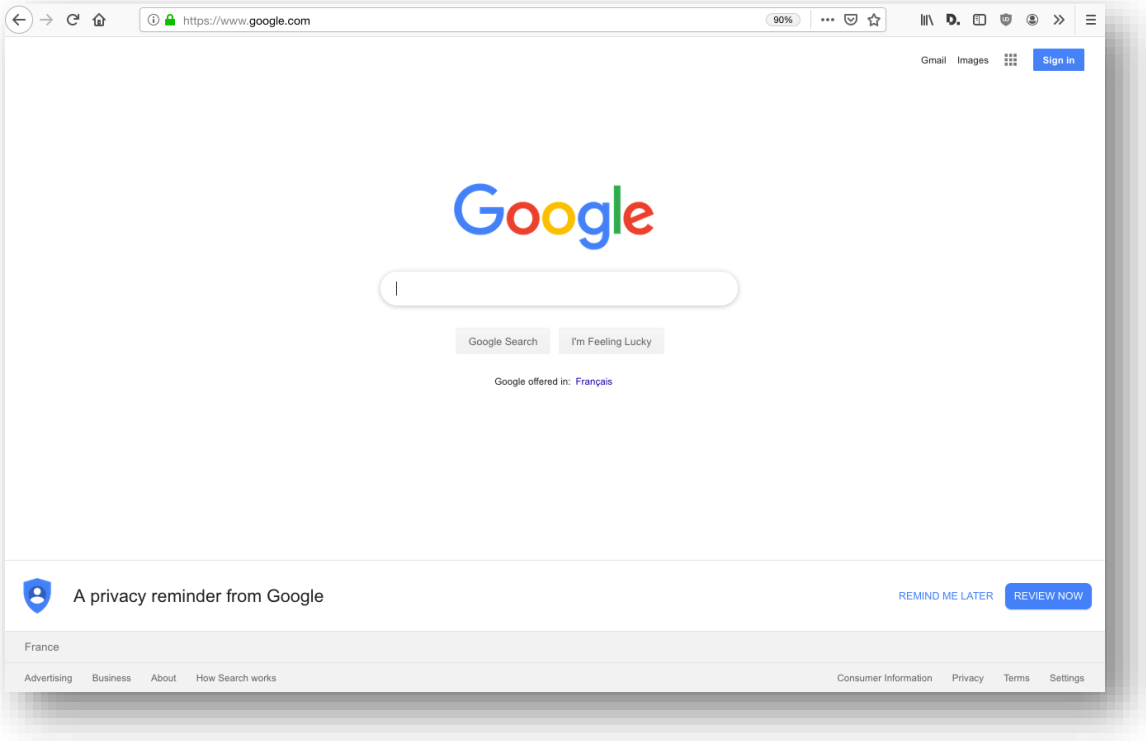


Fig. 12. Third party requests allowed by privacy protecting browser extensions out of 4,519,975 tracking requests.

Why filter lists miss trackers?

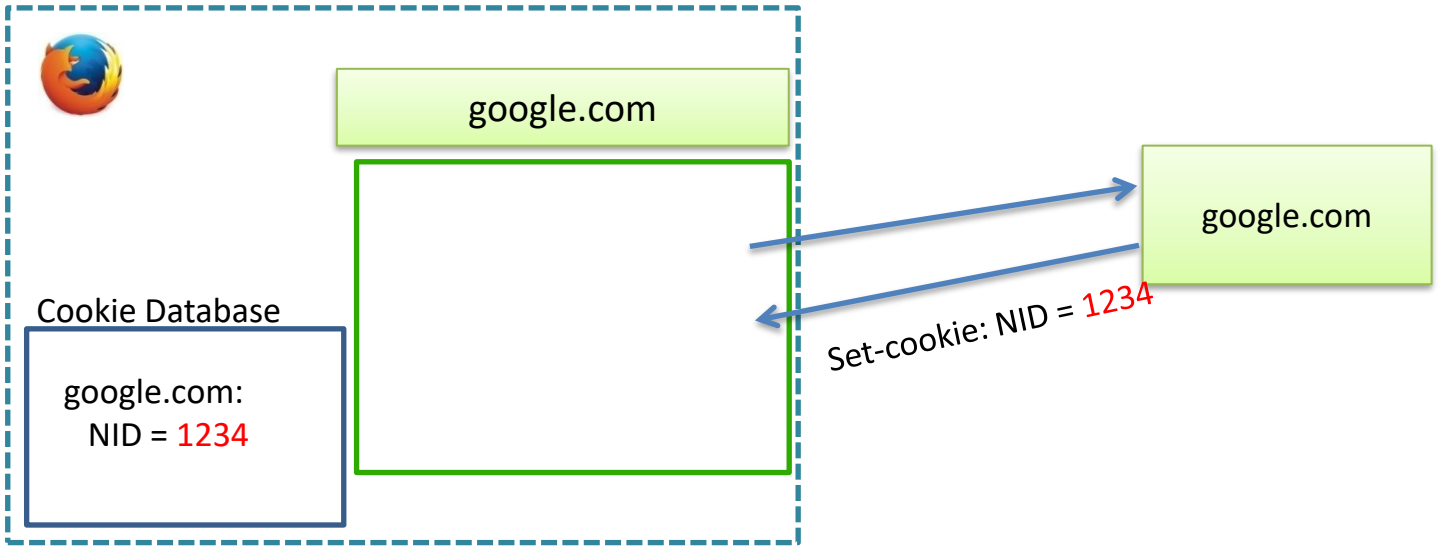
1 - Tracking Enabled by a First-Party Cookie:

User visits google.com, or just opens her browser



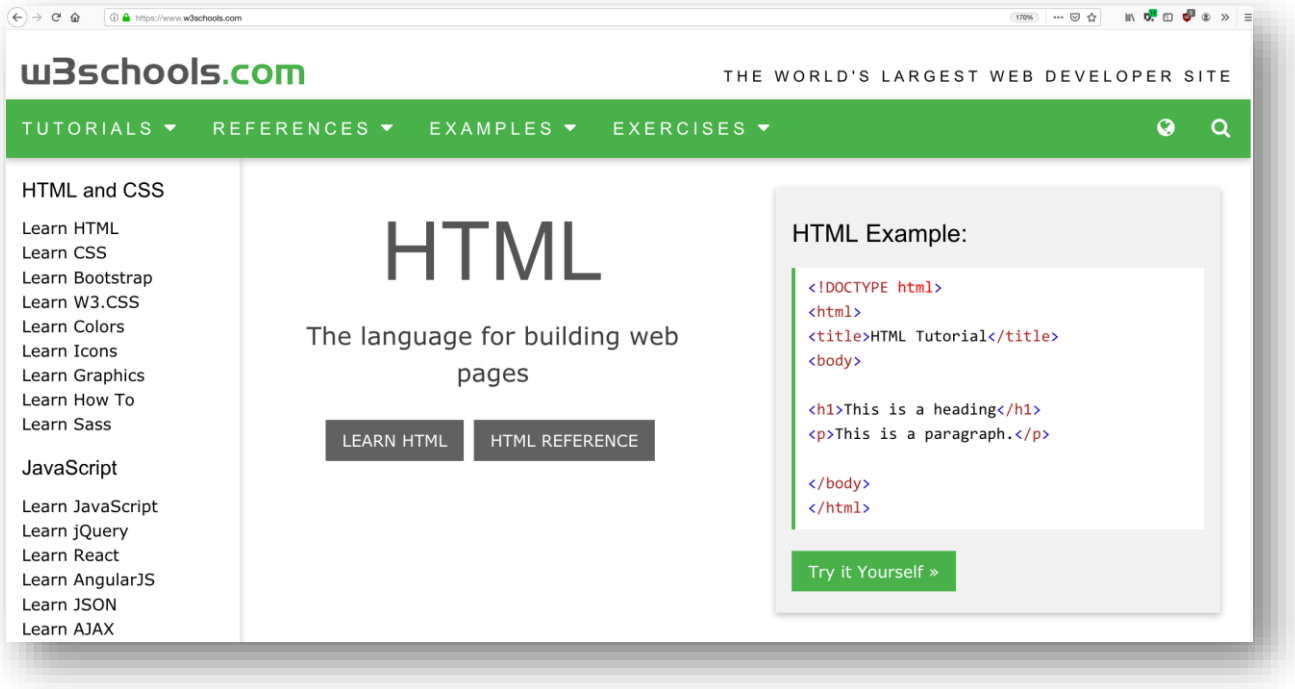
1 - Tracking Enabled by a First-Party Cookie:

Google.com sets a first party cookie in the user's browser



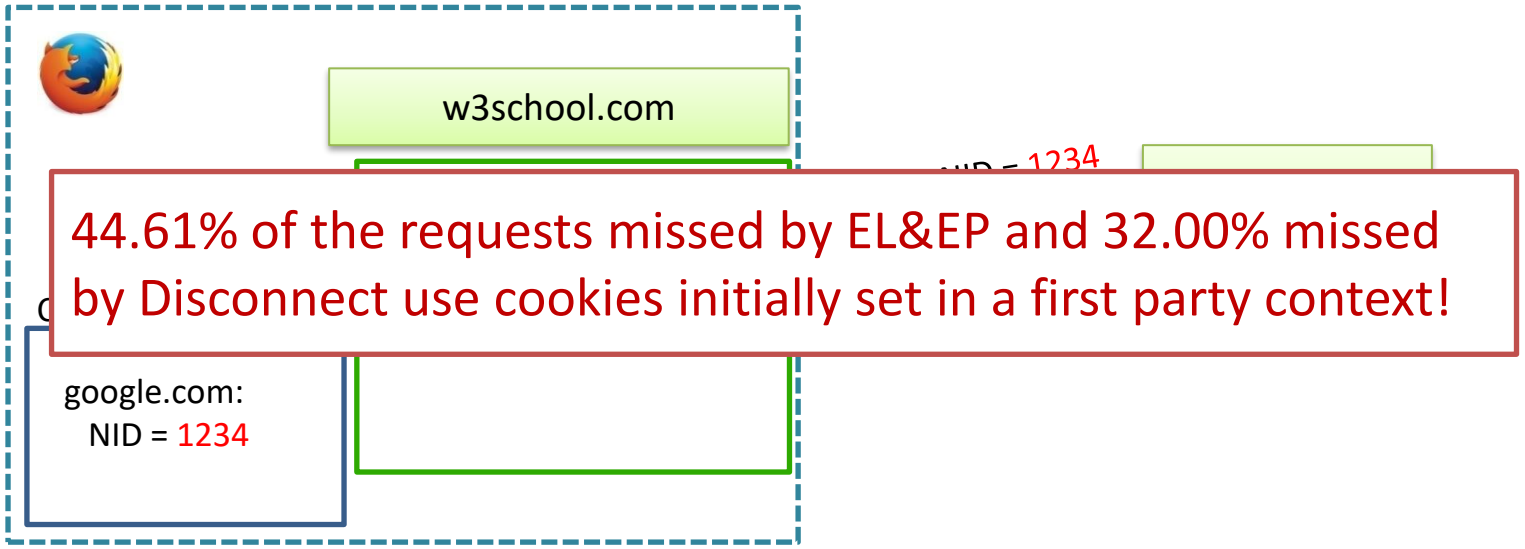
1 - Tracking enabled by a First-Party Cookie:

User visits w3schools.com

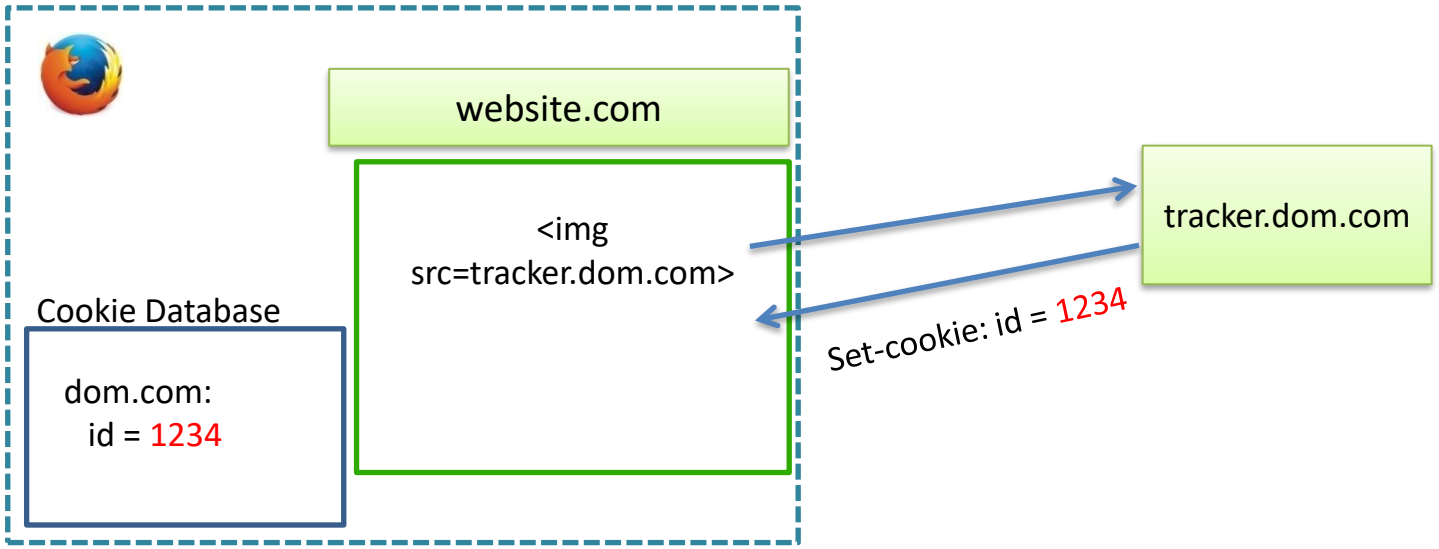


1 - Tracking Enabled by a First-Party Cookie:

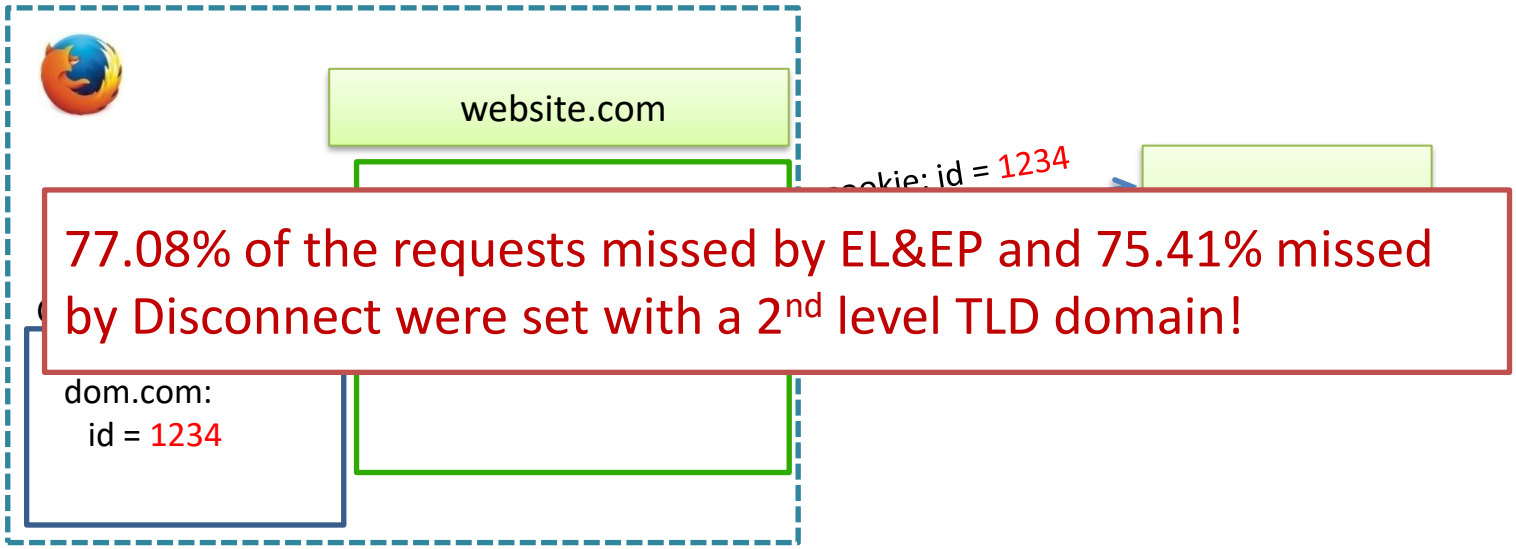
The tracking cookies is automatically sent to cse.google.com.



2 - Tracking enabled by large scope cookies.



2 - Tracking enabled by large scope cookies.



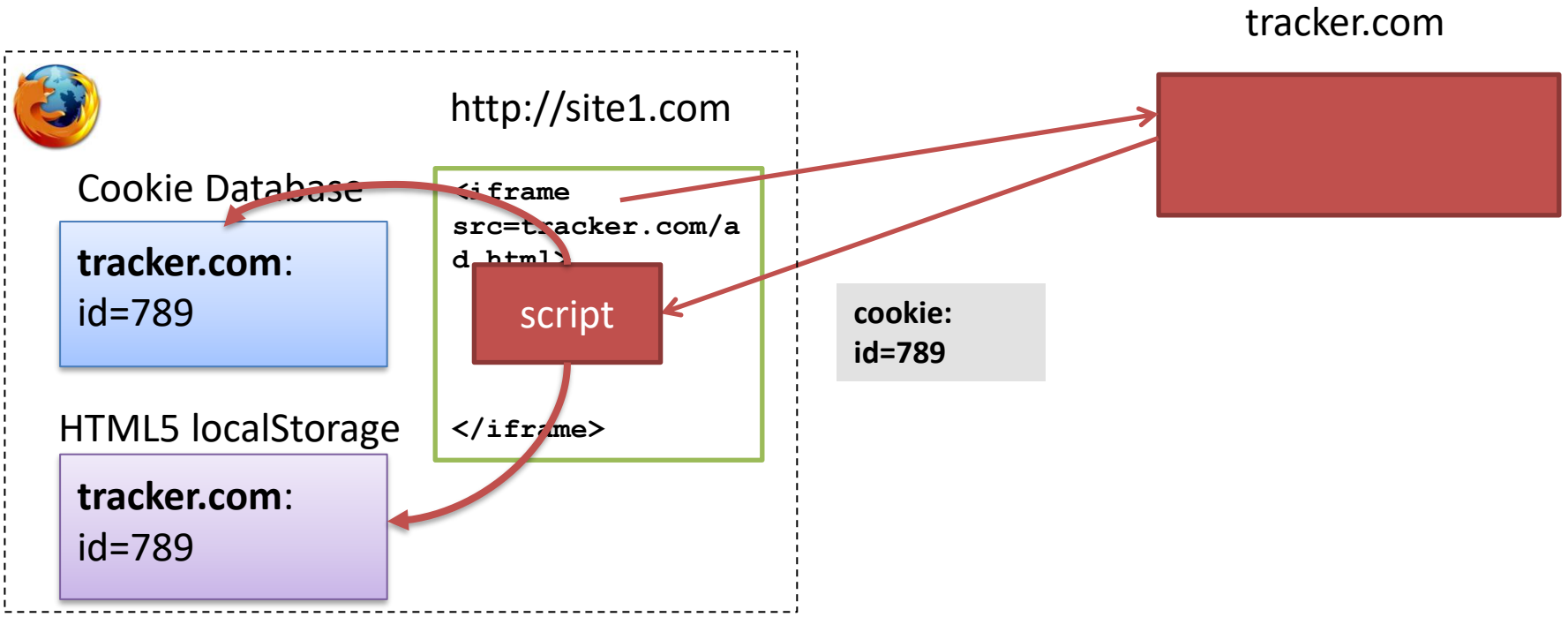
Imane Fouad, Cristiana Santos, Arnaud Legout, Nataliia Bielova. **Did I delete my cookies? Cookies respawning with browser fingerprinting**. *Technical Report (hal-03218403, version 1 - 5 May 2021)*, INRIA, Sophia Antipolis, May 2021.

Wish 1: I want to remove all cookies

Cookie respawning

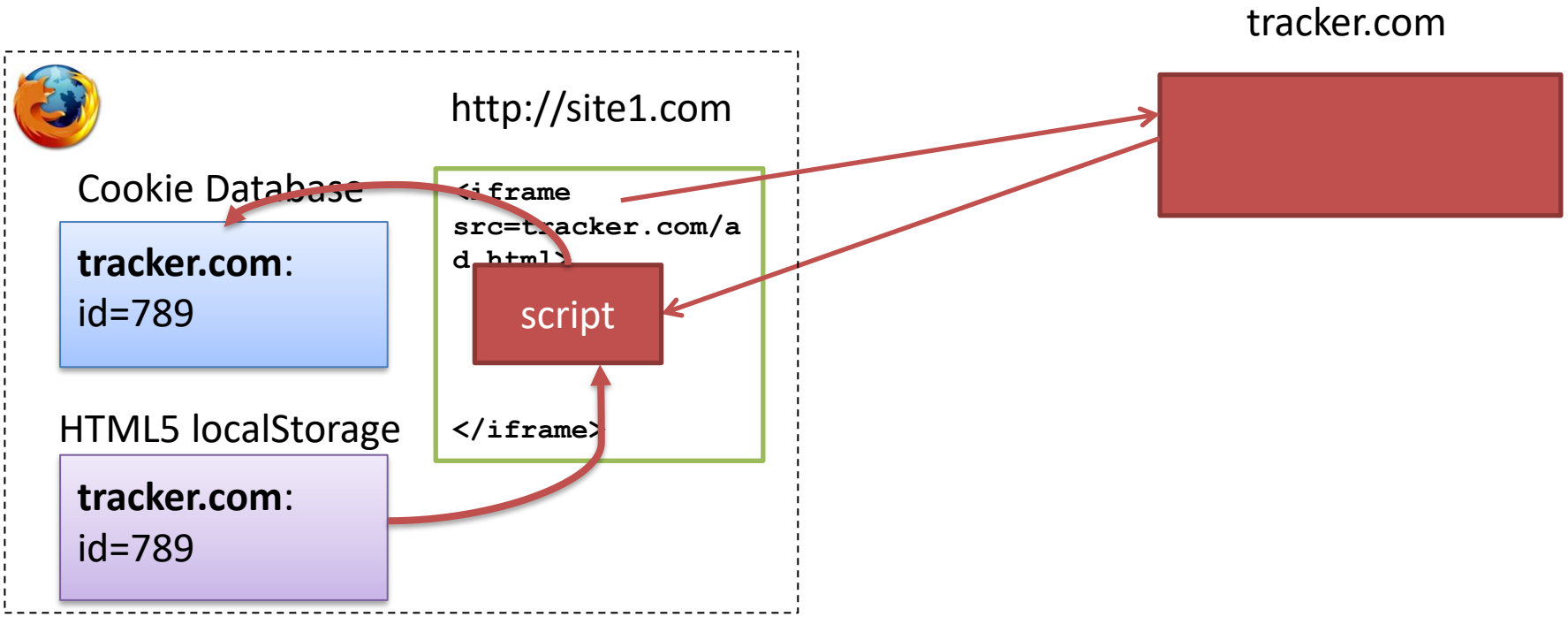
- Cookies **can respawn** even if the user has deleted them

Respawning via HTML5 localStorage



Respawning via HTML5 localStorage

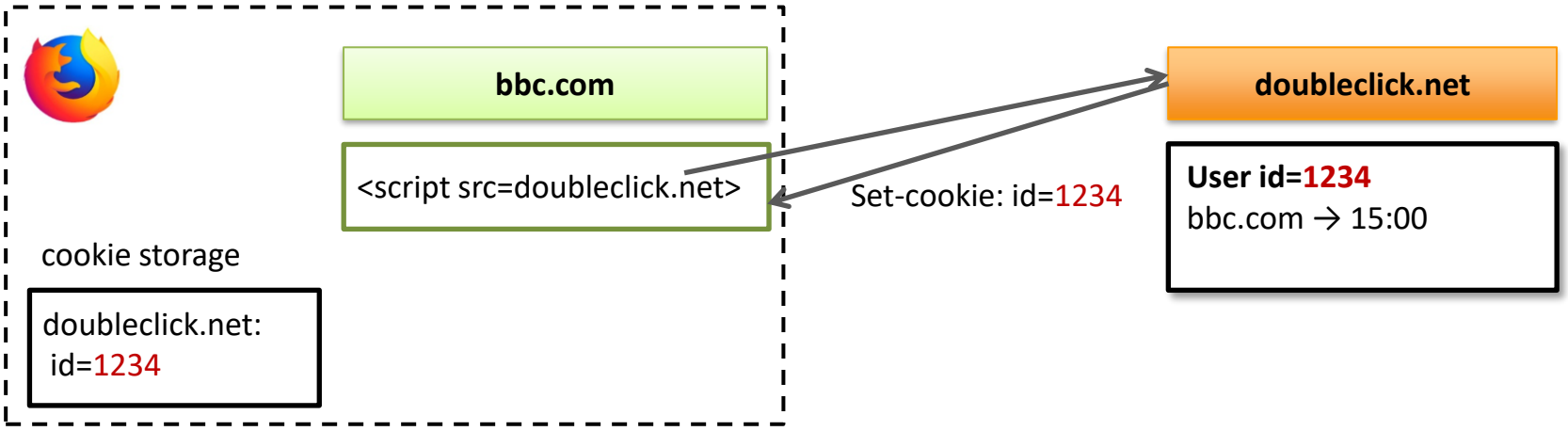
User deletes all the cookies!



Wish 2: I want to clean all browser storage

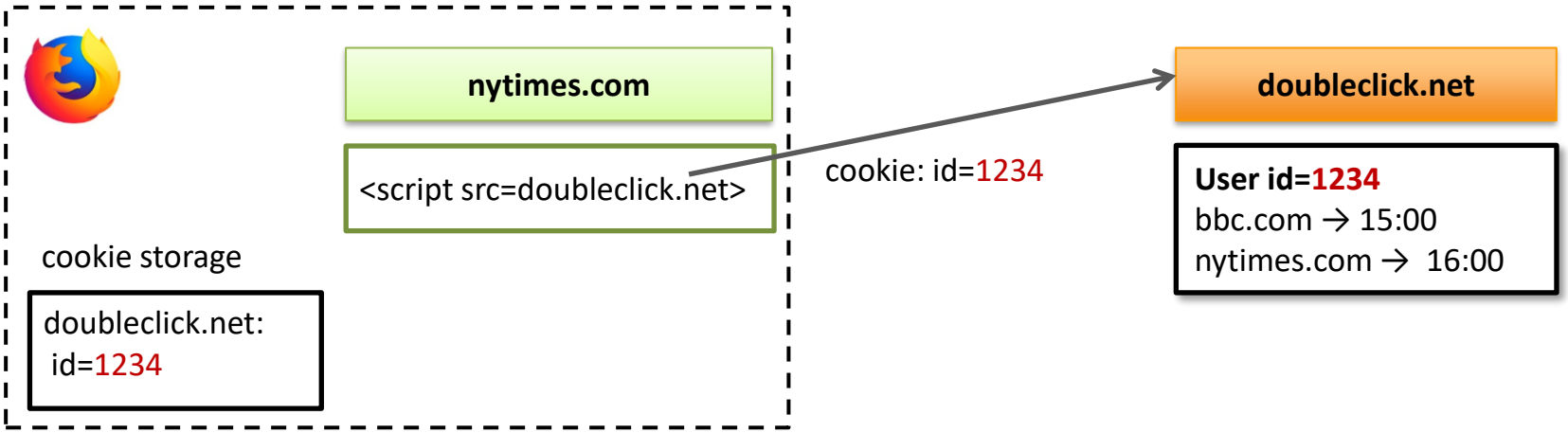
1- Statefull tracking

- **Analytics**: Based on first party cookies, it is used to track repeat visits to a site
- **Cross-site tracking:** Based on third party cookies, it is used to track user cross websites



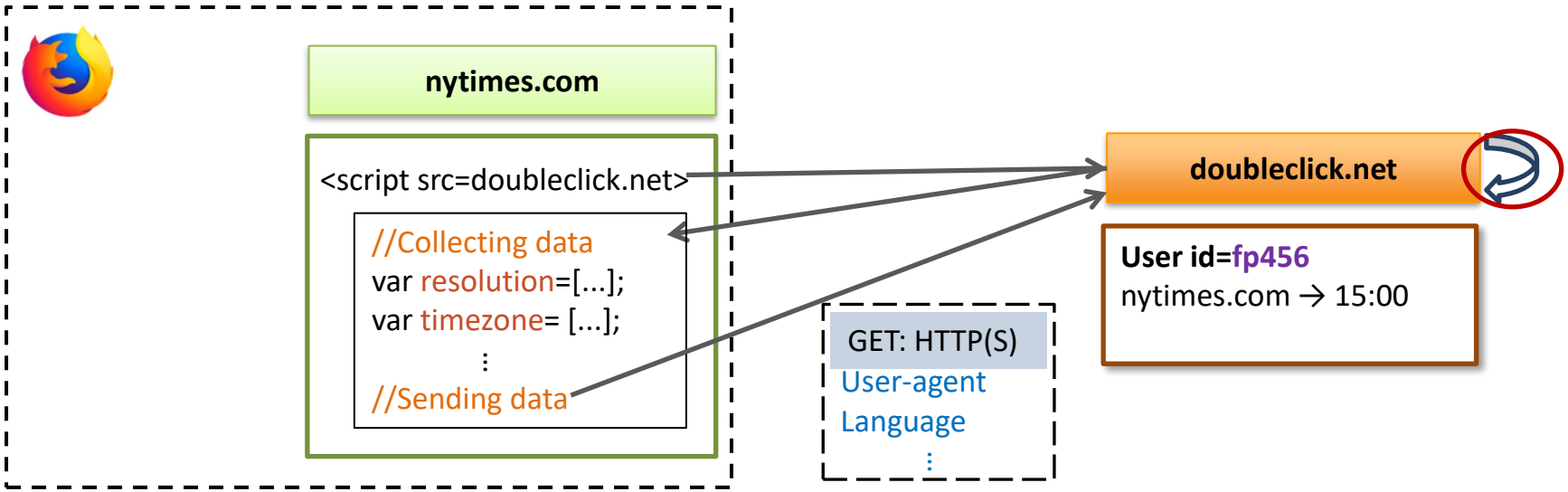
Cross-site tracking enabled by statefull tracking

- **Analytics**: Based on first party cookies, it is used to track repeat visits to a site
- **Cross-site tracking:** Based on third party cookies, it is used to track user cross websites



2- Stateless tracking

Active and passive features are used for fingerprinting



Stateful tracking

(+) Stable way to track a web user until she cleans cookies and other browser storages.

(-) Requires storage.

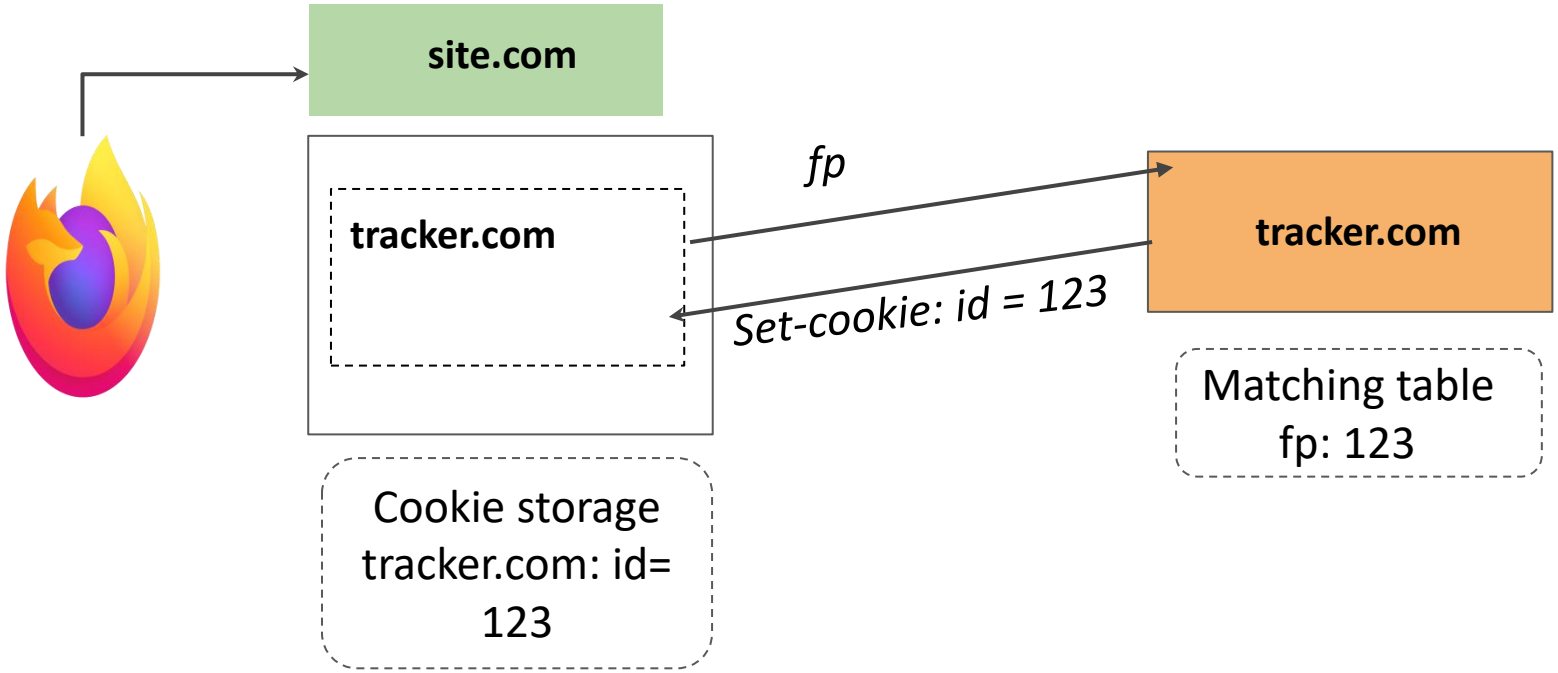
Stateless tracking

(+) Does not require any storage and can't be easily stopped by the user.

(-) Not stable over time

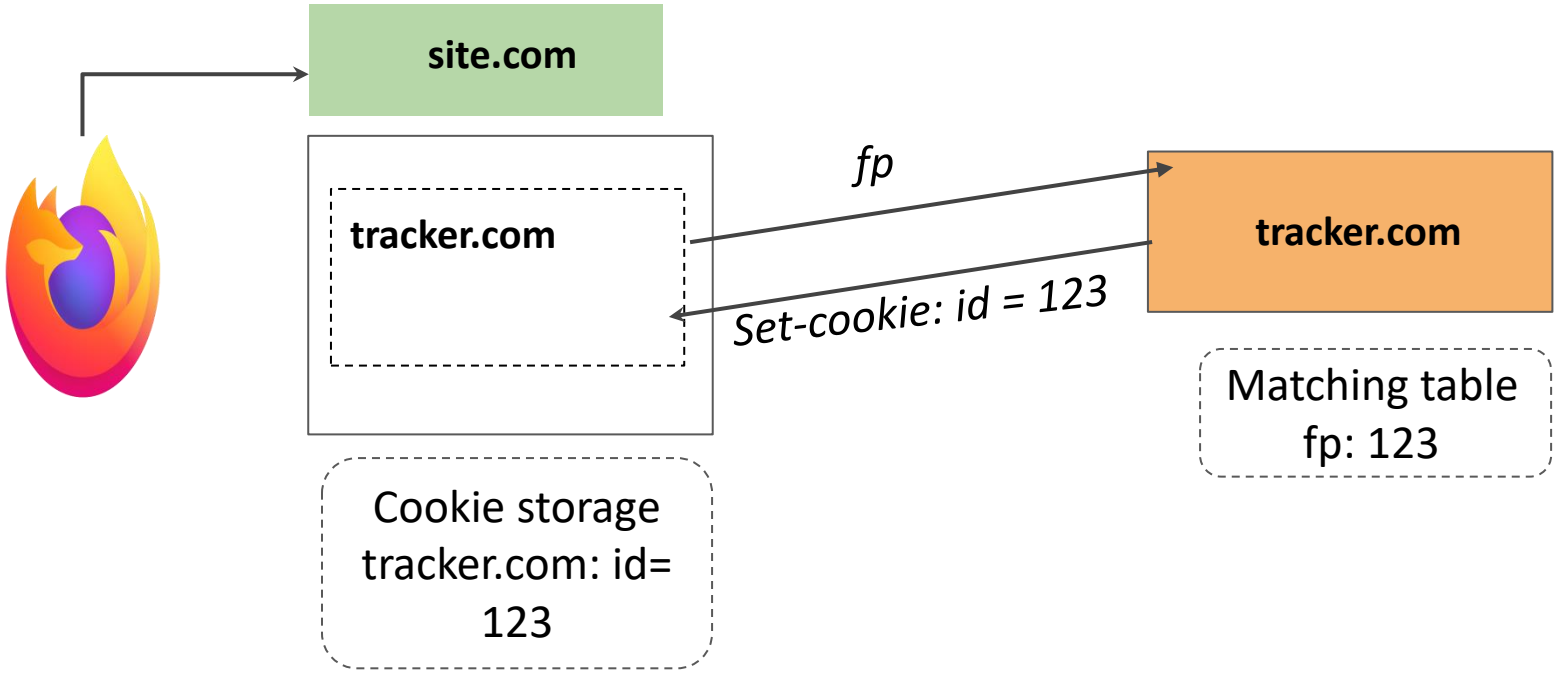
Cookie respawning via browser fingerprinting

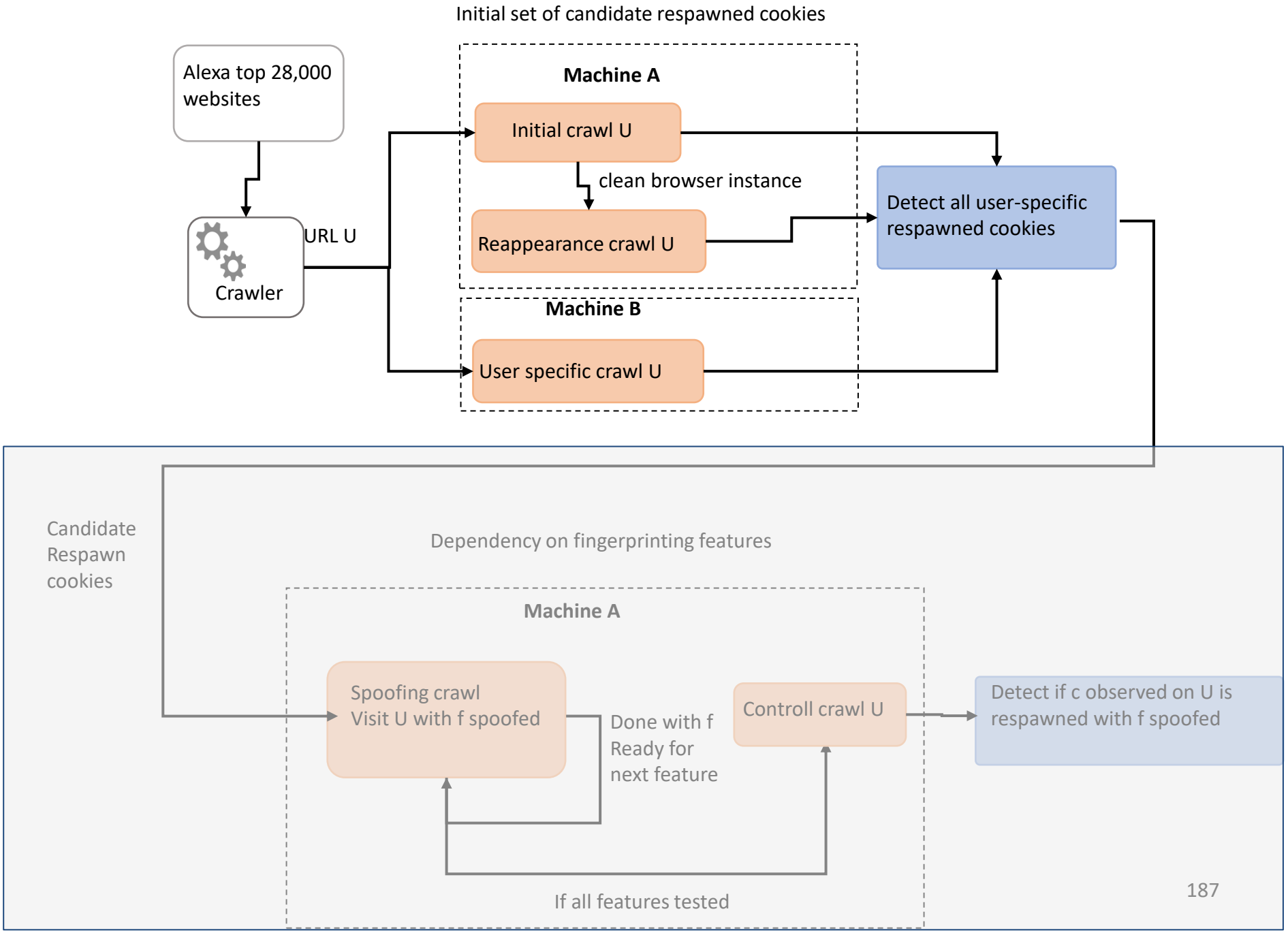
Cookie respawning via browser fingerprinting



Cookie respawning via browser fingerprinting

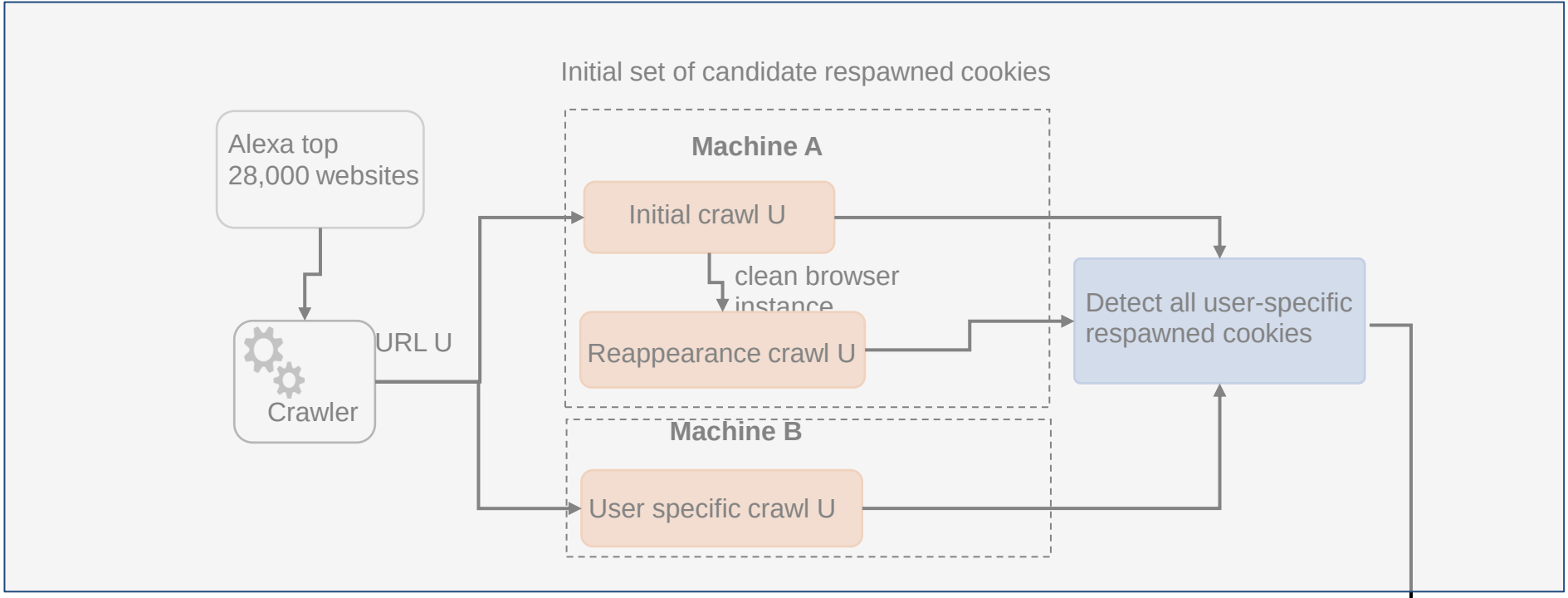
The user cleans her browser




Initial set of candidate respawned cookies
Alexa top 28,000 websites
Machine A
Initial crawl U
clean browser instance
Reappearance crawl U
Crawler
URL U
Machine B
User specific crawl U
Detect all user-specific respawned cookies
Candidate Respawn cookies
Dependency on fingerprinting features
Machine A
Spoofing crawl
Visit U with f spoofed
Done with f
Ready for
next feature
Controll crawl U
Detect if c observed on U is respawned with f spoofed
If all features tested

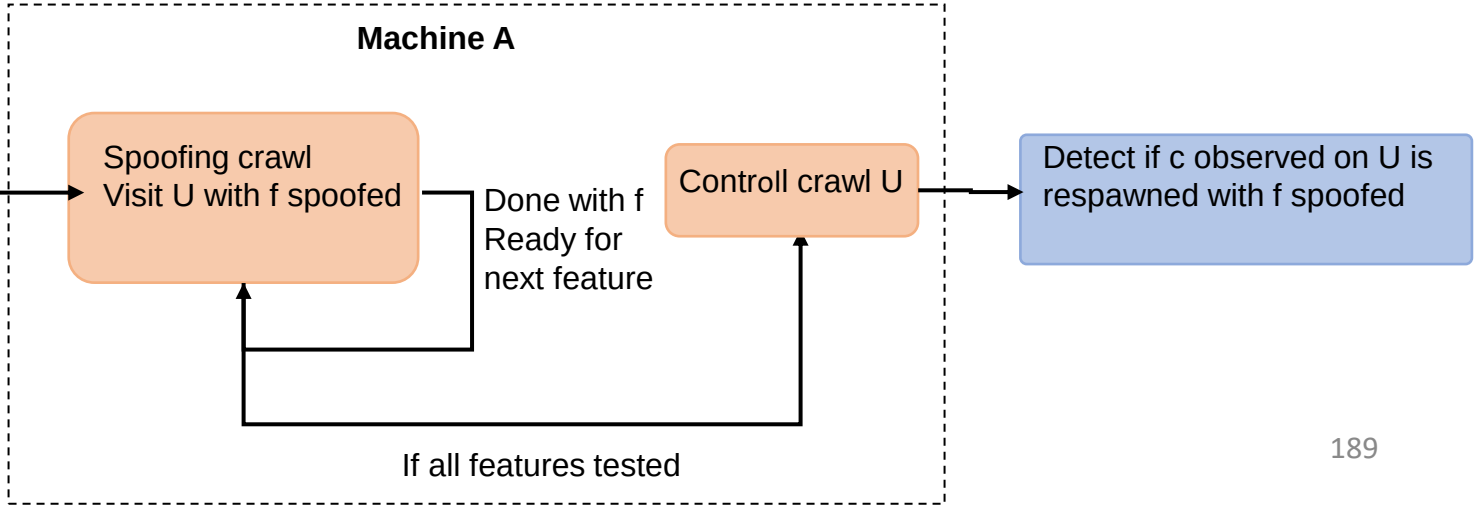
Cookie respawning via browser fingerprinting

Crawls	*Initial*	*Reappear-ance*	*User specific*	*Control*
Collected cookies	541,691	84,956	5,547	1,883
Occurrence on websites	28,500	13,782	3,674	1,781



Candidate Respawn cookies

Dependency on fingerprinting features

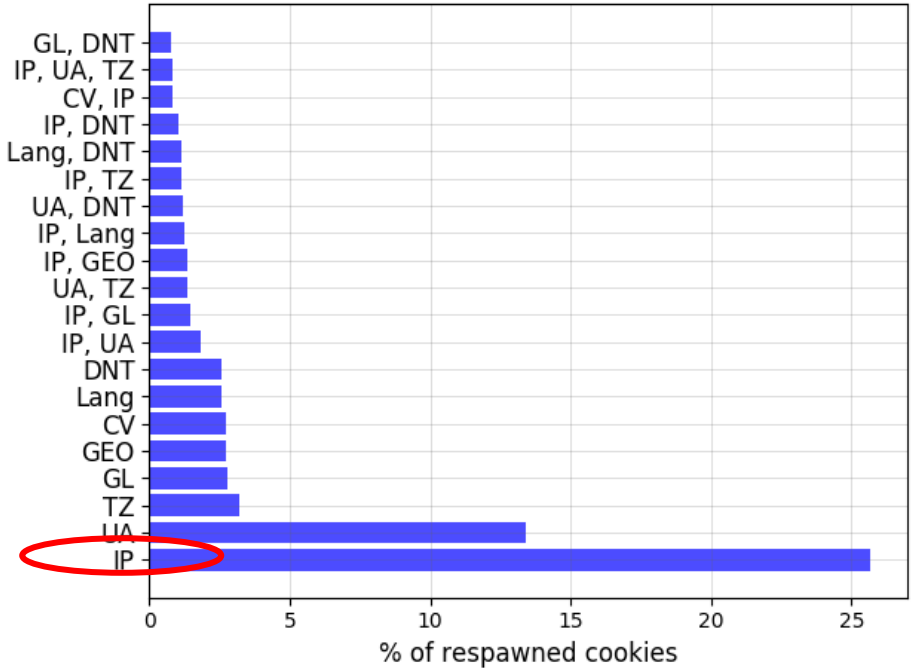


Cookie respawning via browser fingerprinting

Crawls	*Initial*	*Reappear-ance*	*User specific*	*Control*
Collected cookies	541,691	84,956	5,547	1,883
Occurrence on websites	28,500	13,782	3,674	1,781

How?

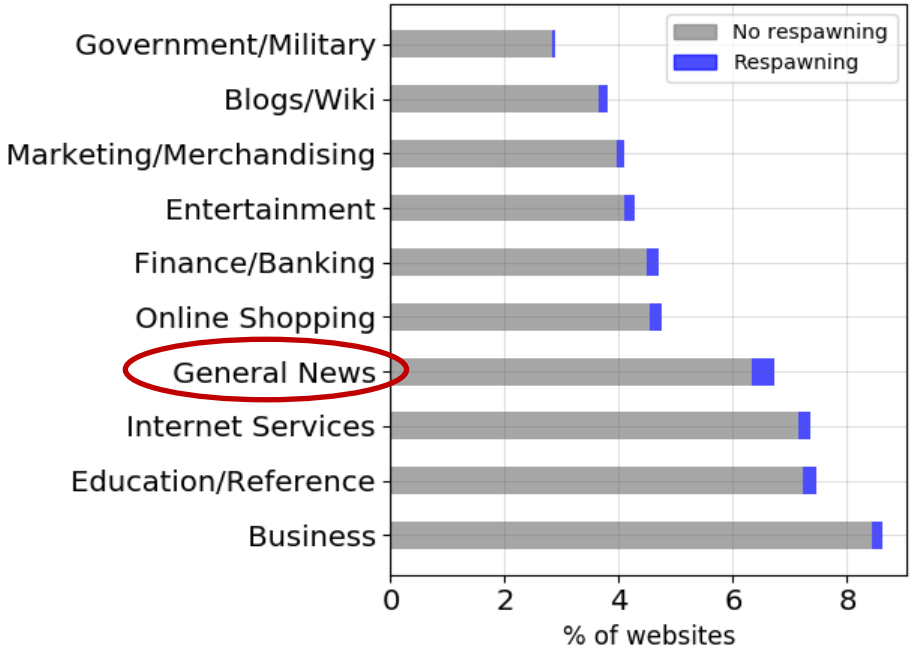
- Cookies are respawned with 184 distinct sets of features
- The IP address is used alone to respawn 366 (25.68%) cookies



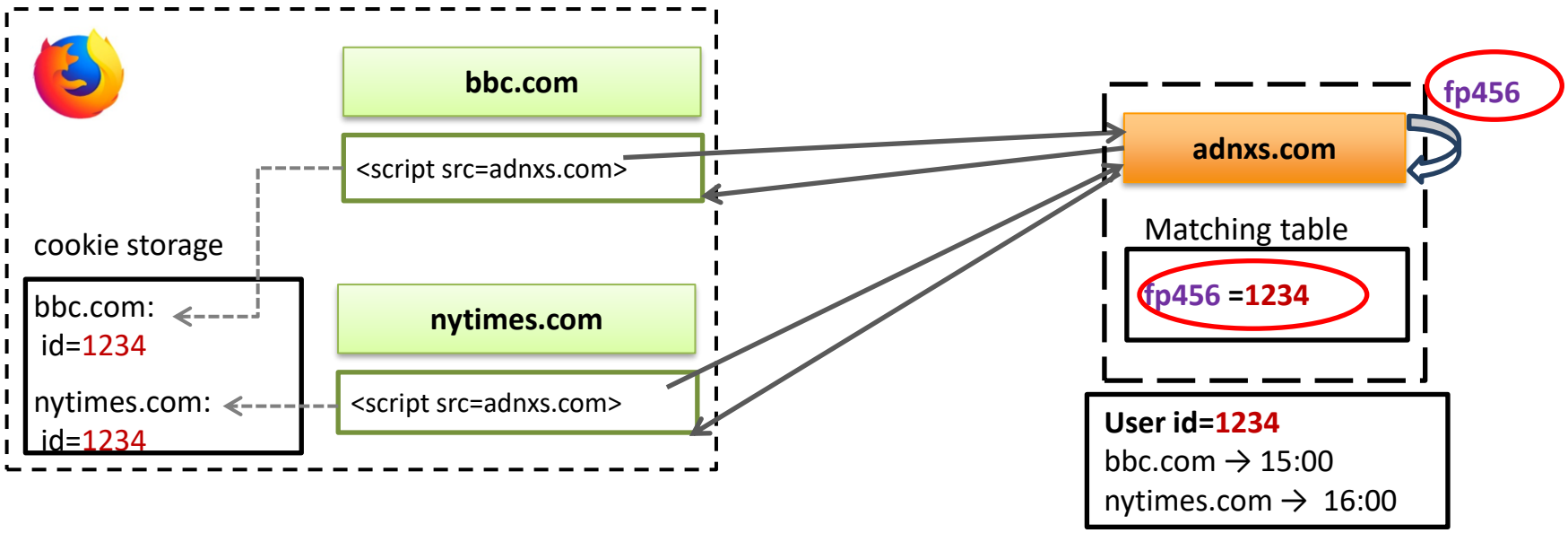
Mishra et al. [48] showed that 87% of participants (out of 2,230) retain at least one IP address for more than a month

Where?

- We identified respawning on 143 distinct categories
- 5.95% of News websites contain at least one respawned cookie
- 21 cookies are respawned on adult websites



Tracking consequences!



Cookie respawning with browser fingerprinting enables a stable and persistent cross site tracking using first-party cookies

Legal consequences!

Cookie respawning with browser fingerprinting violates the following principles

- **Fairness principle:** Personal data must be processed fairly (Article 5(1)(a))
  - **All 1,425 respawned cookies**

- **Transparency principle:** Personal data processing must be handled in a transparent manner in relation to the user (Article 5(1)(a))
  - **Top 10 popular respawned cookies owners**

- **Lawfulness principle:** Websites should obtain user consent (Articles 5(1)(a) and 6(1))
  - **38.69% of the respawned cookies**

Not all Cookies are Subject to Consent

Three criteria are used to evaluate whether cookies are exempted or subject to consent

- Purpose
- Duration (Session or persistent)
- Context (first or third party cookies)

=> Cookie respawning with browser fingerprinting **bypass** criteria used to identify the need of user's consent

31.30% of the respawned cookies are session cookies

“If you're not paying for the product, you are the product”

--Andrew Lewis--

Thank you for attending this course

Sources of this Presentation

Note: References in bold are highly recommended reads.

- [1] Cache Logic http://www.cachelogic.com/research/
- [2] Keith W. Ross and Dan Rubenstein "P2P Systems". Infocom 2004 tutorial. http://cis.poly.edu/~ross/tutorials/P2PtutorialInfocom.pdf
- [3] S. Sen and Jia Wang "Analysing peer-to-peer traffic across large networks". ACM SIGCOMM'02
- [4] T. Karagiannis, A. Broido, M. Faloutsos, Kc Claffy "Transport Layer Identification of P2P Traffic". ACM IMC'04
- **[5] X. Yang and G. de Veciana "Service Capacity of Peer to Peer Networks". IEEE Infocom'04**
- **[6] D. Qiu and R. Srikant "Modeling and Performance Analysis of BitTorrent-Like Peer-to-Peer Networks". ACM SIGCOMM'04**
- **[7] J. H. Saltzer, D. P. Reed, and D. D. Clark "End-to-end arguments in system design". *ACM Transactions on Computer Systems 2*, 4 (November 1984) pages 277-288**
- **[8] P. Rodriguez, E. W. Biersack "Dynamic Parallel Access to Replicated Content in the Internet". *IEEE/ACM Transactions on Networking,* August 2002 *(Also in IEEE/Infocom 2000)***
- [9] E. W. Biersack, P. Rodriguez, P. Felber "Performance Analysis of Peer-to-Peer Networks for File Distribution". Research Report RR-04-108. April 2004.
- [10] P. A. Felber and E. W. Biersack. Self-scaling Networks for Content Distributions. In Ozalp Babaoglu et~al., editors, *Self-Star Properties in Complex Information Systems*, volume 3460 of *Lecture Notes in Computer Science*. Springer-Verlag, 2005.
- [11] E. K. Lua et al. "A Survey and Comparison of Peer-to-Peer Overlay Network Schemes", IEEE Communications survey and tutorial, March 2004.
- [12] Jian Liang, Rakesh Kumar, Keith Ross,"The KaZaA Overlay: A Measurement Study", Computer Networks (Special Issue on Overlays), to appear.
- [13] Y. Kulbak, D. Bickson "The eMule Protocol Specification" January 2005
- [14] C. Gkantsidis, P. Rodriguez "Network Coding for Large Scale Content Distribution"

Sources of this Presentation

Note: References in bold are highly recommended reads.

- [15] Dejan Kostic, A. Rodriguez, J. Albrecht, and A. Vahdat "Bullet: High Bandwidth Data Dissemination Using an Overlay Mesh" SOSP'03, October 2003.
- [16] T. Klingberg, R. Manfredi , "Gnutella Protocol Development v0.6", June 2002
- [17] T. Klingberg, "Partial File Sharing Protocol", August 2002
- [18] A. Legout, G. Urvoy-Keller, and P. Michiardi. ''Understanding BitTorrent: An Experimental Perspective''. *Technical Report (inria-00000156, version 3 - 9 November 2005)*, INRIA, Sophia Antipolis, November 2005.
- [18] BitTorrent Protocol Specification v1.0. http://wiki.theory.org/BitTorrentSpecification
- **[19] Bram Cohen, "Incentives Build Robustness in BitTorrent", May 2003**
- [20] J.A Pouwelse et al., "The BitTorrent P2P File-Sharing System: Measurements and Analysis", IPTPS 2005
- [21] M. Izal et al., "Dissecting BitTorrent: Five Months in a Torrent's Lifetime", PAM 2004
- [22] L. Guo et al., "Measurements, Analysis, and Modeling of BitTorrent-like Systems" IMC 2005
- **[23] Shamir "How to Share a Secret" Communications of the ACM, 1979**
- **[24] D. L. Chaum "Untraceable Electronic Mail, Return Addresses and Digital Pseudonyms" Communications of the ACM, 1981**
- **[25]Stoica et al. "Chord: A Scalable Peer-to-Peer Lookup Service for Internet Applications" ACM SIGCOMM'01**
- [26]Rowstron and Druschel "Pastry: Scalable, decentralized object location and routing for large-scale peer-to-peer systems"
- [27] Maymounkov and Mazières "Kademlia: A peer-to-peer Information System Based on the XOR Metric"
- [28] Ratnasamy et al. "A Scalable Content-Addressable Network", SIGCOMM'01
- **[29] Gummadi et al. "The Impact of DHT Routing Geometry on Resilience Proximity"**

Sources of this Presentation

Note: References in bold are highly recommended reads.

- [30] C. Fragouli, J.-Y. Le Boudec and J. Widmer "Network Coding: An Instant Primer". *ACM Sigcomm Computer Communication Review*, Vol. 36, Nr. 1, pp. 63-68, 2006.
- [31] C. Gkantsidis, J. Miller, P. Rodriguez "Comprehensive view of a Live Network Coding P2P system". ACM SIGCOMM/USENIX IMC'06, Brasil. Oct 2006.
- [32] C. Gkantsidis, P. Rodriguez "Cooperative Security for Network Coding File Distribution". IEEE/INFOCOM'06, Barcelona, April 2006.
- [33] A. R. Bharambe, C. Herley, and V. N. Padmanabhan "Analyzing and Improving a BitTorrent Network's Performance Mechanisms". In Proc. of Infocom'06, Barcelona, Spain, April 2006.
- **[34] A. Legout, G. Urvoy-Keller, and P. Michiardi "Rarest First and Choke Algorithms Are Enough". In *Proc. of ACM SIGCOMM/USENIX IMC'2006*, Rio de Janeiro, Brazil, October 2006.**
- **[35] A. Legout, N. Liogkas, E. Kohler, and L. Zhang "Clustering and Sharing Incentives in BitTorrent Systems". *Technical Report (inria-00112066, version 1 - 21 November 2006)*, INRIA, Sophia Antipolis, November 2006.**
- [36] S. Jun and M. Ahamad."Incentives in BitTorrent Induce Free Riding". In Proc. of the Workshop on Economics of Peer-to-Peer Systems (P2PEcon'05), Philadelphia, PA, August 2005.
- [37] N. Liogkas, R. Nelson, E. Kohler, and L. Zhang. "Exploiting Bittorrent For Fun (But Not Profit)". In Proc. of IPTPS'06, Santa Barbara, CA, February 2006.
- [38] T. Locher, P. Moor, S. Schmid, and R.Wattenhofer. "Free Riding in BitTorrent is Cheap." In Proc. of HotNets-V, Irvine, CA, November 2006.
- [39] J. Shneidman, D. Parkes, and L. Massoulie. "Faithfulness in Internet Algorithms". In Proc. of theWorkshop on Practice and Theory of Incentives and Game Theory in Networked Systems (PINS'04), Portland, OR, September 2004.
- [40] B Fan, DM chiu and JCS Lui, "The Delicate Tradeoff of BitTorrent-like File Sharing Protocol Design", IEEE ICNP 2006

Sources of this Presentation

Note: References in bold are highly recommended reads.

- [41] Kenjiro Cho, Kensuke Fukuda, Hiroshi Esaki,Akira Kato “Observing Slow Crustal Movement in Residential User Traffic”. CoNext’2008, December 2008.
- [42] C. Zhang, P. Dunghel, D. Wu, K.W. Ross, “Unraveling the BitTorrent Ecosystem”. To appear in IEEE Transactions on Parallel and Distributed Systems.
- [43] Stevens Le Blond, Arnaud Legout, Walid Dabbous. “Pushing BitTorrent Locality to the Limit”. *C*omputer Networks, October 2010, ISSN 1389-1286, DOI: 10.1016/j.comnet.2010.09.014.
- [44] Anwar Al Hamra, Nikitas Liogkas, Arnaud Legout, Chadi Barakat. “Swarming Overlay Construction Strategies”. In *Proc. of ICCCN'2009*, August 2--6, 2009, San Francisco, CA, USA.
- [45] T. Karagiannis, P. Rodriguez, and K. Papagiannaki. “Should internet service providers fear peer-assisted content distribution?” In Proc. of IMC'05, Berkeley, CA, USA, October 2005.
- [46] H. Xie, Y. R. Yang, A. Krishnamurthy, Y. Liu, and A. Silberschatz. “P4p: Provider portal for applications.” In Proc. of ACM SIGCOMM, Seattle, WA, USA, August 2008.
- [47] D. R. Choffnes and F. E. Bustamante. “Taming the torrent: A practical approach to reducing cross-isp traffic in p2p systems.” In Proc. of ACM SIGCOMM, Seattle, WA, USA, August 2008.
- [48] The BitTorrent Protocol Specification, BEP 3, http://www.bittorrent.org/beps/bep_0003.html
- [49] Peer ID Conventions , BEP 20, http://www.bittorrent.org/beps/bep_0020.html
- [50] R. Dingledine, N. Mathewson, P. Syverson. “Tor: The Second-Generation Onion Router.” In proc. of Usenix Security’2004, San Diego, CA, USA, August 2004.
- [51] Niels Ferguson, Bruce Schneier, Tadayoshi Kohno. “Cryptography Engineering.” 2010, Wiley.

Sources of this Presentation

Note: References in bold are highly recommended reads.

- [52] Stevens Le Blond, Arnaud Legout, Fabrice Lefessant, Walid Dabbous, Mohamed Ali Kaafar. “Spying the World from your Laptop - Identifying and Profiling Content Providers and Big Downloaders in BitTorrent.” In *Proc. of LEET'10*, April 27, 2010, San Jose, CA, USA.
- [53] Sandvine. “Global Internet Phenomena Report”, Spring 2011.
- [54] Stevens Le Blond, Chao Zhang, Arnaud Legout, Keith Ross, and Walid Dabbous. “ I Know Where You are and What You are Sharing: Exploiting P2P Communications to Invade Users' Privacy. ” In *Proc. of ACM SIGCOMM/USENIX IMC'11*, Nov. 2--3, 2011, Berlin, Germany.
- [55] Stevens Le Blond, Pere Manils, Abdelberi Chaabane, Mohamed Ali Kaafar, Claude Castelluccia, Arnaud Legout, Walid Dabbous. “ One Bad Apple Spoils the Bunch: Exploiting P2P Applications to Trace and Profile Tor Users. ” In *Proc. of LEET'11*, March 29, 2011, Boston, MA, USA.
- [56] Sandvine. “Global Internet Phenomena Report”, 1H 2014

Heartbleed attack

- Disclosed in April 2014
  - Issue introduced in December 2012
  - https://en.wikipedia.org/wiki/Heartbleed
- Only one contributor to open SSL in 2014
  - https://github.com/openssl/openssl/graphs/contributors
  - https://www.buzzfeed.com/chrisstokelwalker/the-internet-is-being-protected-by-two-guys-named-st

SolarWinds attack

- Russian cyberespionage campaign
- What a powerful and competent adversary can do!
- Step1: find a company providing network-management tools (easy)
  - SolarWinds
  - Many critical clients including private sector entities and government agencies
  - 18k customers received this update

SolarWinds attack

- Step 2: access the build environment (easy)
  - Most likely targeted phishing attack to get credentials
- Step 3: inject a malware into the product updates (hard)
  - 18k customers received this update
  - Do nothing for weeks to do not be detected and decorrelate a possible attack to the source (the product update)
  - Uninstall everything at SolarWinds

SolarWinds attack

- Step 4: decide where to exploit the hole (easy)
  - Get location of victims to decide on their value
    - Communication only during legitimate communication of the product (hard to detect)
  - Target only high value victims, do not take the risk to be detected on low value victims
    - Uninstall everything on low value victims
- Step 5: install trojan on high value victims (hard)
  - Scan for possible hard to escape security products
    - If risk of being detected, uninstall everything
    - Otherwise, download a trojan

SolarWinds attack

- Step 6: use a commercial malware penetration testing product: Cobalt Strike Beacon
  - Sold by the Minnesota company HelpSystems
  - Customize a loader to prevent detection of the installation
- Step 7: access the Microsoft 365 cloud environment for all compromised victims
  - By gaining access through the malware of the Active Directory Federation Services

https://www.lawfaremedia.org/article/solarwinds-and-holiday-bear-campaign-case-study-classroom

Internet Timeline

From 1962 to 1991

Disclaimer: I found this timeline long ago, but I don't remember where. If you did it and want to be credited for it, send me an email.

Internet Timeline

- 1962 **Kleinrock thesis describes underlying principles of packet-switching technology**
- 1966 **ARPANET project**
  - Larry Roberts of MIT's Lincoln Lab is hired to manage the ARPANET project.
  - ARPA computer network, a packet-switched network with minicomputers acting as gateways for each node using a standard interface.
- 1967 **Packet switching**
  - Donald Davies, of the National Physical Laboratory in Middlesex, England, coins the term *packet switching* to describe the lab's experimental data transmission.

Internet Timeline

- 1968 **Interface message processors**
  - Bolt Beranek and Newman, Inc. (BBN) wins a DARPA contract to develop the packet switches called interface message processors (IMPs).
- 1969 **DARPA deploys the IMPs**
  - First transmission between UCLA and Stanford: “lo”
- 1970 **Initial ARPANET host-to-host protocol**
  - Network Working Group (NWG), formed at UCLA by Steve Crocker, deploys the initial ARPANET host-to-host protocol, called the Network Control Protocol (NCP). The primary function of the NCP is to establish connections, break connections, switch connections, and control flow over the ARPANET, which grows at the rate of one new node per month.

Internet Timeline

- 1972 **First e-mail program**
  - Ray Tomlinson at BBN writes the first e-mail program to send messages across the ARPANET. In sending the first message to himself to test it out, he uses the @ sign—the first time it appears in an e-mail address.
- 1972 **First public demonstration of the new network technology**
  - Robert Kahn at BBN, who is responsible for the ARPANET's system design, organizes the first public demonstration of the new network technology at the International Conference on Computer Communications in Washington, D.C., linking 40 machines and a Terminal Interface Processor to the ARPANET.
- 1973 **Paper describes basic design of the Internet and TCP**
  - Robert Kahn and Vinton Cerf, "A Protocol for Packet Network Interconnection" in *IEEE Transactions on Communications*.
- 1974 F.F. Kuo "ALOHA System", January 1974

Internet Timeline

- 1976 **TCP/IP incorporated in** Berkeley Unix
- 1977 **Demonstration of independent networks to communicate**
  - Cerf and Kahn organize a demonstration of the ability of three independent networks to communicate with each other using TCP protocol.
- 1981 **TCP/IP standard adopted**
  - Postel, J., "Internet Protocol", STD 5, RFC 791, September 1981.
  - Postel, J., "Transmission Control Protocol", STD 7, RFC 793, September 1981.
- 1982 **ARPANET hosts convert to new TCP/IP protocols**
  - All hosts connected to ARPANET are required to convert to the new TCP/IP protocols by January 1, 1983. The interconnected TCP/IP networks are generally known as the Internet.

Internet Timeline

- 1983 **UNIX scientific workstation introduced**
  - Sun Microsystems introduces its UNIX scientific workstation. TCP/IP, now known as the Internet protocol suite, is included, initiating broad diffusion of the Internet into the scientific and engineering research communities
- 1983 **The Internet**
  - ARPANET, and all networks attached to it, officially adopts the TCP/IP networking protocol. From now on, all networks that use TCP/IP are collectively known as the Internet. The number of Internet sites and users grow exponentially
- 1984 **Advent of Domain Name Service. D**eveloped by Paul Mockapetris and Craig Partridge
- 1984 J. H. Saltzer, D. P. Reed, and D. D. Clark "End-to-end arguments in system design" *ACM Transactions on Computer Systems,* November 1984

Internet Timeline

- 1984 John Nagle "Congestion Control in IP/TCP Internetworks" October 1984
- October 1986 First Congestion Collapse
  - From 32 Kbps to 40 bps
- 1988 Van Jacobson "Congestion Avoidance and Control" SIGCOMM'88, August 1988
- 1991 **World Wide Web software developed**
  - CERN releases the World Wide Web software developed earlier by Tim Berners-Lee. Specifications for HTML (hypertext markup language), URL (uniform resource locator), and HTTP (hypertext transfer protocol) launch a new era for content distribution.