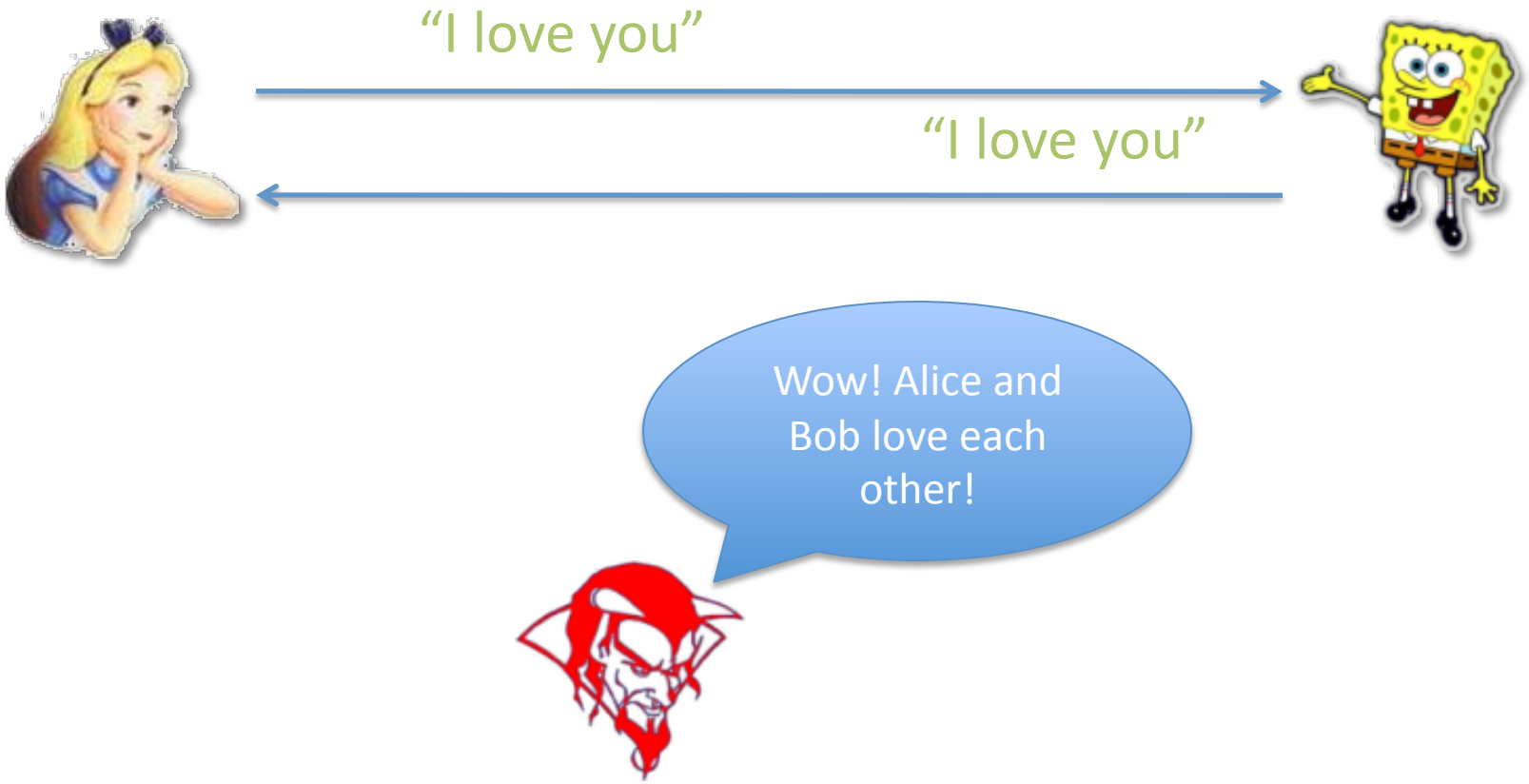
CSE 461: Privacy

Ben Greenstein

Privacy

- “the quality or state of being apart from company or observation”
  - Merriam-Webster

Network privacy threat



Is providing privacy easy?

E_k("I love you")
→ wnvq31#2v.

D_k(wnvq31#2v.)
→ "I love you"

wnvq31#2v.

wnvq31#2v.

????

Is providing privacy easy?

E_k("I love you")
→ wnvq31#2v.

D_k(wnvq31#2v.)
→ "I love you"

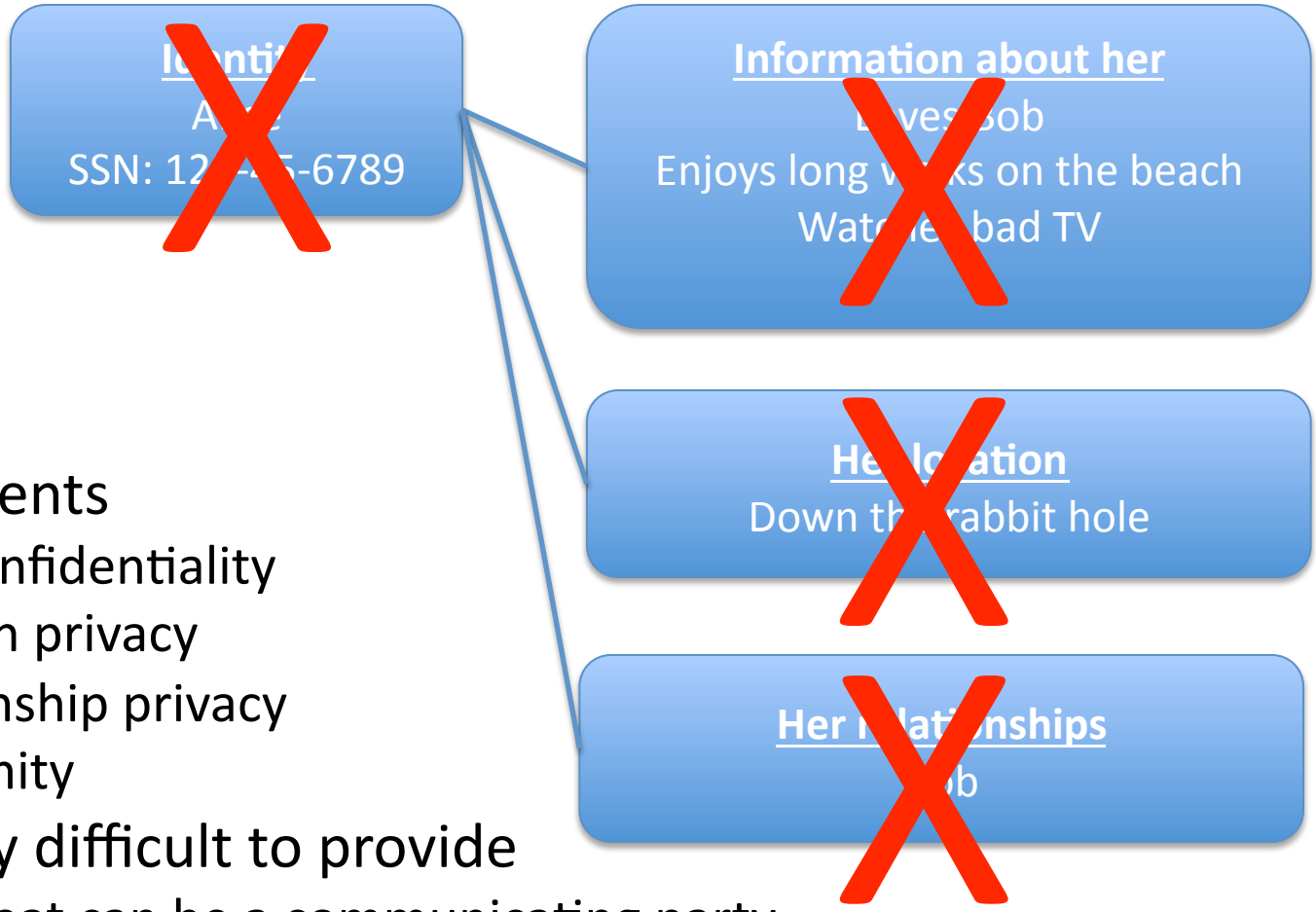
wnvq31#2v.

wnvq31#2v.

Hmm, why are Alice and Bob talking? Maybe they love each other

Obviously platonic, given Bob is sending messages from the SF Gay Men's Chorus Bldg.

Providing Privacy



- Requirements
  - Data confidentiality
  - Location privacy
  - Relationship privacy
  - Anonymity
- Often very difficult to provide
  - The threat can be a communicating party

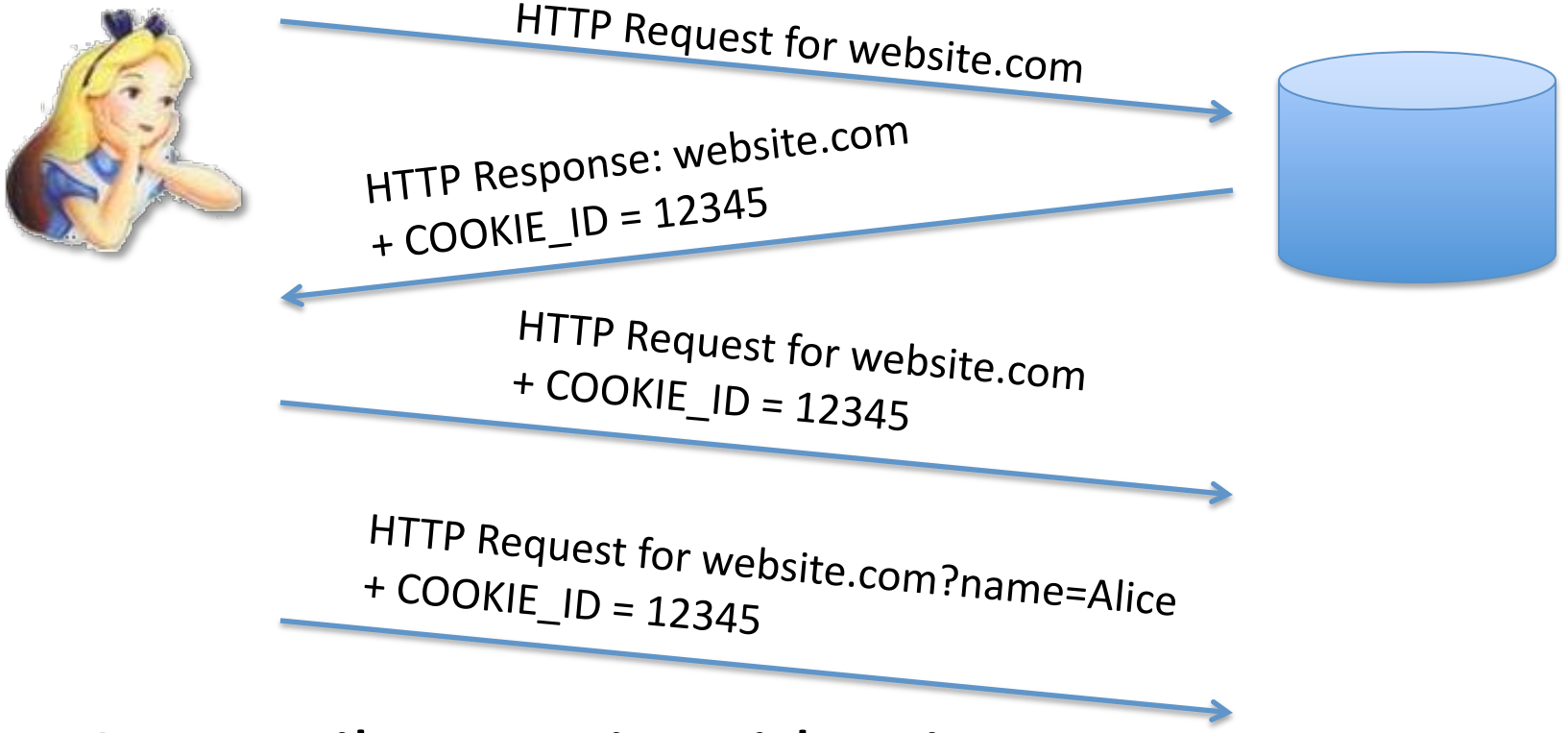
Outline

- Application layer threats and challenges
  - Solution ???
- Network layer threats and challenges
  - One solution: TOR
- Link Layer threats and challenges
  - One solution: SlyFi

The application layer

The dreaded cookie

- Chief mechanism for connecting user sessions



- Can easily associate identity

Threat: Web sites

- User gives personal data to web sites
  - "Privacy is for old people. This is the MySpace generation. People publish every detail of their lives for all the world to see." -- Mary Baker, HP, 2007
- Web sites don't do enough to respect privacy
  - "We cannot guarantee that only authorized persons will view your information. We cannot ensure that information you share on Facebook will not become publicly available. We are not responsible for third party circumvention of any privacy settings or security measures on Facebook." - http://www.facebook.com/policy.php

Consequence

Teachers and Facebook: Privacy vs. standards

As CMS plans to clarify policy, teacher's attorney says she never intended posting to be public.

By Fred Clasen-Kelly
frkelly@charlotteobserver.com

Posted: Friday, Nov. 14, 2008

An attorney for a suspended Charlotte-Mecklenburg Schools teacher said Thursday she never intended for the public to view negative comments she made about students on Facebook.

She now faces possible firing for listing "teaching chitlins in the ghetto of Charlotte" among her activities.

Consequence

Sears Installs Spyware

Posted by kdawson on Thu Jan 03, 2008 11:35 AM
from the **naughty-naughty** dept.

Gandalf_the_Beardy writes in with news that's been around a while but is getting more attention lately. Last month Benjamin Googins, a security researcher at CA, determined that Sears Holding Corp. installed ComScore spyware without adequate disclosure. Sears said, yes we tell people about tracking their browsing. On Jan. 1 spyware researcher Ben Edelman weighed in, noting that Sears' notice occurs on page 10 of a 54-page privacy statement, and twits Sears because its installation identifies the software as "VoiceFive" and later claims it's coming from a company called "TMRG, Inc." even though a packet sniffer confirms the software belongs to ComScore, adding "These confusing name-changes fit the trend among spyware vendors."

Btw, the threat doesn't have to be intentional

AOL's Massive Data Leak

Take Action: Were You Exposed By AOL's Data Leak?

Spread the word: Get buttons for your blog and email friends

In August 2006, AOL publicly released three months of search queries by 650,000 AOL users. Though AOL has removed the data from its site and rightly apologized, the grave damage is already done. The data quickly became available all over the Net, and AOL may have violated its own privacy policy as well as existing federal law. Both companies like AOL and Congress should heed the lessons of this Data Valdez and enhance protections for your privacy. On August 14, EFF asked [PDF] the Federal Trade Commission (FTC) to investigate AOL and require changes in its privacy practices.

AOL's actions demonstrate a shocking disregard for user privacy. Search terms can expose the most intimate details of a person's life. These details can be embarrassing and even cause great harm. Would you want strangers to know where you or your child work or go to school? How about everyone seeing search queries that reference your financial information, medical history, sexual orientation, or religious affiliation?"

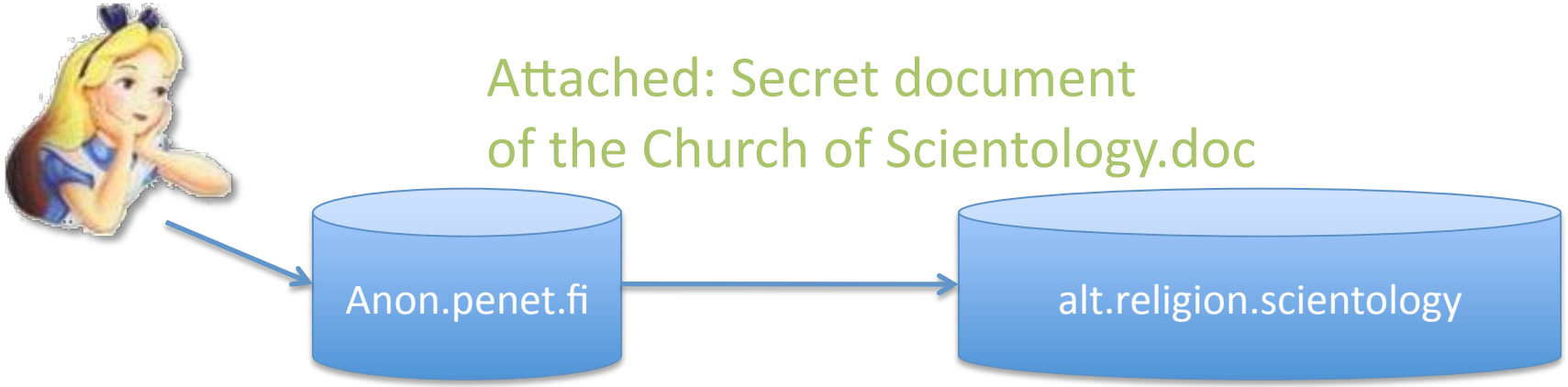
Though the data was associated with random ID numbers, that information could still be connected back to an individual given enough clues, as this NY Times article clearly demonstrates. Whether it's because of vanity searches for your name or MySpace profile or searches related to your city and neighborhood, your search history could create a trail of breadcrumbs that ultimately leads to your doorstep.

Challenges

- How can we learn what information about us is collected?
  - Limited visibility into where your data is going:
    E.g., evite shares your zip code with doubleclick.net, adbureau.net, abmr.net, voicefive.com, target.com and atdmt.com
- How do we ensure anonymity when we want it?
  - E.g., when searching.
  - Note: blocking cookies breaks apps
- How do we control the exposure of personal information?
- What's the right balance between functionality and privacy?

The network layer

Threat: Nodes in between

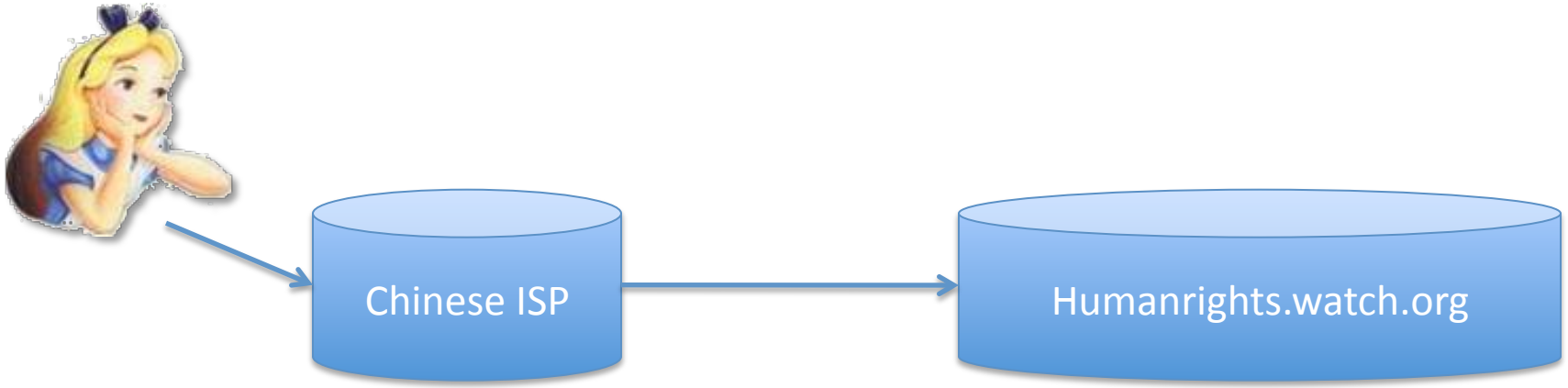


- Address linked back to anonymous remailer
- Remailer gives up Alice's identity

Overeager governments simplify tracking

- EU directive 2006/24/EC: 3 year data retention
  - For ALL traffic, **requires EU ISPs to record**:
    - Sufficient information to identify endpoints
      (both legal entities and natural persons)
    - Session duration
    - … but not session contents
  - Make available to law enforcement
    - … but penalties for transfer or other access to data

Threat: Nodes in between



- Same problem with different nouns…
  - Remailer (anon.penet.fi) → ISP
  - Newsgroup (alt.religion.scientology) → website.

- ISP sees Alice's destination
- ISP blocks access
  - Also might notify government

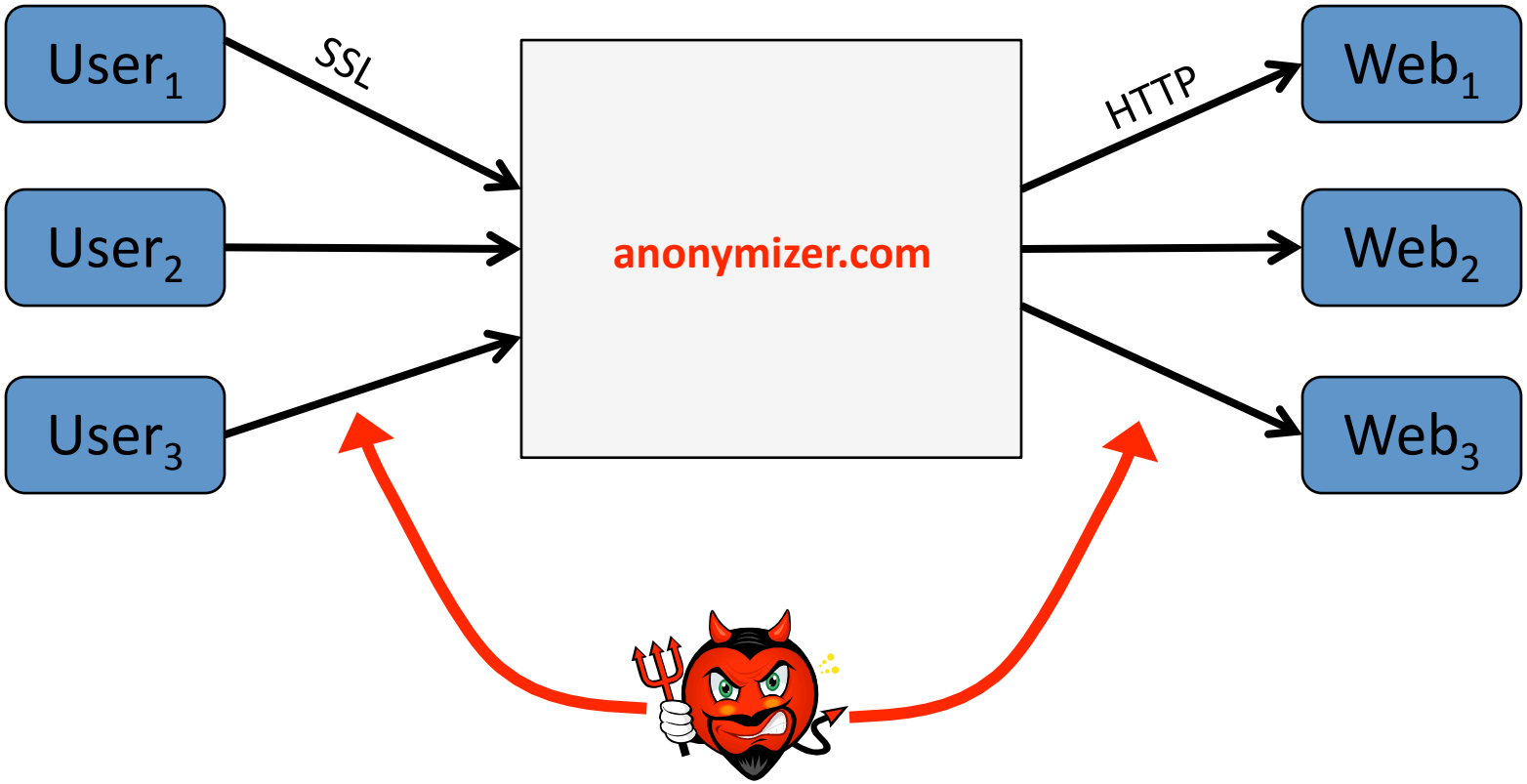
- Identities easier to determine when using IPv6 autoconfiguration
  - Binds MAC (a globally unique identifier) to your IP address

TOR: For anonymous Web browsing

- Why?
  - Discuss health issues and financial matters anonymously, conceal interactions with gambling sites
  - Bypass Internet censorship in parts of the world
  - Law enforcement
- Two goals:
  - Hide user identity from target web site
  - Hide browsing pattern from employer or ISP

1st attempt: anonymizing proxy

HTTPS:// anonymizer.com ? URL=target



Anonymizing proxy: security

- Monitoring ONE link: eavesdropper gets nothing
- Monitoring TWO links:
  - Eavesdropper can do traffic analysis
  - More difficult if lots of traffic through proxy

- Trust: proxy is a single point of failure
  - Can be corrupt or subpoenaed
    - Example: The Church of Scientology vs. anon.penet.fi

- Protocol issues:
  - Long-lived cookies make connections to site **linkable**

How proxy works

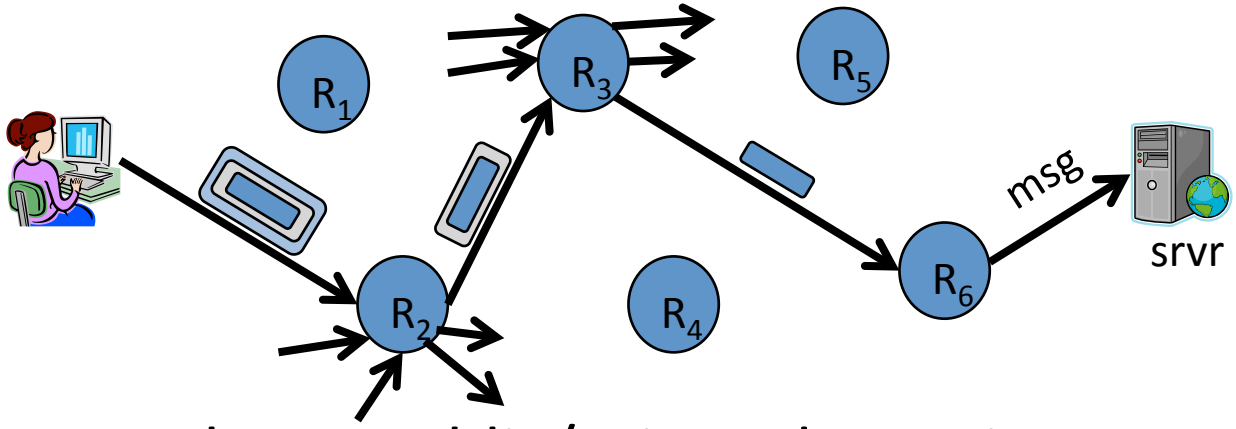
- Proxy rewrites all links in response from web site
  - Updated links point to anonymizer.com
    - Ensures all subsequent clicks are anonymized
- Proxy rewrites/removes cookies and some HTTP headers

- Proxy IP address:
  - if a single address, could be blocked by site or ISP
  - anonymizer.com consists of >20,000 addresses
    - Globally distributed, registered to multiple domains
    - Note: chinese firewall blocks ALL anonymizer.com addresses

- Other issues: attacks (click fraud) through proxy

2nd Attempt: Mix Nets [C'81]

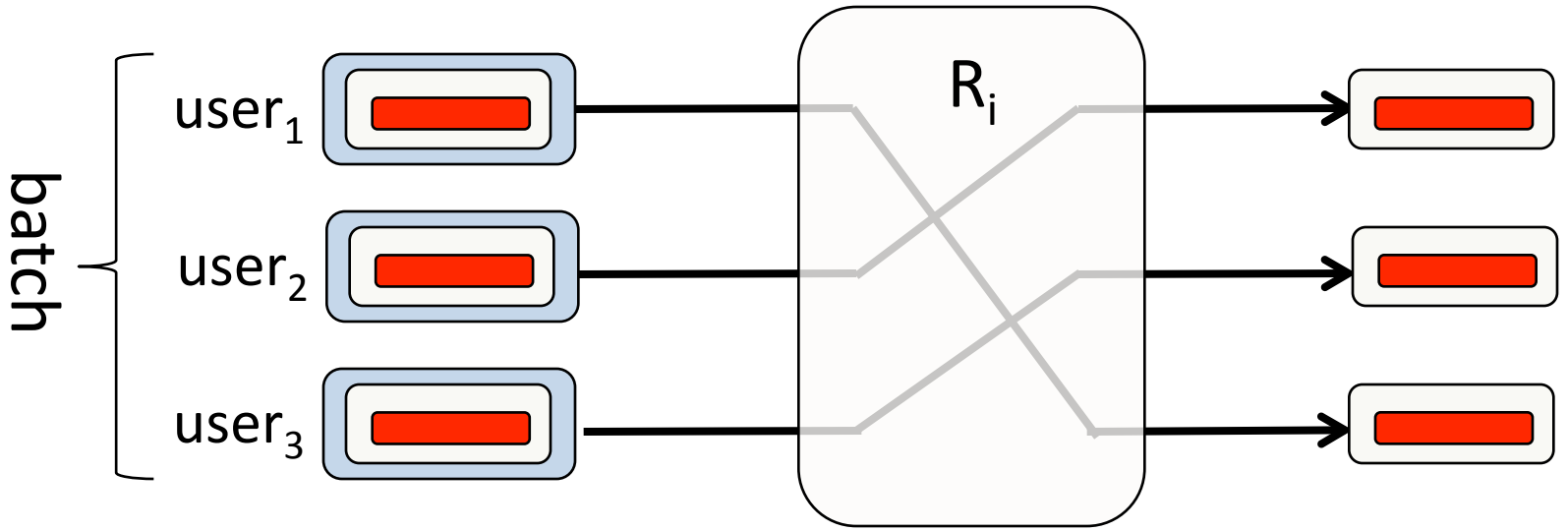
- Goal: No single point of failure



- Every router has a public/private key pair
  - Sender knows all public keys
- To send packet:
  - Pick random route: $R_2 \rightarrow R_3 \rightarrow R_6 \rightarrow$ srvr
  - Prepare **onion packet**:

packet = $E_{pk_2}(R_3,\ E_{pk_3}(R_6,\ E_{pk_6}(\text{srvr}, \text{msg})))$

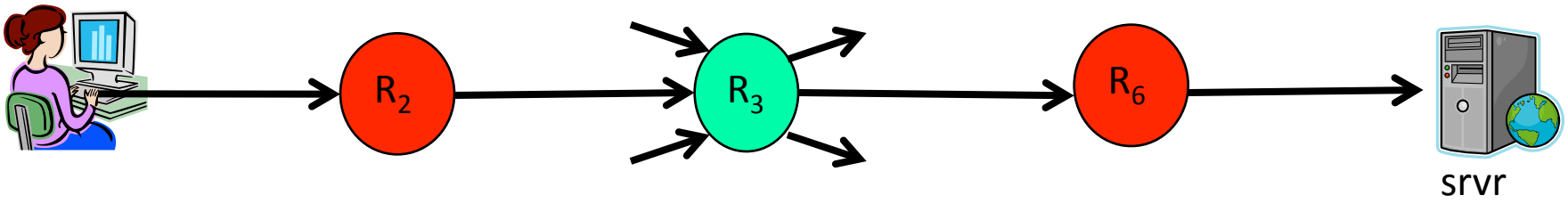
Eavesdropper's view at a single MIX



- Eavesdropper observes incoming and outgoing traffic
- Crypto prevents linking input/output pairs
  - Assuming enough packets in incoming batch
  - If variable length packets then must pad all to max length

Performance

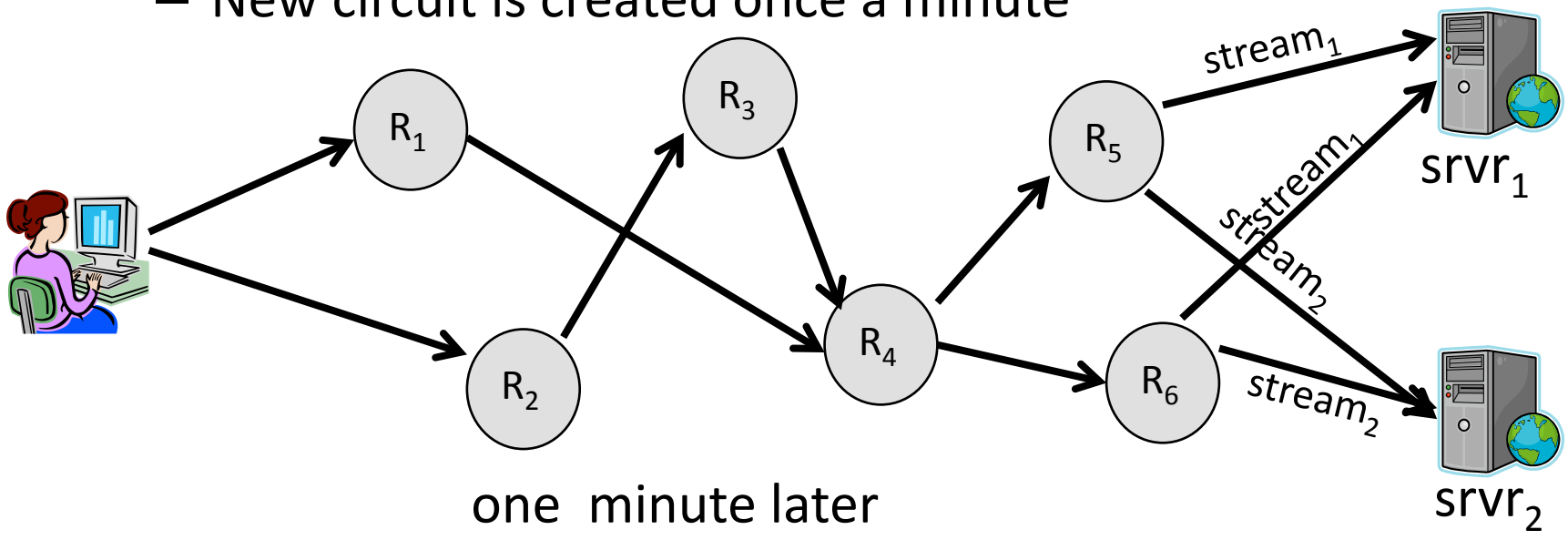
- Main benefit:
  - Privacy as long as **at least one** honest router on path



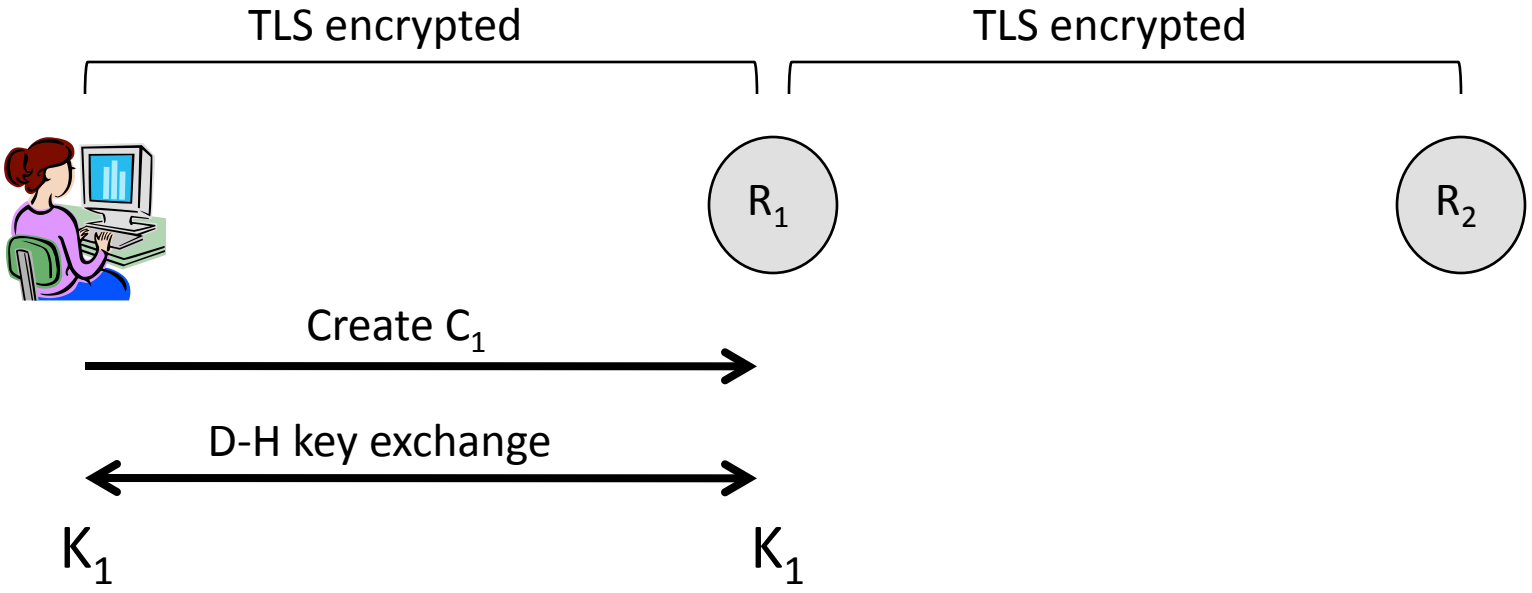
- Problems:
  - High latency (lots of public key ops)
    - Inappropriate for interactive sessions
    - May be OK for email (e.g. Babel system)
  - No forward security
- How does server respond?
  - hint: user includes “response onion” in forward packet

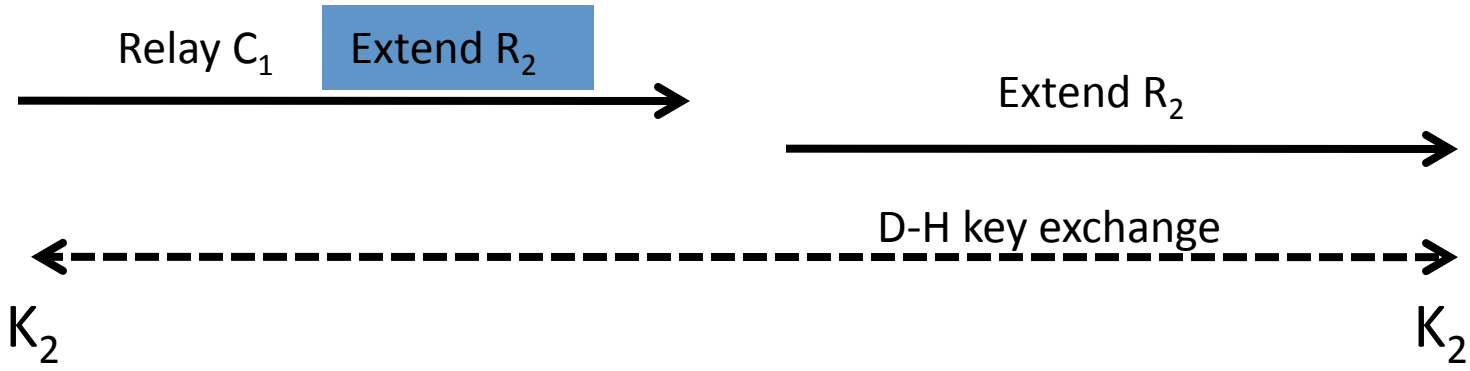
3rd Attempt: Tor Mix circuit-based design

- Goals: No single failure point & better performance
- Trusted directory contains list of Tor routers
- User's machine preemptively creates a circuit
  - Used for many TCP streams
  - New circuit is created once a minute

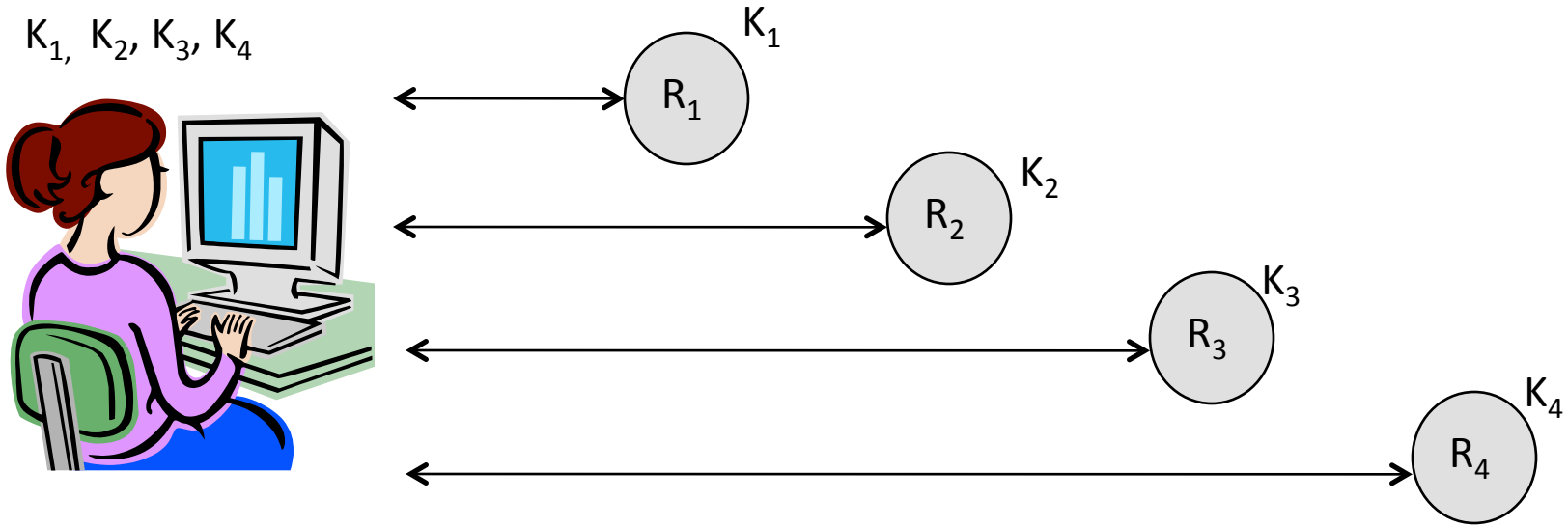


Creating circuits



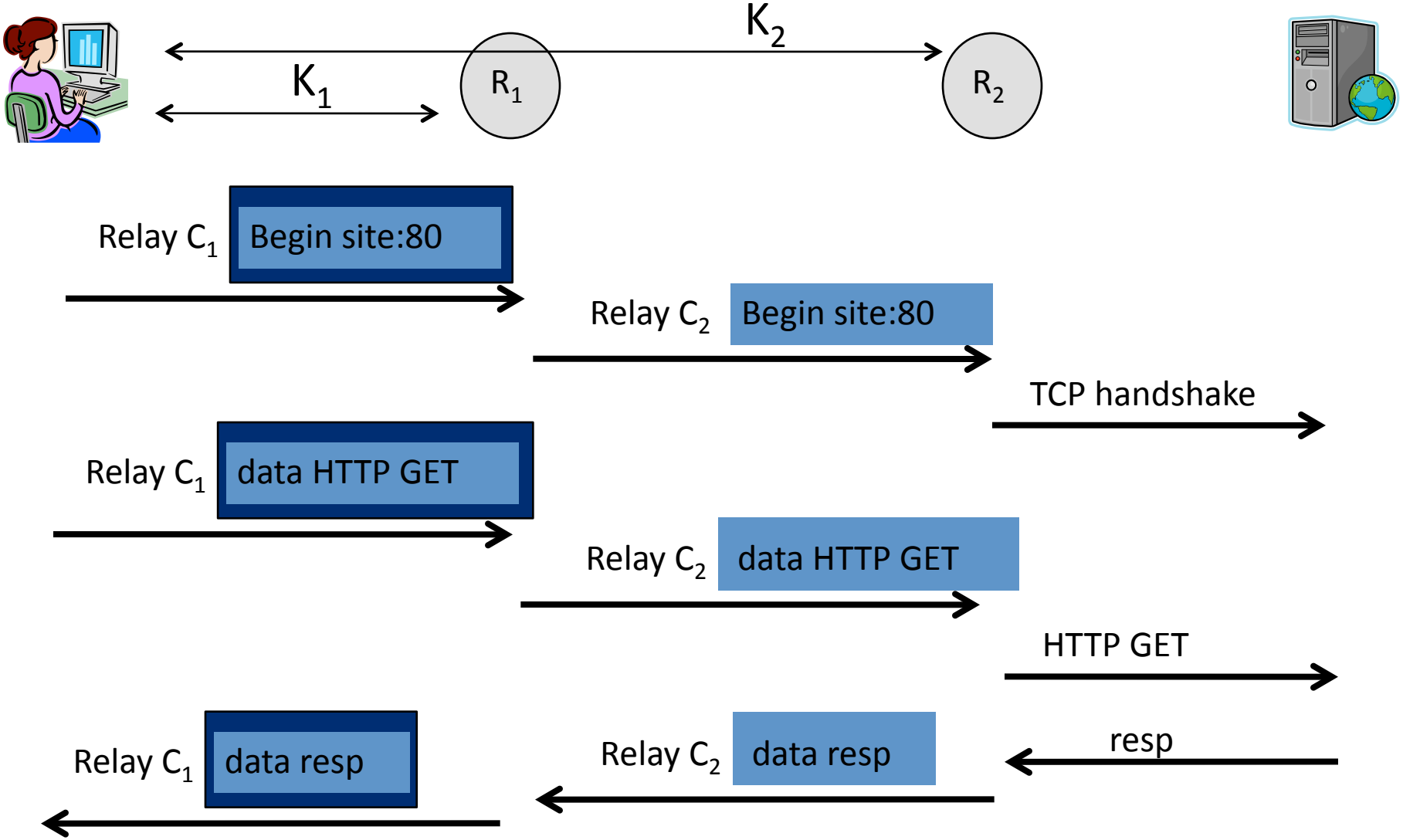


Once circuit is created



- User has shared key with each router in circuit
- Routers only know ID of successor and predecessor

Sending data



Properties

- Performance:
  - Fast connection time: circuit is pre-established
  - Traffic encrypted with AES: no pub-key on traffic
- Tor crypto:
  - provides end-to-end integrity for traffic
  - Forward secrecy via TLS
- Downside:
  - Routers must maintain state per circuit
  - Each router can link multiple streams via `CircuitID`
    - all steams in one minute interval share same `CircuitID`

The Link Layer (and link-local threats)

Bonjour

jill yetman's iBook G4 [00:17:f2:c8:46:8e]._workstation._tcp.local

Gary Yngve's MacBook Pro._postgresql._tcp.local

Benjamin Melton's M

Chantri P

Marianne

Darin Trav

Ja

Ch

Andrew

Rosslyn Lu

Gary Yngve is a graduate student in computer science at UW. For his doctorate, to be completed this summer, he is developing interacive visualizations for biological models, ontologies, and simulations. He greatly enjoys teaching, and he has had the pleasure of instructing a course and serving as head TA several times.

Originally from Florida and Georgia, he spent his undergrad years at Georgia Tech. He now feels at home in the Pacific Northwest, undeterred by the cold and rain. In his spare time, he enjoys engaging in human-powered leave-no-trace activity in the mountains (e.g. climbing, skiing), playing the cello, cooking, and writing about himself in the third-person.

Acquis
Marian
mkedla
Wester

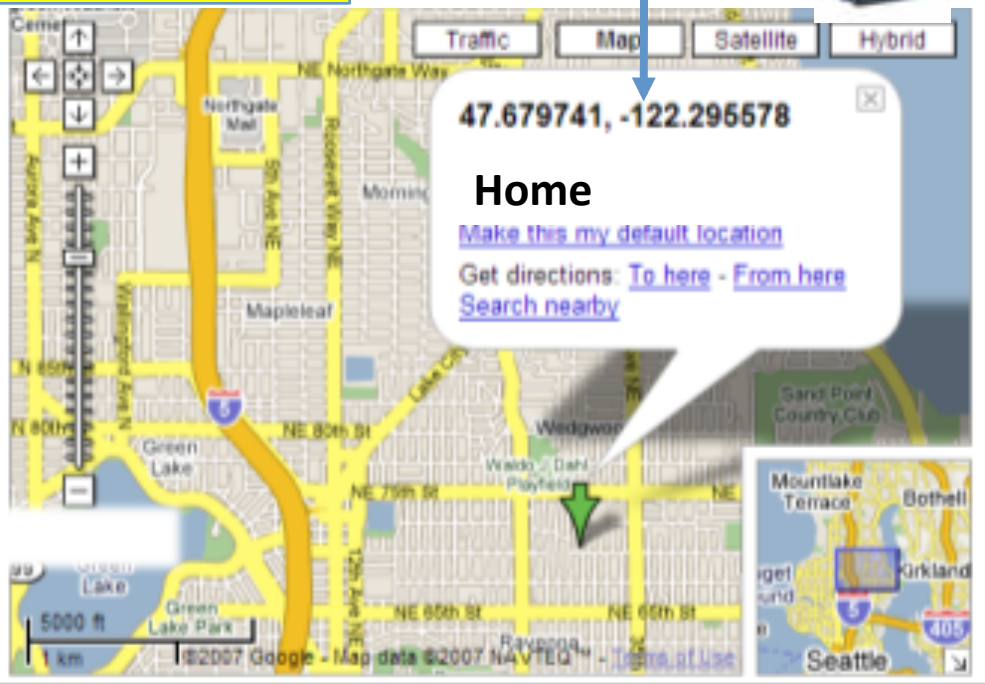
eck
she
3,
of

Processes) | (roy@ee)

Analogous privacy issues with link layer service discovery

802.11 header	Is “djw” here?	“djw” is here

www.wigle.net



Problem: Anonymity compromised by MAC Addresses

802.11 beacon	Alice's iPod is here
MAC: 12:34:56:78:90:ab	

802.11 beacon	Alice's iPod is here
MAC: 12:34:56:78:90:ab	

802.11 probe	Is Alice's iPod here?

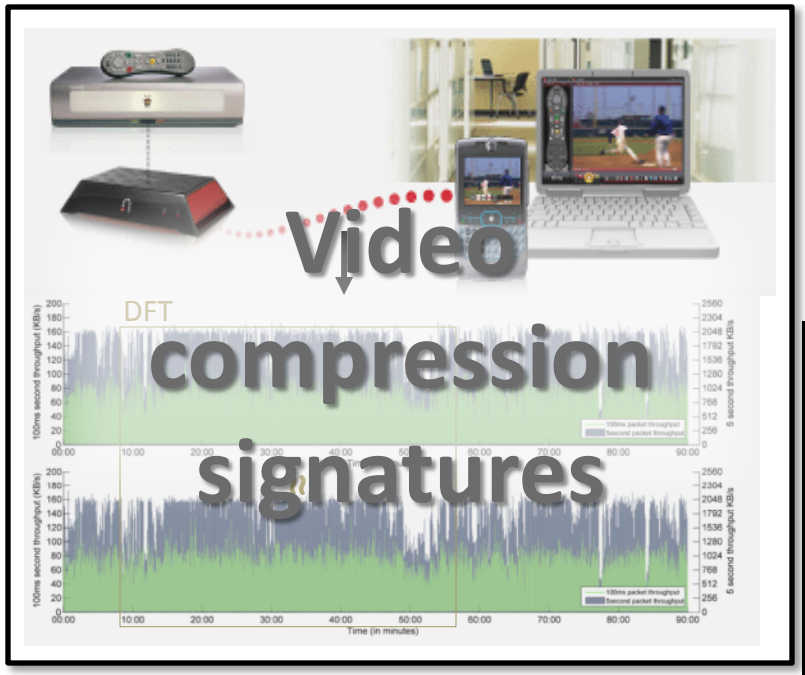
Easy to identify and relate devices over time

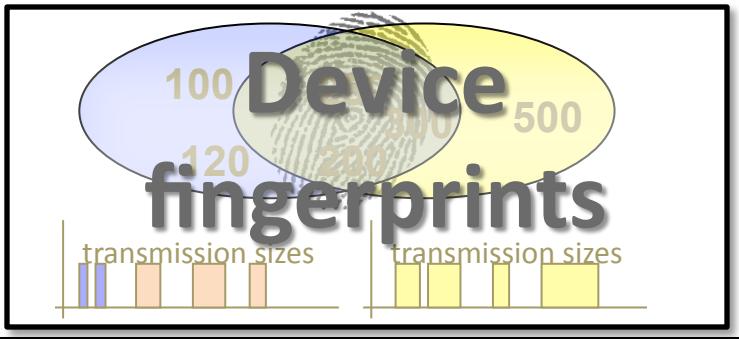
Problem: Side channel analysis via MAC address linkages

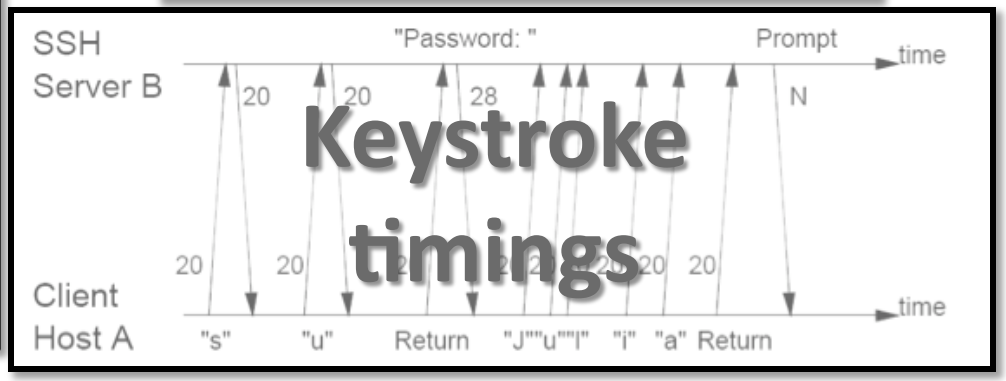
Isolated data streams are more susceptible to side-channel analysis on packet sizes and timing

– Exposes keystrokes, VoIP calls, webpages, movies, …

[Liberatore, *CCS* '06; Pang, *MobiCom* '07; Saponas, *Usenix Security* '07; Song, *Usenix Security* '01; Wright, *IEEE S&P '08*; Wright, *Usenix Security* '07]







SlyFi: A link layer with better privacy

- Why?
  - Easy for eavesdropper to track and profile users via their wireless transmissions
- Two goals:
  - Conceal all identifiers from 3rd parties
  - Make the protocol efficient
  - http://www.seattle.intel-research.net/pubs/mobisys08-slyfi.pdf

Best Security Practices

Bootstrap

Username: Alice
Key: 0x348190…

SSID: Bob's Network
Key: 0x2384949…

Out-of-band (e.g., password, WiFi Protected Setup)

Discover

802.11 probe	Is Bob's Network here?
802.11 beacon	Bob's Network is here

Authenticate and Bind

802.11 auth	Proof that I'm Alice
802.11 auth	Proof that I'm Bob

Send Data

802.11 header	
802.11 header	

- Confidentiality
- Authenticity
- Integrity

Privacy Problems Remain

Many exposed bits are (or can be used as) identifiers that are linked over time

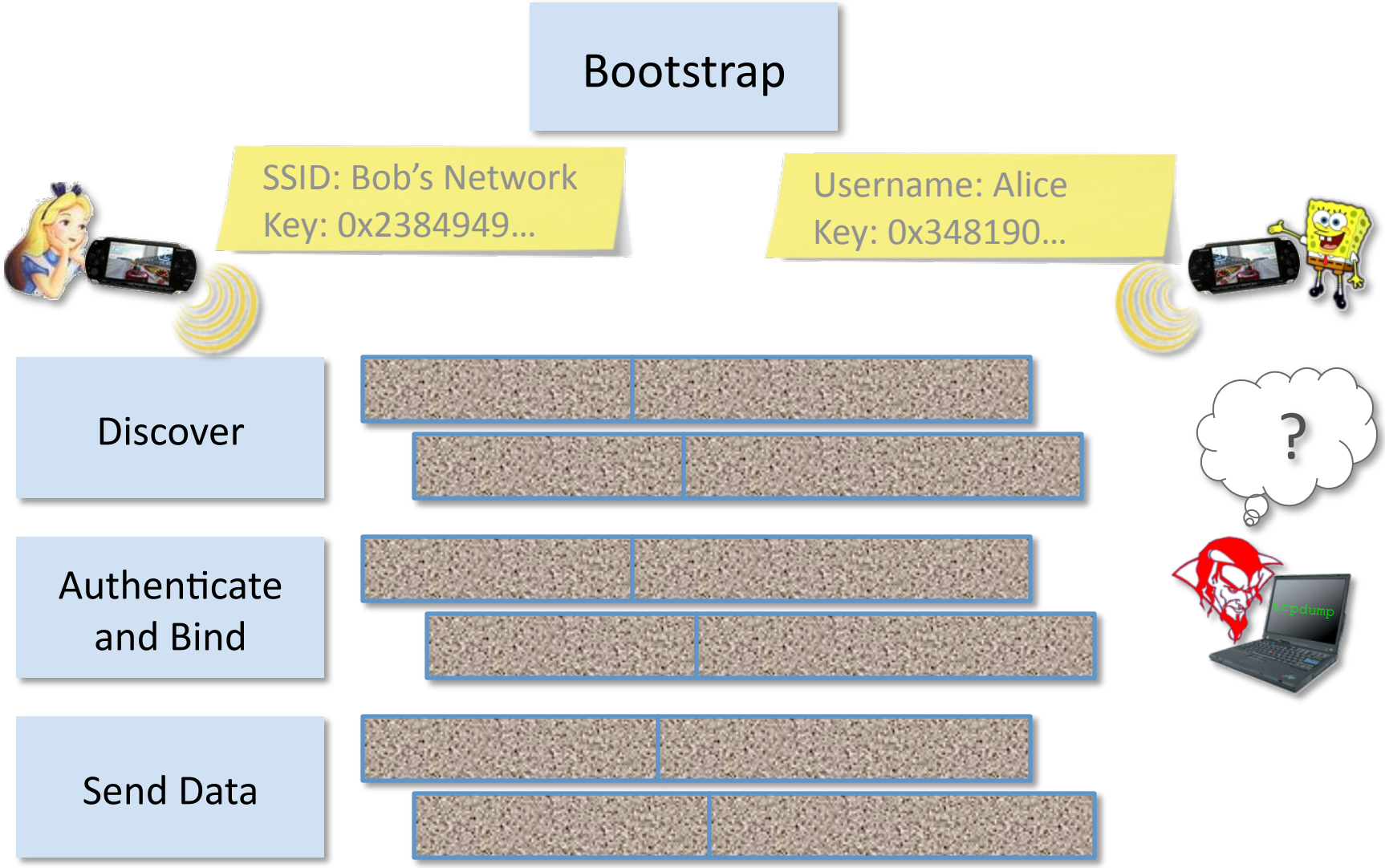
Secret: 0x2384949…

Secret: 0x348190…

Discover	802.11 probe	Is Bob's Network here?
	802.11 beacon	Bob's Network is here
Authenticate and Bind	802.11 auth	Proof that I'm Alice
	802.11 auth	Proof that I'm Bob
Send Data	MAC addr, seqno, …	
	MAC addr, seqno, …	

- Confidentiality
- Authenticity
- Integrity

Goal: all bits appear random to outsiders



Challenge is to make the protocol work when all bits are hidden

Which packets are mine?

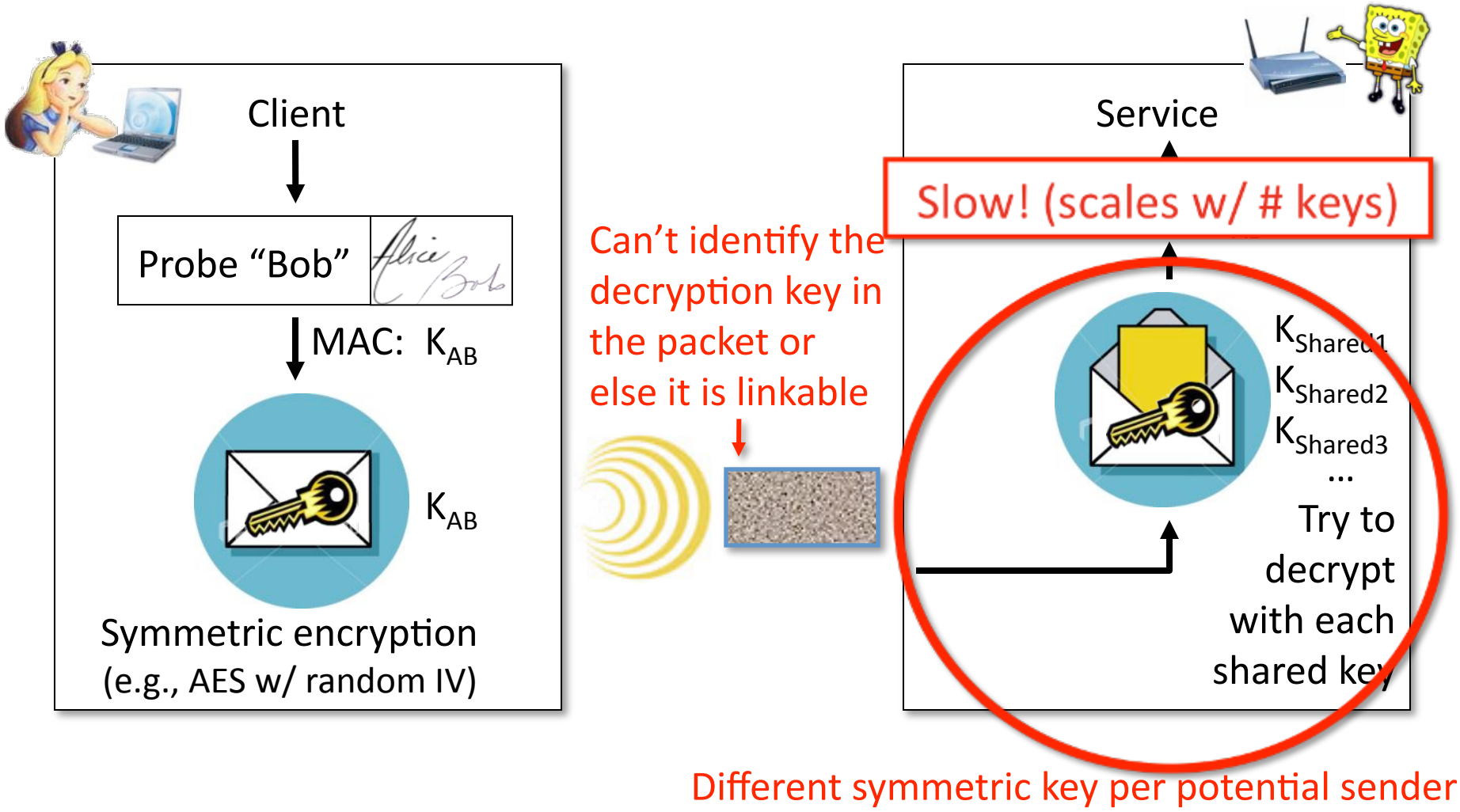
Which packets are mine?

- Filtering without Identifiers
- Without changing the usage model
- Without breaking services
- Without changing authentication
- While staying just as efficient

Straw man: MAC Pseudonyms

- **Idea**: change MAC address periodically
  - Per session or when idle [Gruteser '05, Jiang '07]
- Other fields remain (e.g., in discovery/binding)
  - No mechanism for data authentication/encryption
  - Doesn't hide network names during discovery or credentials during authentication
- Pseudonyms are linkable in the short-term
  - Same MAC must be used for each association
  - Data streams still vulnerable to side-channel leaks

Naïve approach (symmetric encryption of all bits) is slow



SlyFi insight: split encryption to provide "one-time addresses"

- Symmetric key almost works, but tension between:
  - Unlinkability: can't expose the identity of the key
  - Efficiency: need to identify the key to avoid trying all keys

- **Idea**: Split the encryption to identify the key in an unlinkable way. Provides "one-time addresses" that can be pre-computed for efficient matching

- Approach:
  - Sender **A** and receiver **B** agree on tokens: T_1^{AB}, T_2^{AB}, T_3^{AB}, …
  - **A** attaches T_i^{AB} to encrypted packet for **B**

SlyFi "one time addresses"

Client

Service

Need a shared variable, *i*, that changes often

Main challenge:
Sender and receiver must synchronize *i* without communication

K_{AB}

T_i^{AB}

Lookup T_i^{AB} in a table to get K_{AB}

Symmetric encryption
(e.g., AES w/ random IV)

$T_i^{AB} = AES_{K_{AB}}(i)$

$T_i^{AB} = AES_{K_{AB}}(i)$

Data Transport

= transmission #

- Only sent over established connections
- Expect messages to be delivered

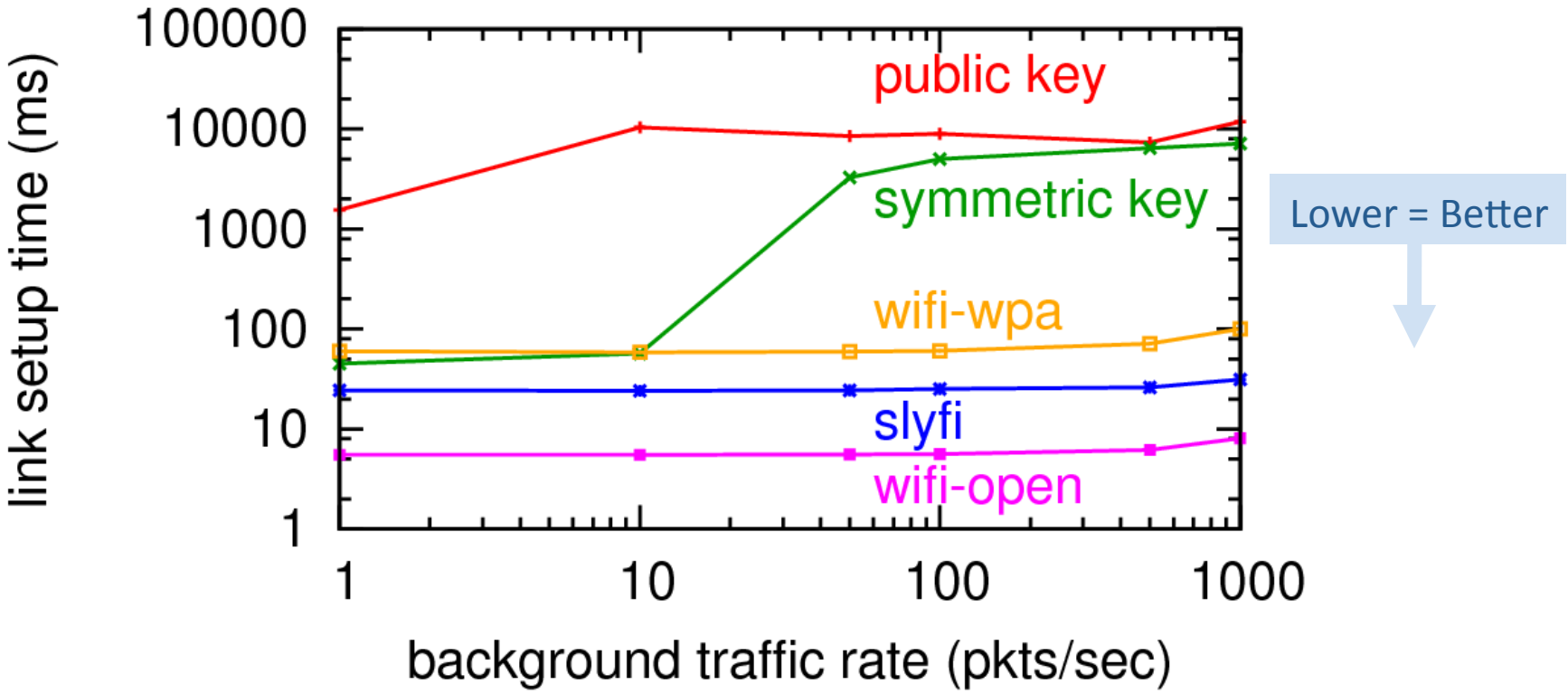
Discovery and Binding

= ⌊current time/5 min⌋

- **Infrequent**: sent when trying to associate
- **Narrow interface**: single application, few side-channels
- Linkability OK at short timescales

- On receipt of $T_i^{AB} = AES_{K_{AB}}(i)$, receiver computes T_{i+1}^{AB}
- Handling message loss or clock skew:
  - On receipt of T_i^{AB} save $T_{i+1}^{AB}, \ldots, T_{i+k}^{AB}$ in table
  - Tolerates k consecutive losses or skew of 5 * k minutes
  - No loss ⇒ compute one token per reception

Prototype Discovery/Binding Time



SlyFi link setup has less overhead than WPA

Epilogue: You

- You won't realize privacy is important until you're ruined by the release of personal information
  - Drunk Rob on MySpace