CSE 484 / CSE M 584: Computer Security and Privacy

Anonymity and Secure Messaging

Fall 2016

Ada (Adam) Lerner
lerner@cs.washington.edu

Thanks to Franzi Roesner, Dan Boneh, Dieter Gollmann, Dan Halperin, Yoshi Kohno, John Manferdelli, John Mitchell, Vitaly Shmatikov, Bennet Yee, and many others for sample slides and materials ...

Cookies

- Alternative/additional technology:
  - Ice cream

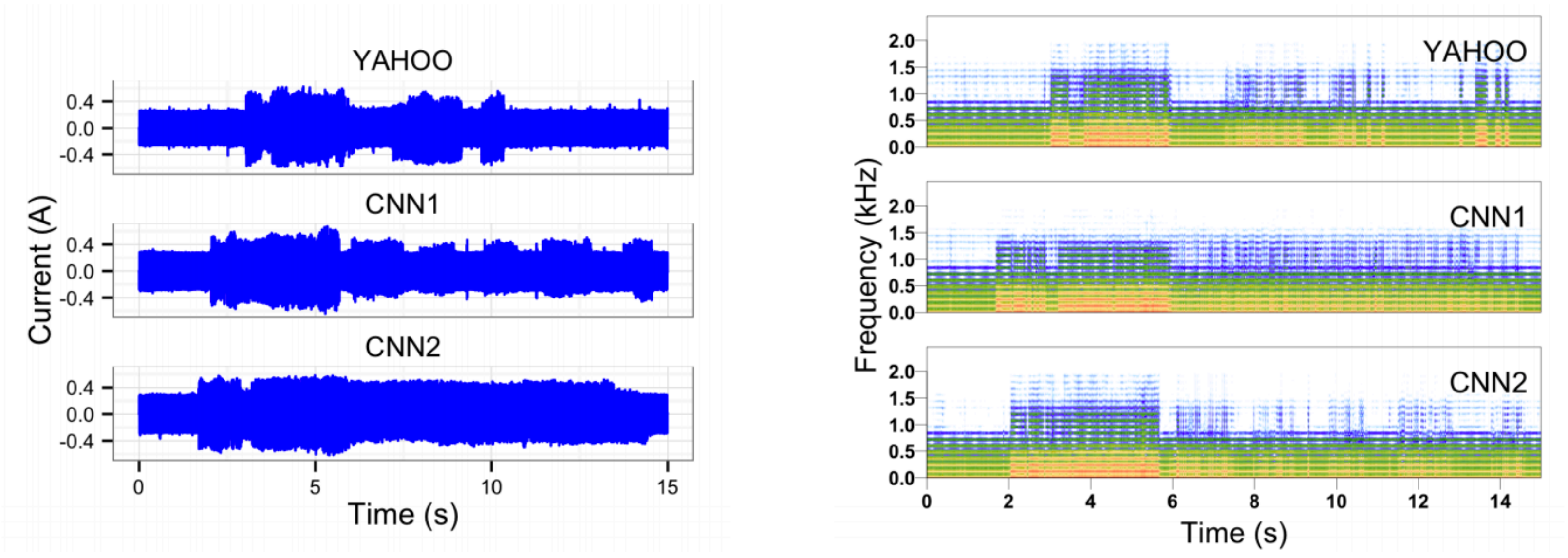
- Some of you asked if we could study these technologies

Cookies

- Section is cancelled, but:
- During section, we'll have a special culinary seminar on the topic of "Delectable Technology"

Security Mindsetish – Reflections on Trusting Trust

Identifying Web Pages: Electrical Outlets



(a) Time-domain plots

(b) Spectrogram plots

Fig. 1: Time- and frequency-domain plots of several power traces as a MacBook loads two different pages. In the frequency domain, brighter colors represent more energy at a given frequency. Despite the lack of obviously characteristic information in the time domain, the classifier correctly identifies all of the above traces.

Clark et al. "Current Events: Identifying Webpages by Tapping the Electrical Outlet" ESORICS 2013

Powerline Eavesdropping

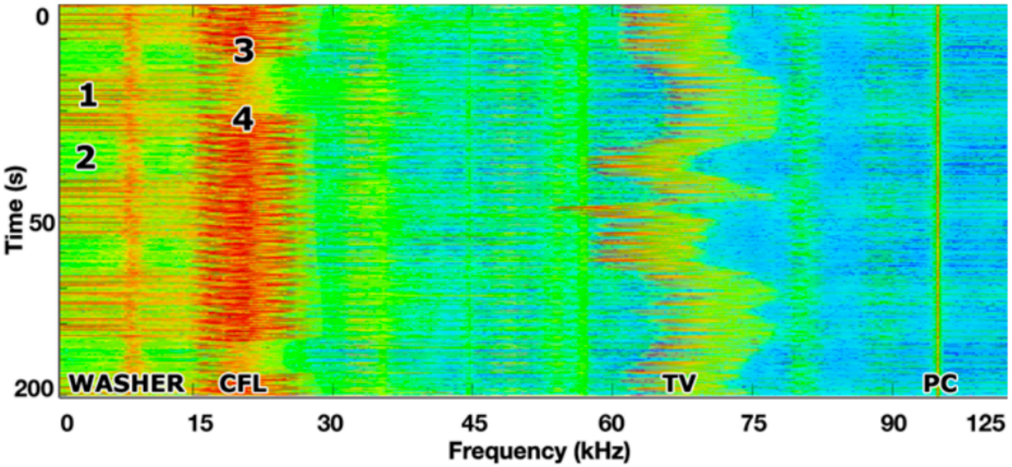


Figure 1: Frequency spectrogram showing various electrical appliances in the home. Washer cycle on (1) and off (2). CFL lamp turning off briefly (3) and then on (4). Note that the TV's (Sharp 42" LCD) EMI shifts in frequency, which happens as screen content changes.

Enev et al.: Televisions, Video Privacy, and Powerline Electromagnetic Interference, CCS 2011

Privacy on Public Networks

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

Questions

Q1: What is anonymity?

Q2: Why might people want anonymity on the Internet?

Q3: Why might people **not** want anonymity on the Internet?

Applications of Anonymity (I)

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

Applications of Anonymity (II)

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

What is Anonymity?

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

Part 1: Anonymity in Datasets

How to release an anonymous dataset?

- Possible approach: remove identifying information from datasets?

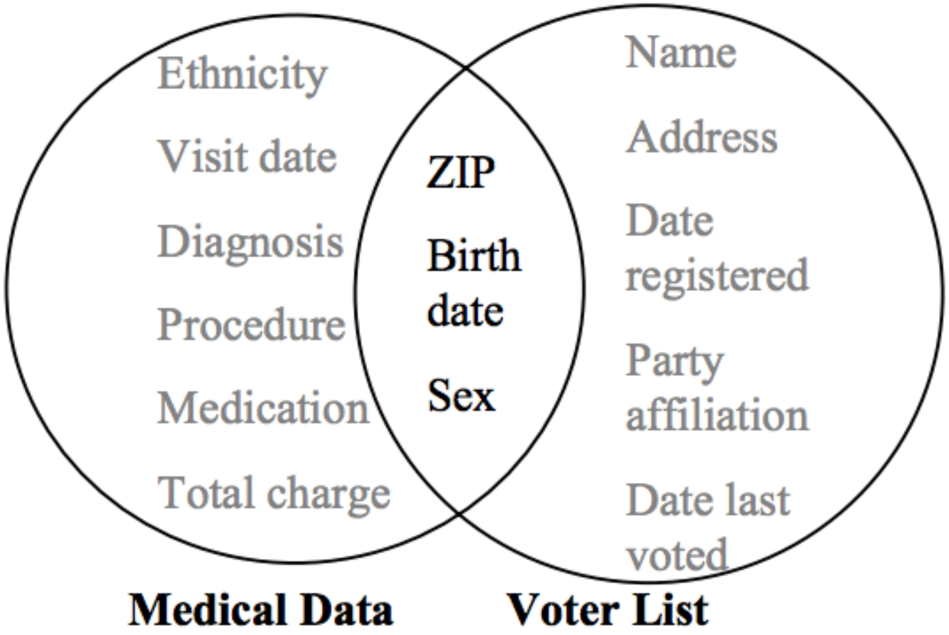
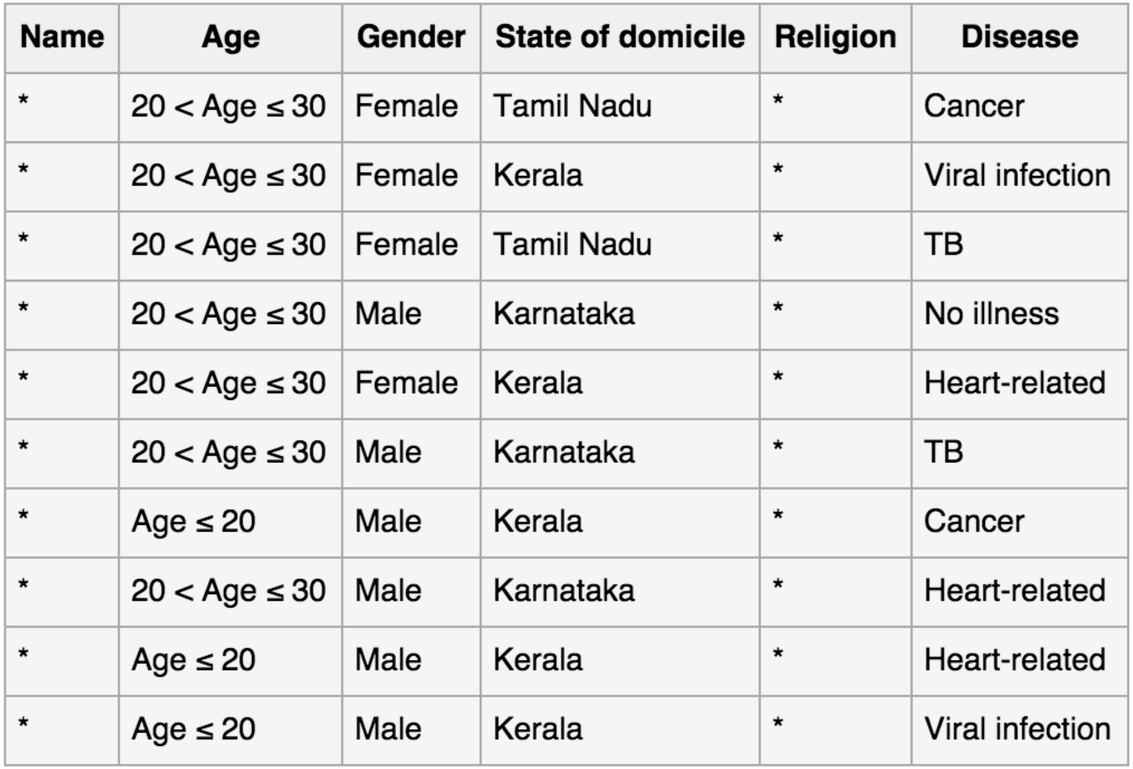


Figure 1 Linking to re-identify data

Massachusetts medical+voter data [Sweeney 1997]

k-Anonymity

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



Name	Age	Gender	State of domicile	Religion	Disease
*	20 < Age ≤ 30	Female	Tamil Nadu	*	Cancer
*	20 < Age ≤ 30	Female	Kerala	*	Viral infection
*	20 < Age ≤ 30	Female	Tamil Nadu	*	TB
*	20 < Age ≤ 30	Male	Karnataka	*	No illness
*	20 < Age ≤ 30	Female	Kerala	*	Heart-related
*	20 < Age ≤ 30	Male	Karnataka	*	TB
*	Age ≤ 20	Male	Kerala	*	Cancer
*	20 < Age ≤ 30	Male	Karnataka	*	Heart-related
*	Age ≤ 20	Male	Kerala	*	Heart-related
*	Age ≤ 20	Male	Kerala	*	Viral infection

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

[Dwork et al.]

Differential Privacy

- **Setting:** Trusted party has a database
- **Goal:** allow queries on the database that are useful but preserve the privacy of individual records
- **Differential privacy intuition:** add noise so that an output is produced with similar probability whether any single input is included or not
- Privacy of the computation, not of the dataset

Part 2: Anonymity in Communication

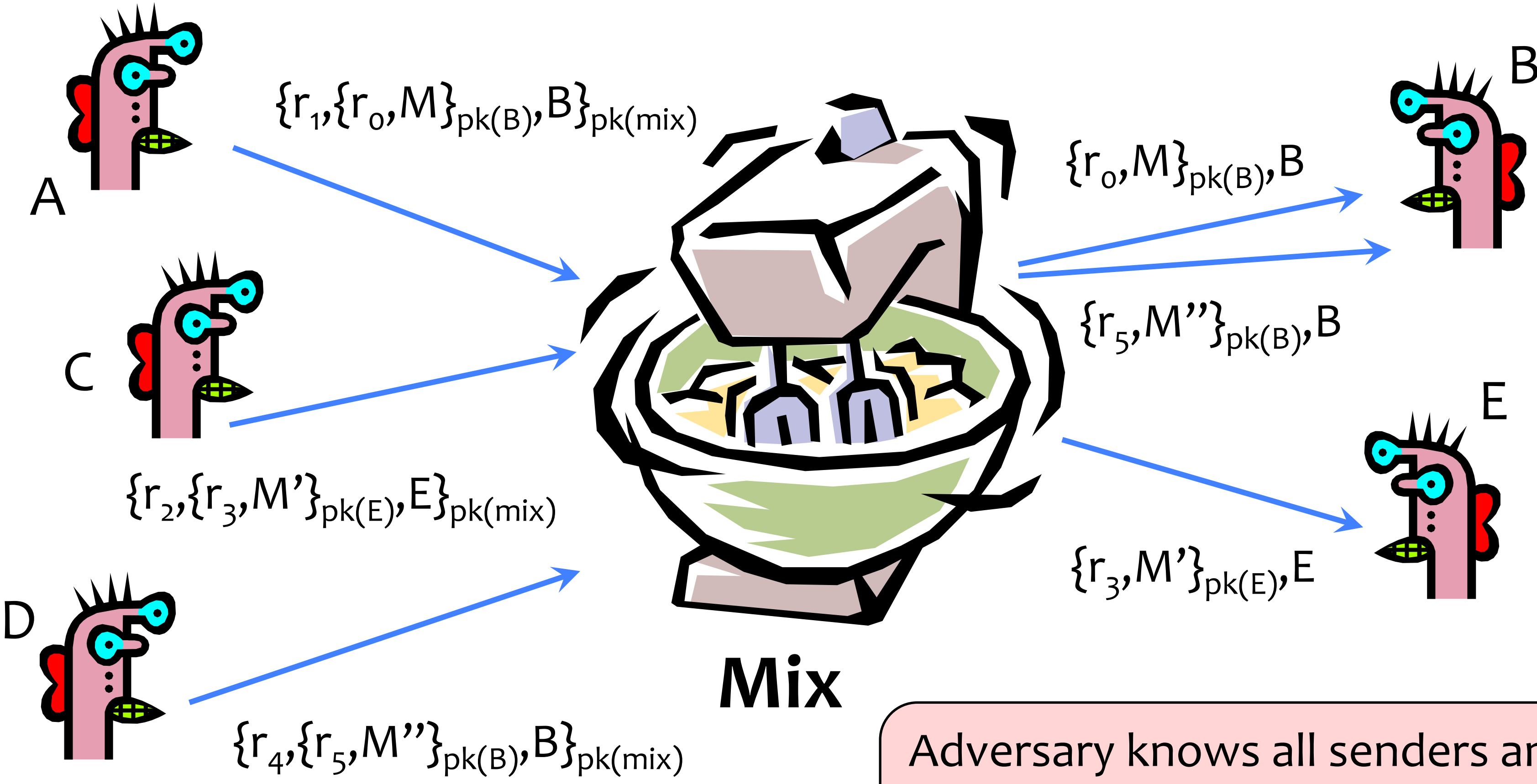
Chaum's Mix

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

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

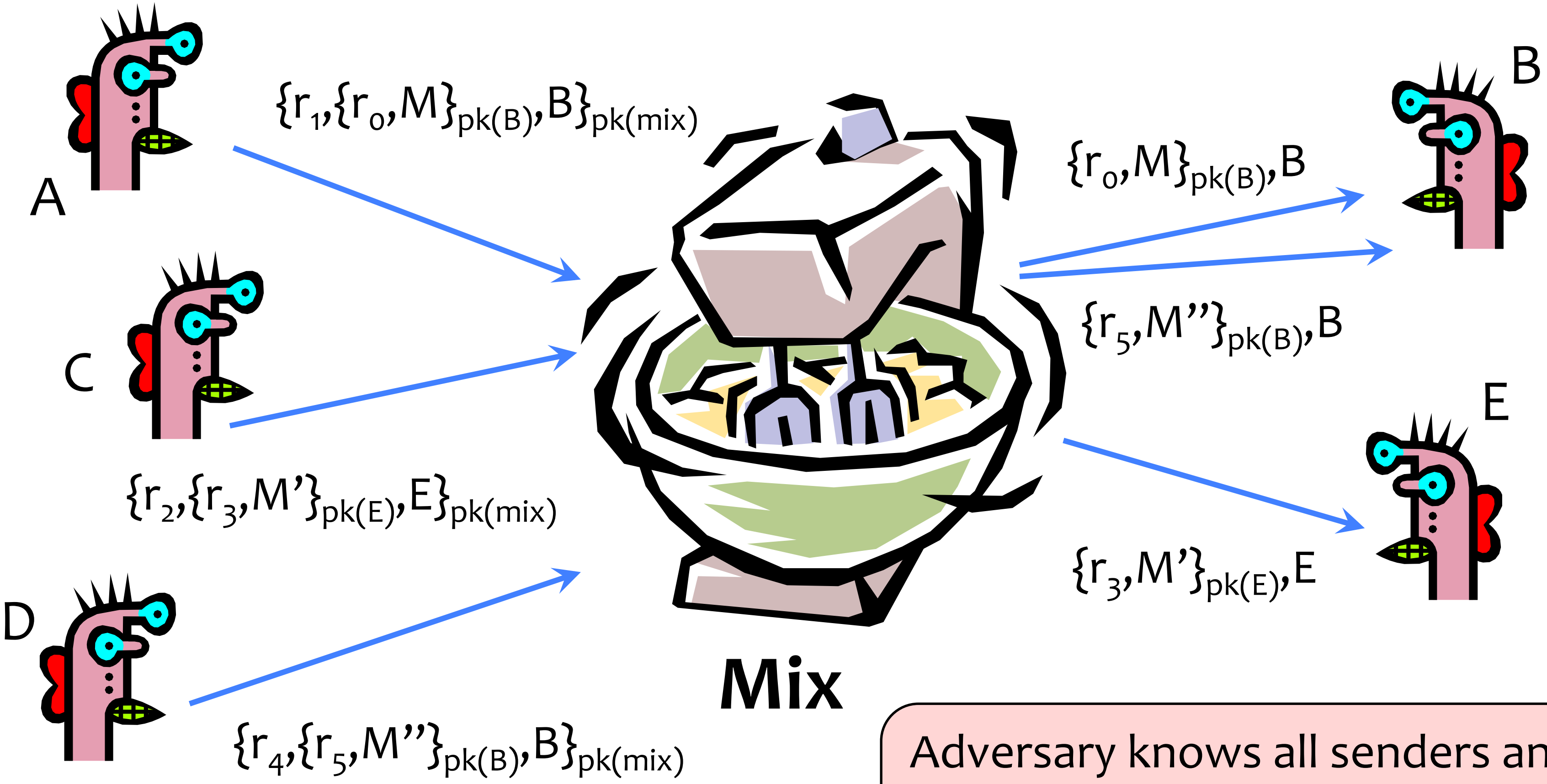
- Public key crypto + trusted re-mailer (Mix)
  - Untrusted communication medium
  - Public keys used as persistent pseudonyms
- Modern anonymity systems use Mix as the basic building block

Basic Mix Design


A
$\{r_1,\{r_0,M\}_{pk(B)},B\}_{pk(mix)}$
C
$\{r_2,\{r_3,M'\}_{pk(E)},E\}_{pk(mix)}$
D
$\{r_4,\{r_5,M''\}_{pk(B)},B\}_{pk(mix)}$
Mix
B
$\{r_0,M\}_{pk(B)},B$
$\{r_5,M''\}_{pk(B)},B$
E
$\{r_3,M'\}_{pk(E)},E$

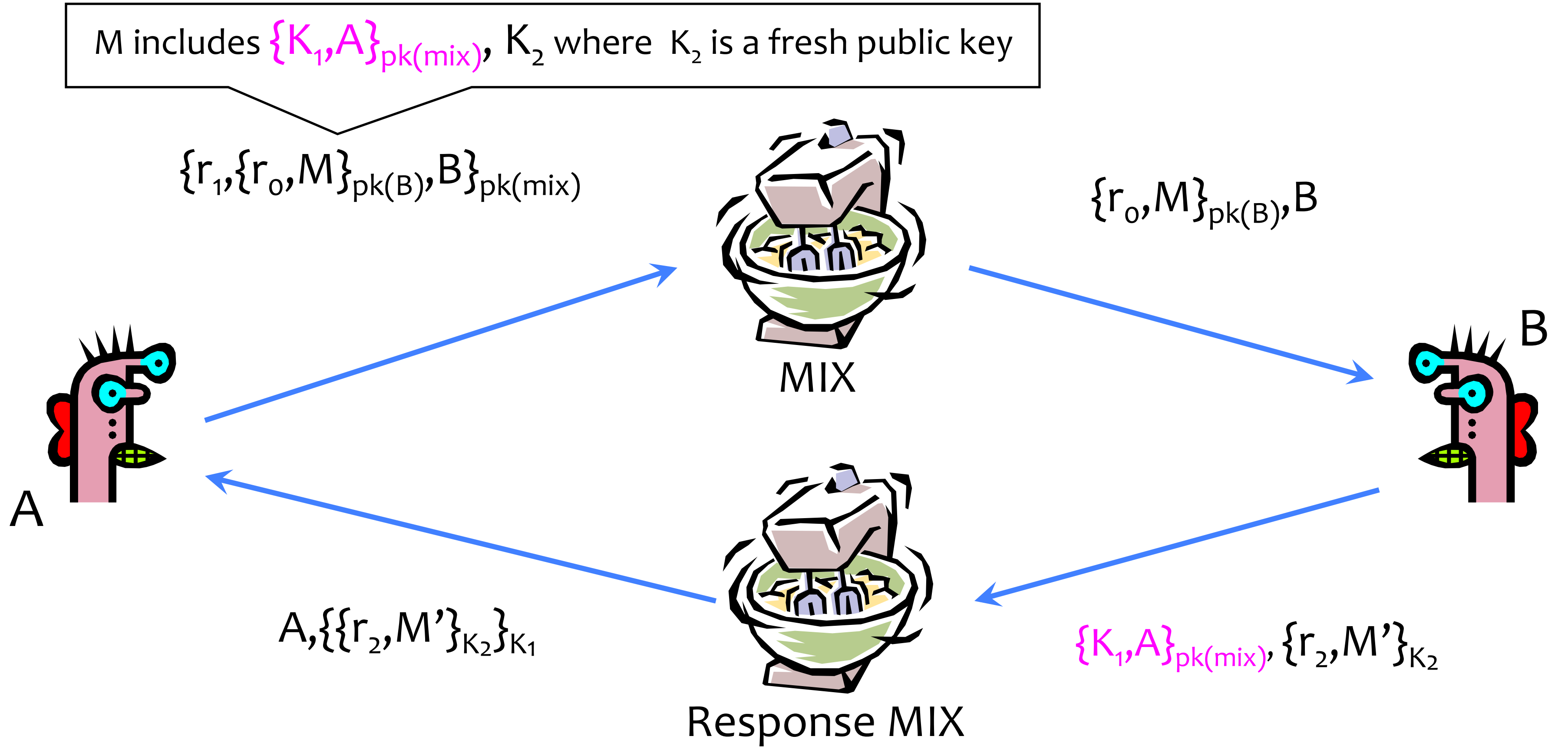
Adversary knows all senders and all receivers, but cannot link a sent message with a received message

Q2


A
$\{r_1,\{r_0,M\}_{pk(B)},B\}_{pk(mix)}$
C
$\{r_2,\{r_3,M'\}_{pk(E)},E\}_{pk(mix)}$
D
$\{r_4,\{r_5,M''\}_{pk(B)},B\}_{pk(mix)}$
Mix
B
$\{r_0,M\}_{pk(B)},B$
$\{r_5,M''\}_{pk(B)},B$
E
$\{r_3,M'\}_{pk(E)},E$

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

Anonymous Return Addresses



M includes $\{K_1,A\}_{pk(mix)}$, K_2 where K_2 is a fresh public key

$\{r_1,\{r_0,M\}_{pk(B)},B\}_{pk(mix)}$

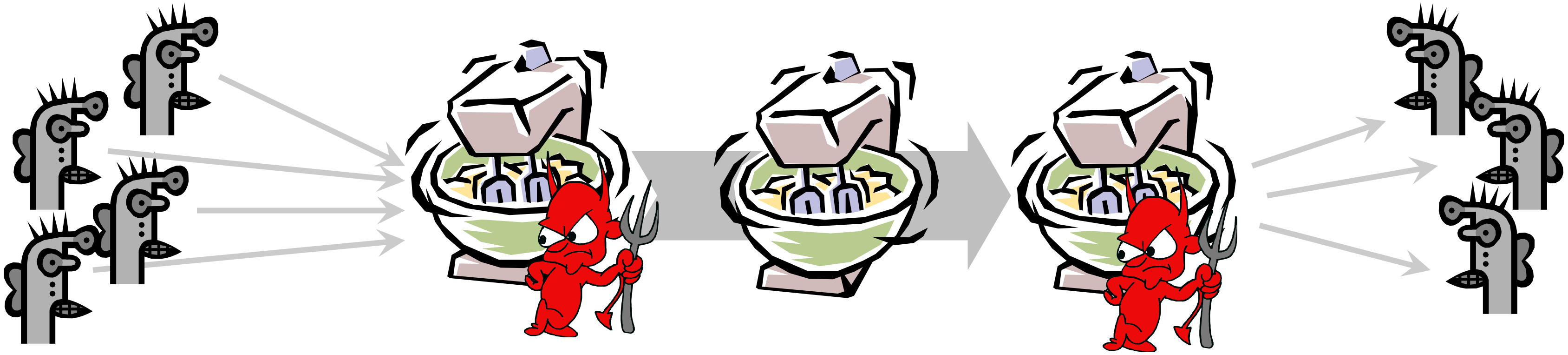
$\{r_0,M\}_{pk(B)},B$

$A,\{\{r_2,M'\}_{K_2}\}_{K_1}$

$\{K_1,A\}_{pk(mix)},\{r_2,M'\}_{K_2}$

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

Mix Cascades and Mixnets

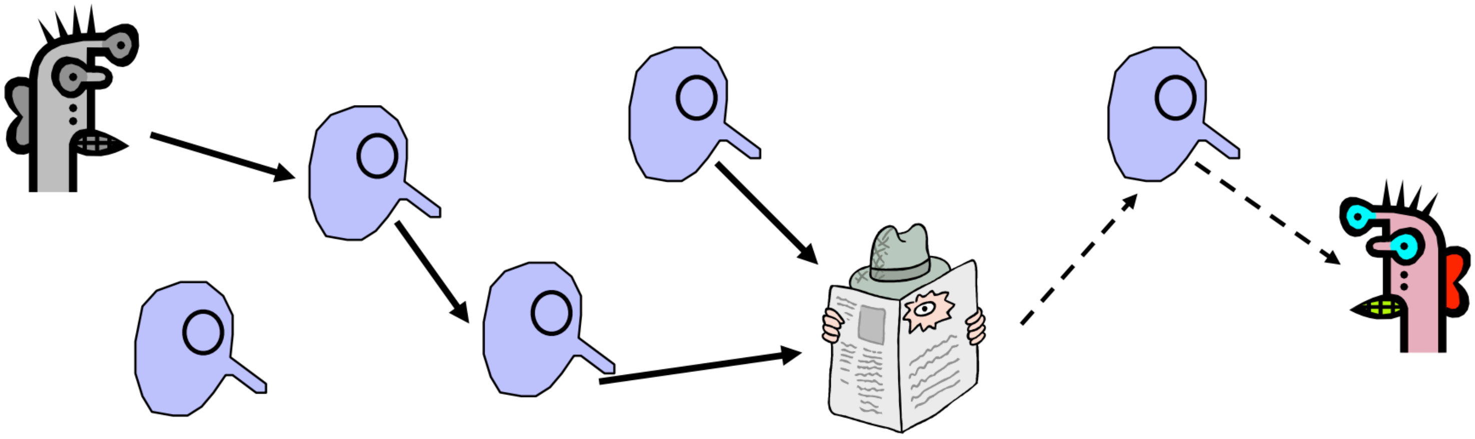


- Messages are sent through a sequence of mixes
  - Can also form an arbitrary network of mixes ("mixnet")
- Some of the mixes may be controlled by attacker, but even a single good mix ensures anonymity
- Pad and buffer traffic to foil correlation attacks

Disadvantages of Basic Mixnets

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

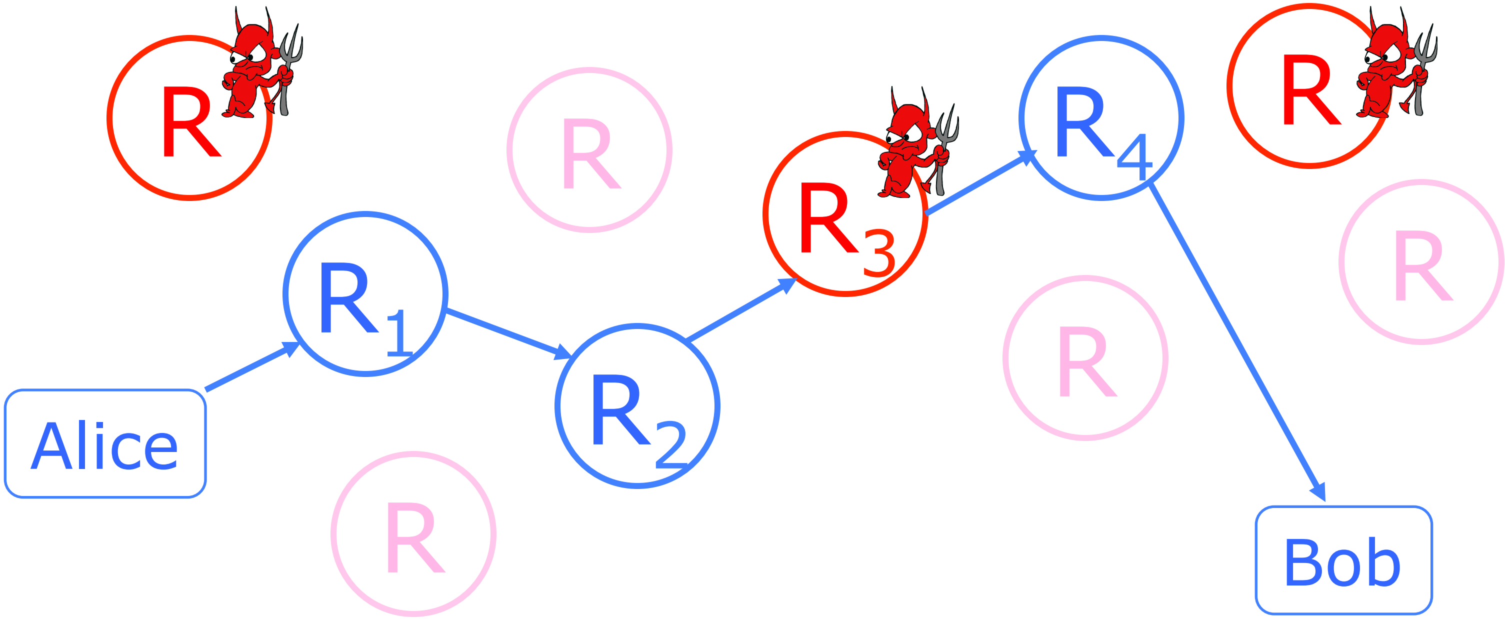
Another Idea: Randomized Routing



- Hide message source by routing it randomly
  - Popular technique: Crowds, Freenet, Onion routing
- Routers don’t know for sure if the apparent source of a message is the true sender or another router

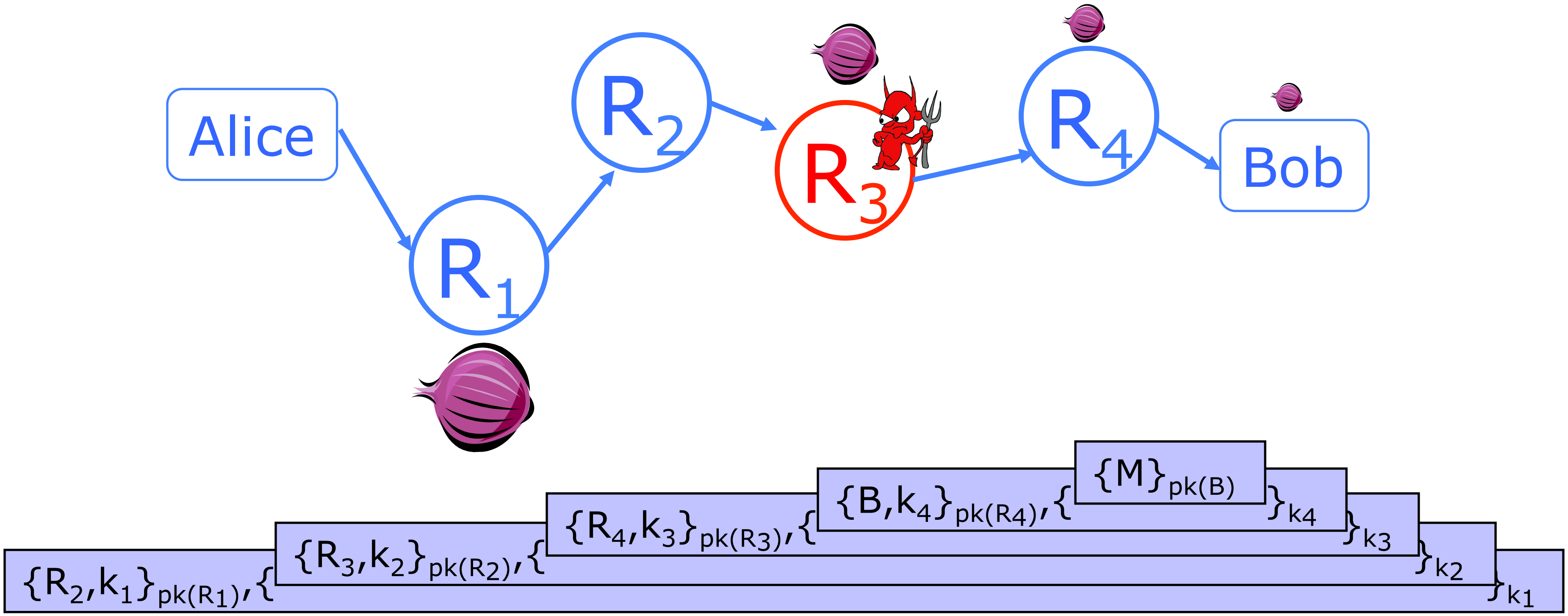
[Reed, Syverson, Goldschlag 1997]

Onion Routing



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

Route Establishment


Alice
R_1
R_2
R_3
R_4
Bob
$\{M\}_{pk(B)}$
$\{B,k_4\}_{pk(R4)},\{$ $\}_{k4}$
$\{R_4,k_3\}_{pk(R3)},\{$ $\}_{k3}$
$\{R_3,k_2\}_{pk(R2)},\{$ $\}_{k2}$
$\{R_2,k_1\}_{pk(R1)},\{$ $\}_{k1}$

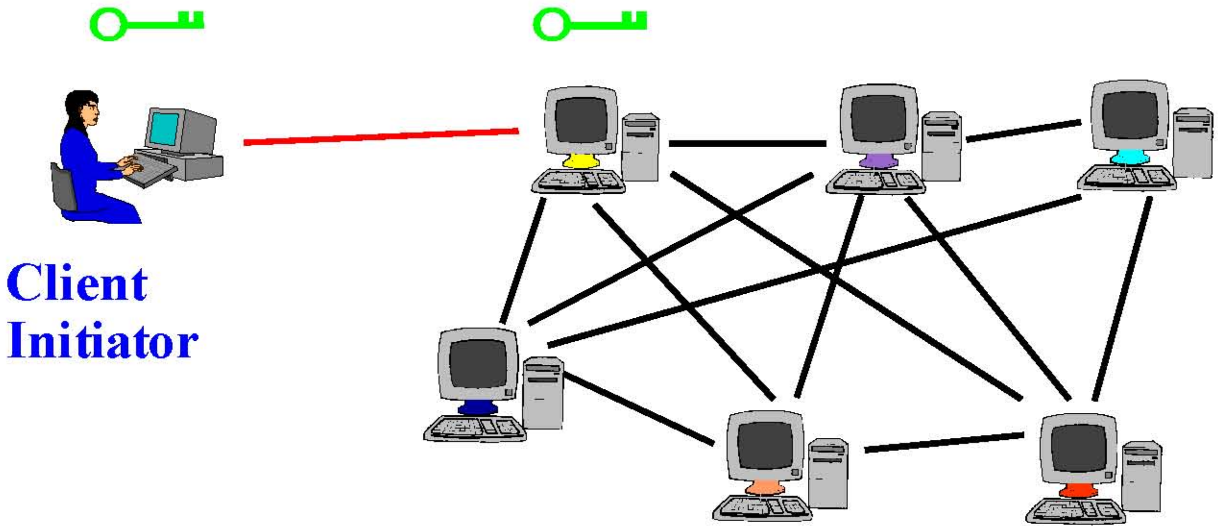
- Routing info for each link encrypted with router’s public key
- Each router learns only the identity of the next router

Tor

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

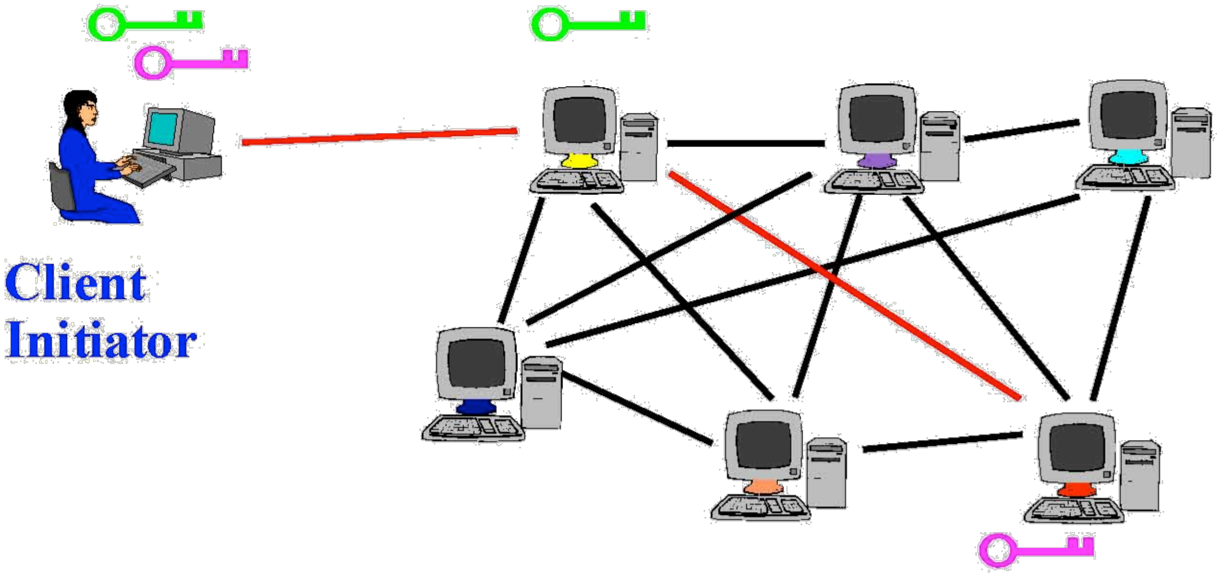
Tor Circuit Setup (1)

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



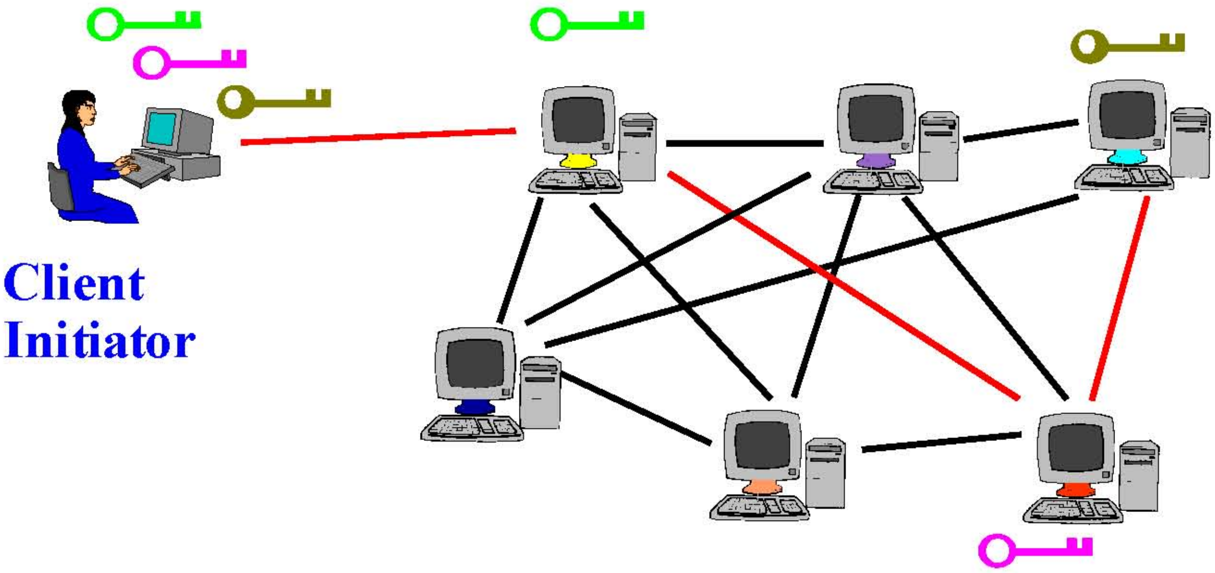
Tor Circuit Setup (2)

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



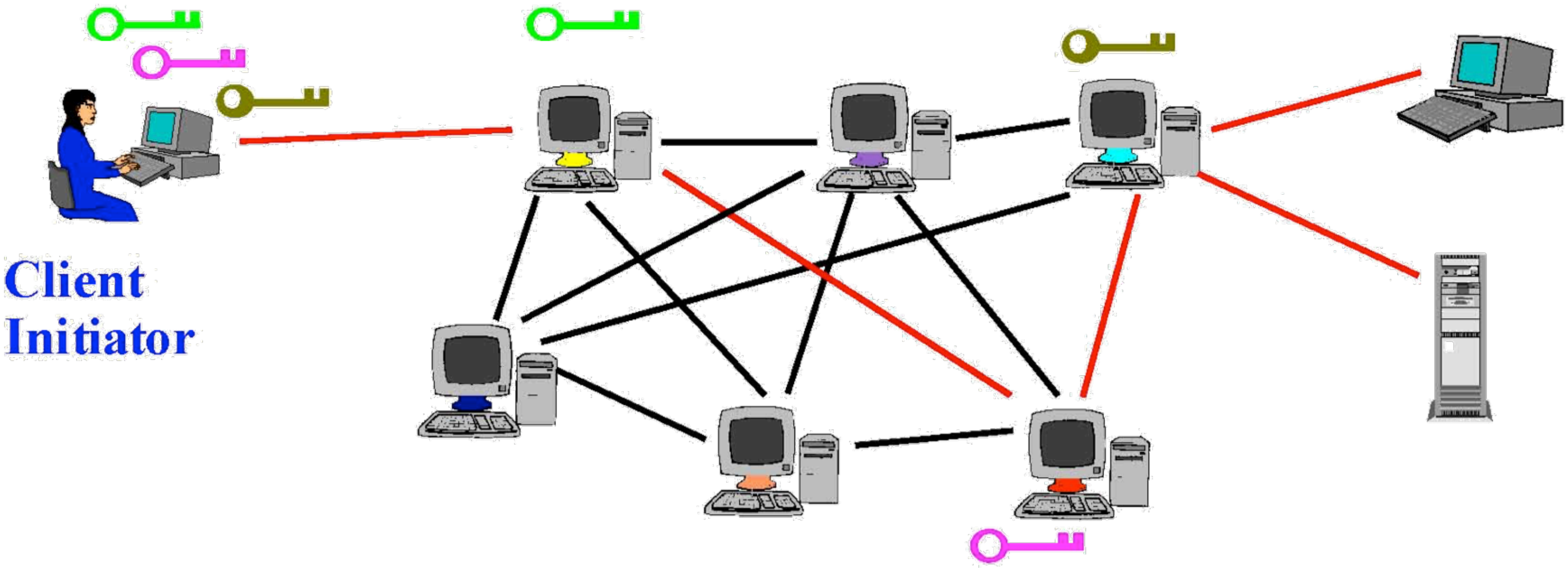
Tor Circuit Setup (3)

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



Using a Tor Circuit

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



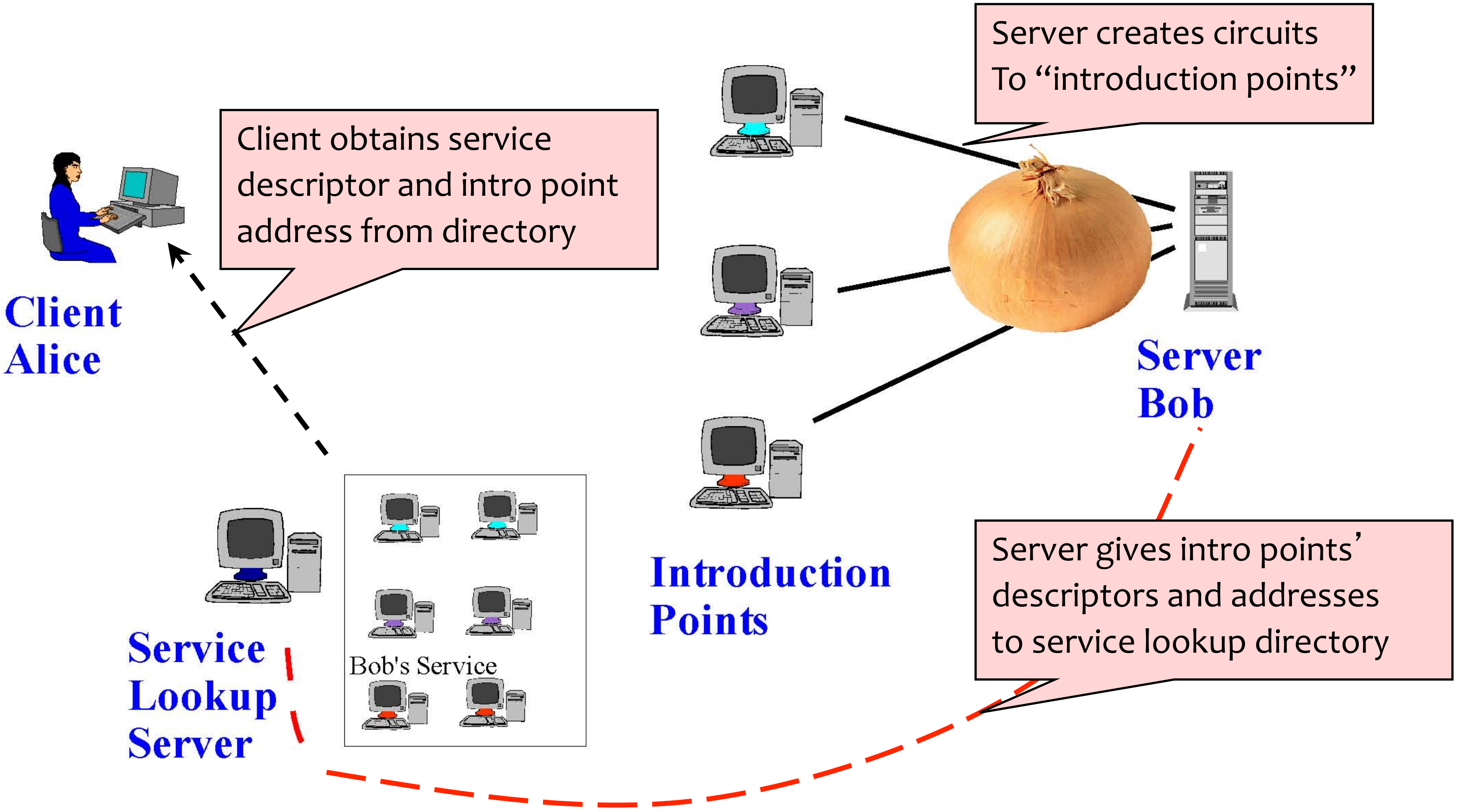
Tor Management Issues

- Many applications can share one circuit
  - Multiple TCP streams over one anonymous connection
- Tor router doesn't need root privileges
  - Encourages people to set up their own routers
  - More participants = better anonymity for everyone
- Directory servers
  - Maintain lists of active onion routers, their locations, current public keys, etc.
  - Control how new routers join the network
    - "Sybil attack": attacker creates a large number of routers
  - Directory servers' keys ship with Tor code

Location Hidden Service

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

Creating a Location Hidden Server

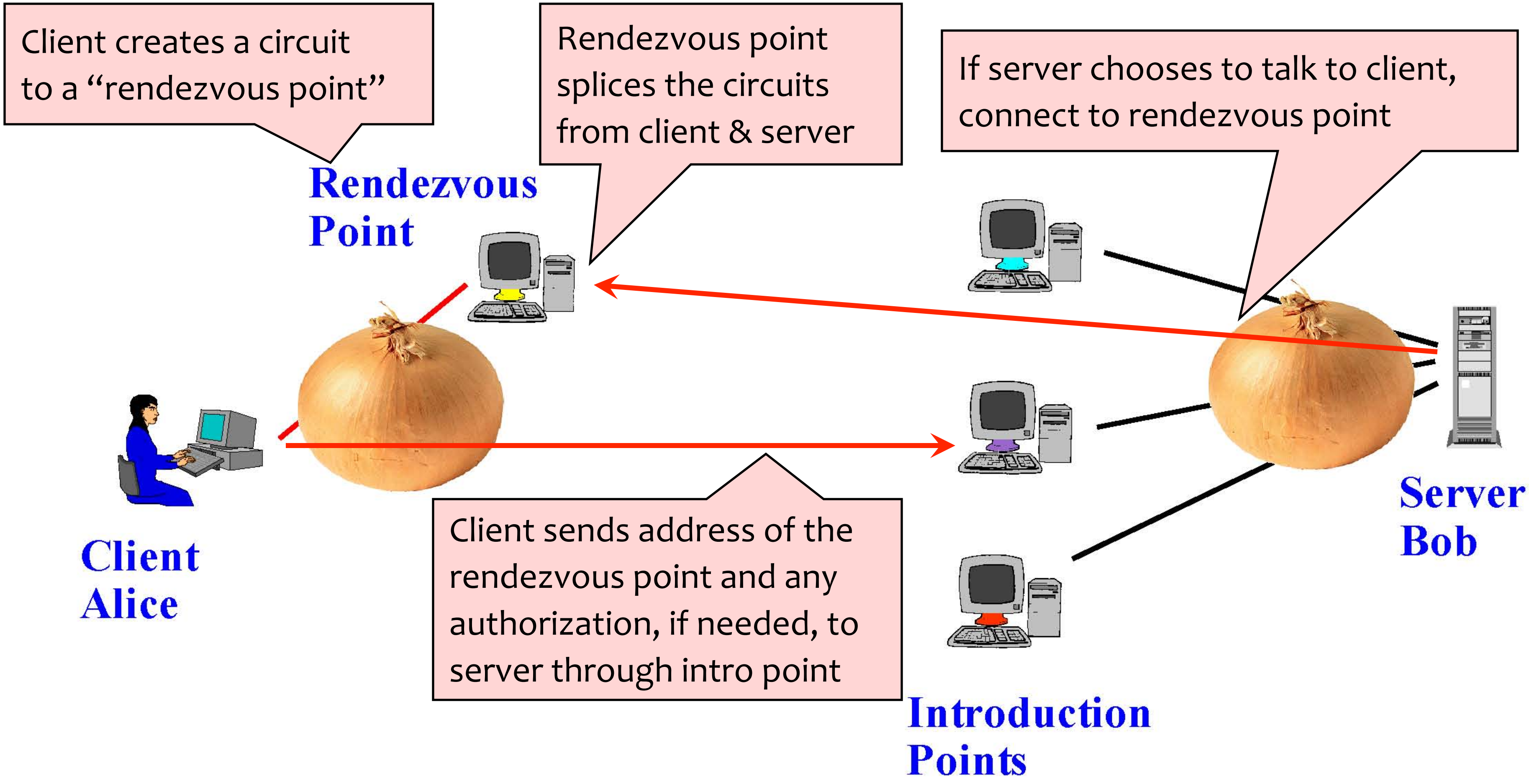


Server creates circuits To "introduction points"

Client obtains service descriptor and intro point address from directory

Server gives intro points' descriptors and addresses to service lookup directory

Using a Location Hidden Server



Client creates a circuit to a “rendezvous point”

Rendezvous point splices the circuits from client & server

If server chooses to talk to client, connect to rendezvous point

Client sends address of the rendezvous point and any authorization, if needed, to server through intro point

Attacks on Anonymity

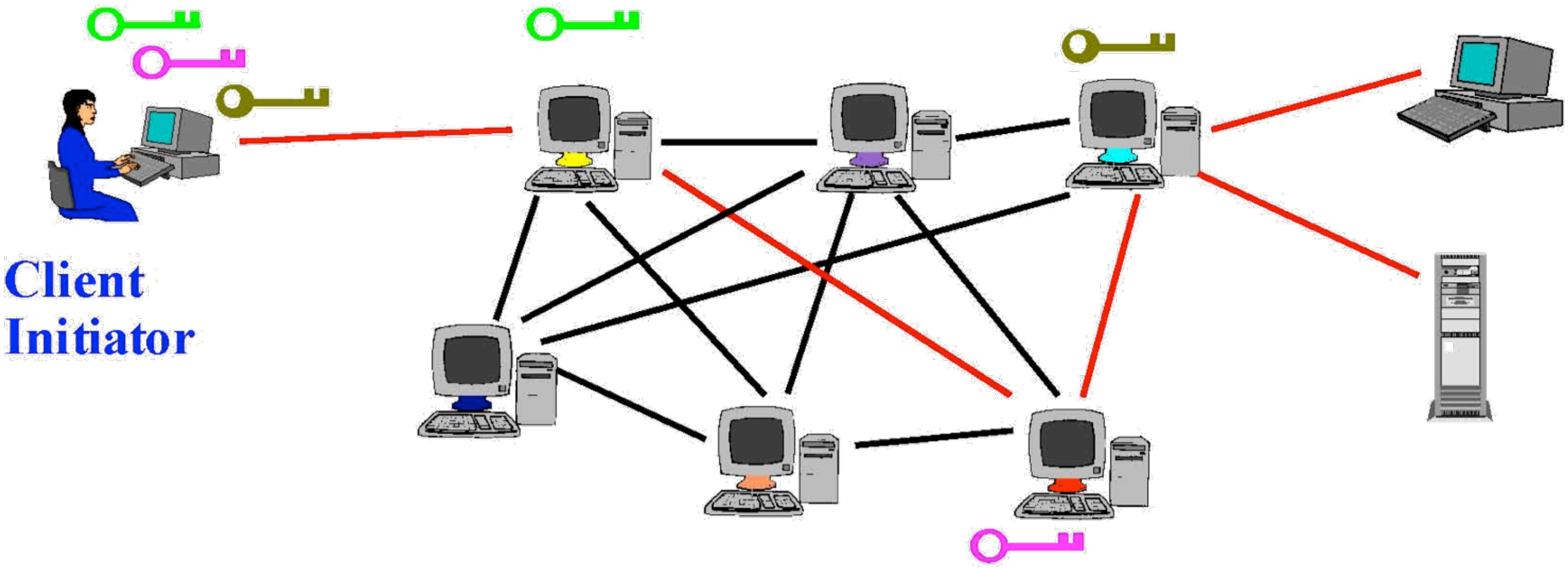
- Passive traffic analysis
  - Infer from network traffic who is talking to whom
  - To hide your traffic, must carry other people's traffic!
- Active traffic analysis
  - Inject packets or put a timing signature on packet flow
- Compromise of network nodes
  - Attacker may compromise some routers
 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

Deployed Anonymity Systems

- Tor (http://tor.eff.org)
  - Overlay circuit-based anonymity network
  - Best for low-latency applications such as anonymous Web browsing
- Mixminion (http://www.mixminion.net)
  - Network of mixes
  - Best for high-latency applications such as anonymous email
- Not: YikYak ☺

Some Caution

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



Identifying Web Pages: Traffic Analysis

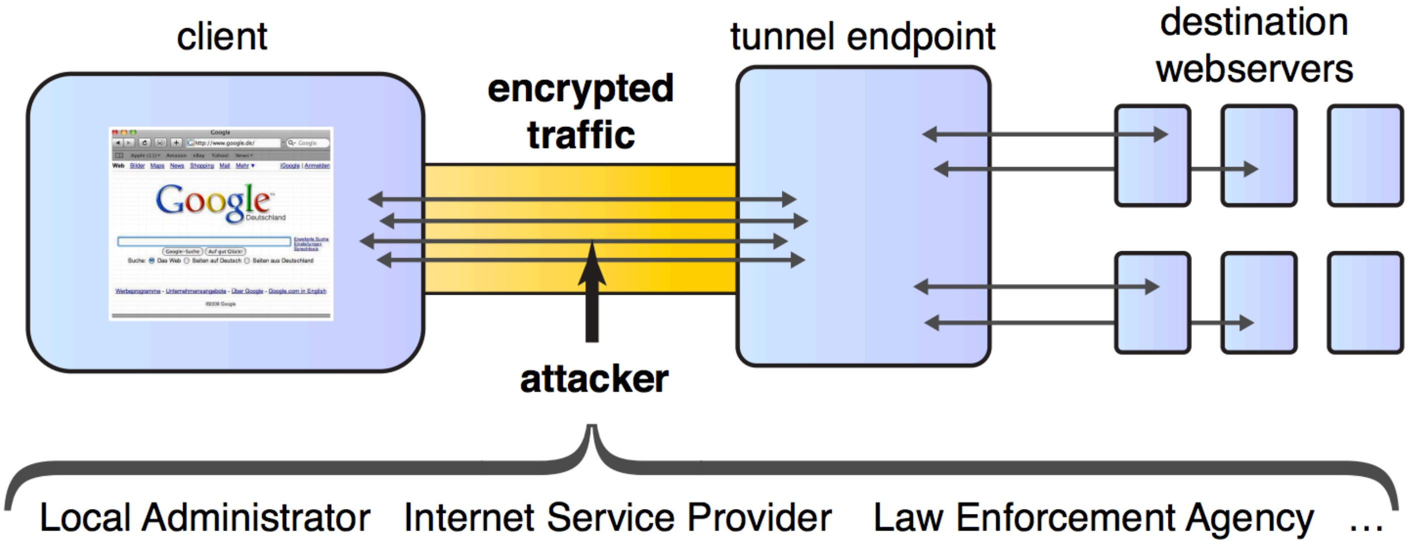


Figure 1: Website fingerprinting scenario and conceivable attackers

Herrmann et al. "Website Fingerprinting: Attacking Popular Privacy Enhancing Technologies with the Multinomial Naïve-Bayes Classifier" CCSW 2009

OTR AND SECURE MESSAGING

OTR – “Off The Record”

- Protocol for end-to-end encrypted instant messaging

- End-to-end: Only the endpoints can read messages.
  - PGP, iMessage, WhatsApp, and a variety of other services provide some form of end-to-end encryption today.

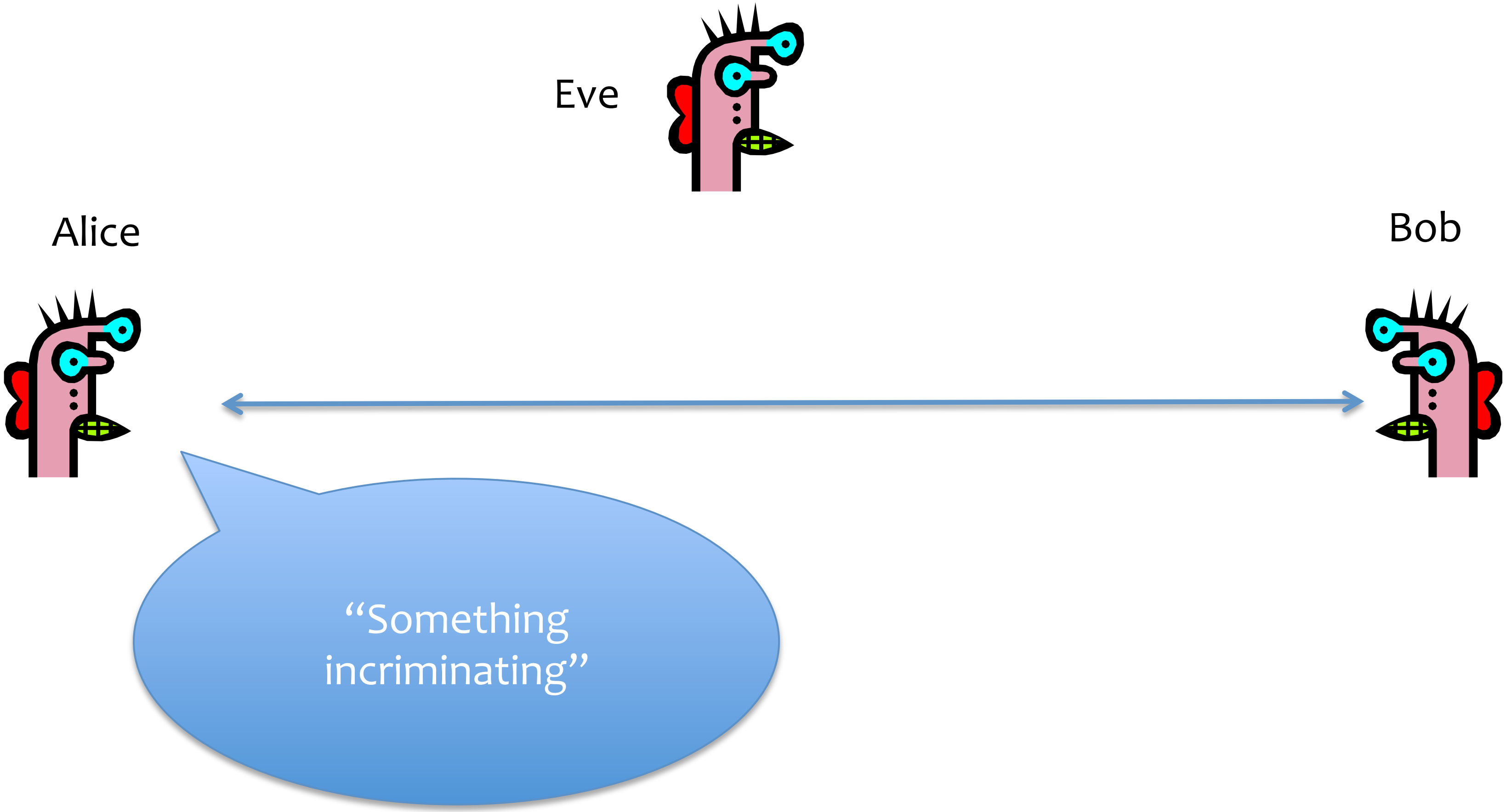
OTR – "Off The Record"

- **End-to-end encryption**
- **Authentication**
- Deniability, *after* the fact
- Perfect Forward Secrecy

OTR – “Off The Record”

- End-to-end encryption
- Authentication
- **Deniability, *after* the fact**
- **Perfect Forward Secrecy**

OTR: Deniability



Eve

Alice

Bob

“Something incriminating”

OTR: Deniability

- During a conversation session, messages are authenticated and unmodified.

- Authentication happens using a MAC derived from a shared secret.

OTR: Deniability

- During a conversation session, messages are authenticated and unmodified.

- Authentication happens using a MAC derived from a shared secret.

- Q1

OTR: Deniability

- Can’t prove the other person sent the message, because you also could have computed the MAC!

OTR: Deniability

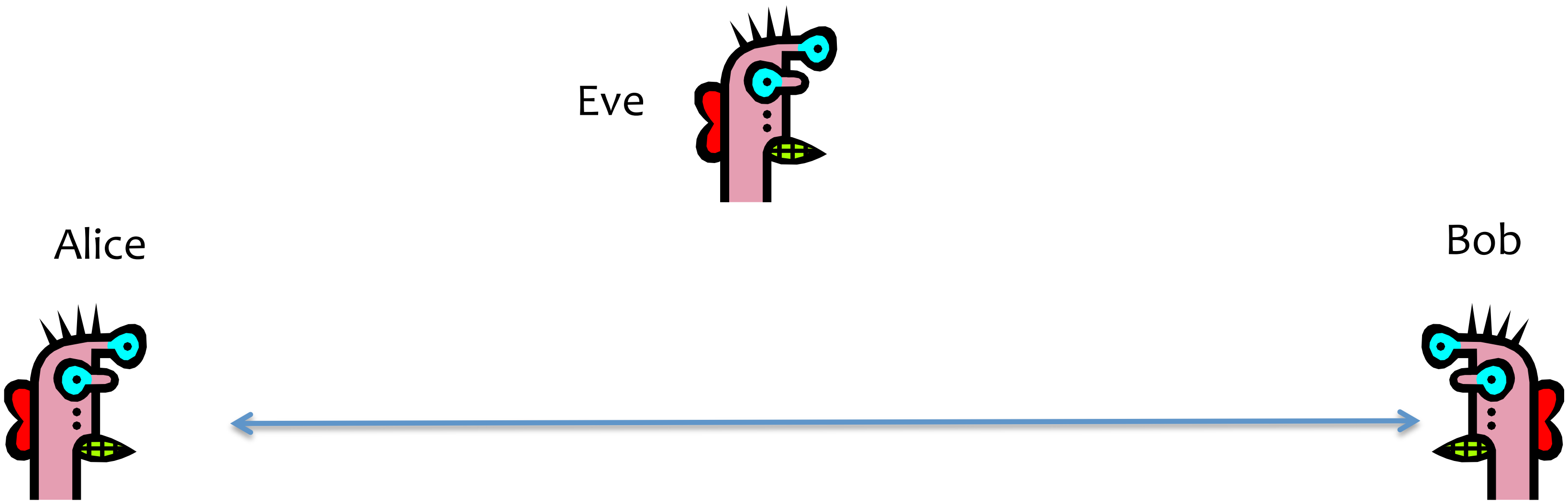
- Can't prove the other person sent the message, because you also could have computed the MAC!

- OTR takes this one step farther: After a messaging session is over, Alice and Bob send the MAC key publicly over the wire!

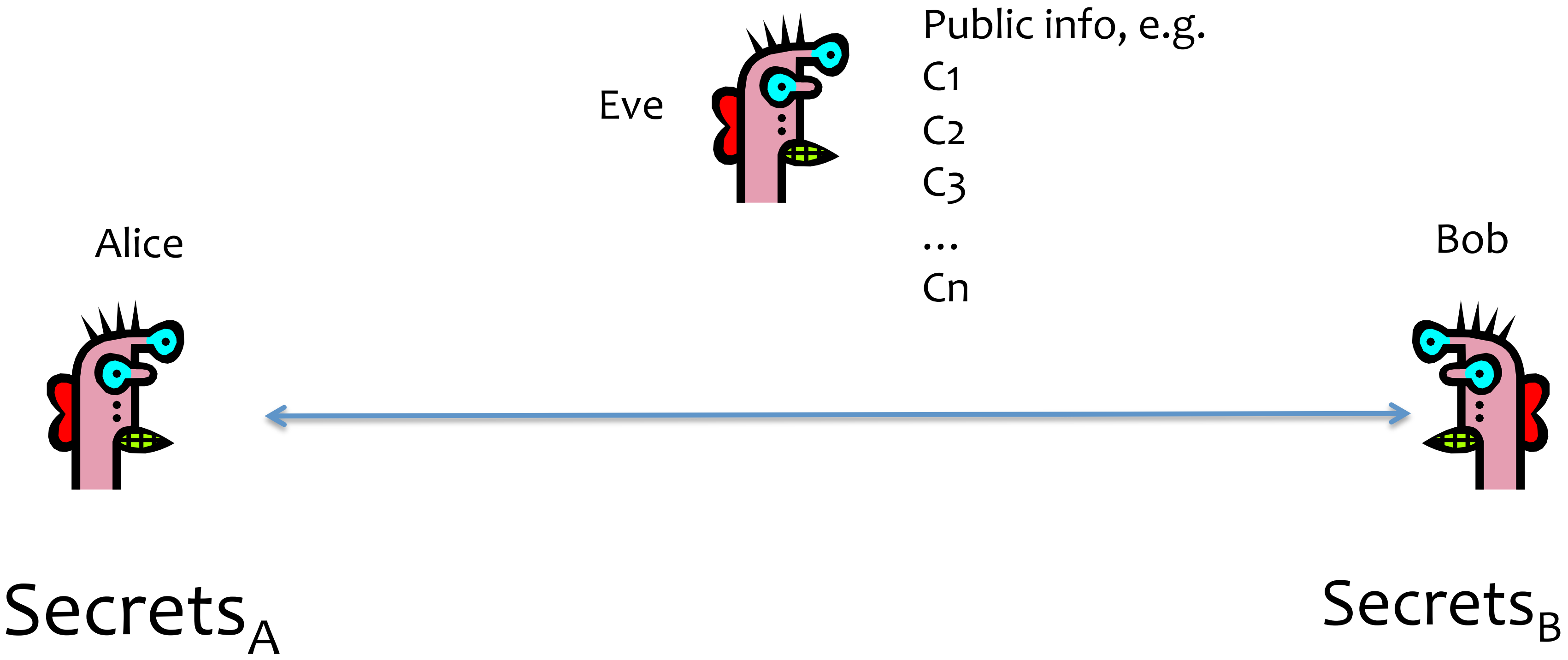
OTR: Deniability

- Eve now knows the MAC key, so technically speaking, she also has the ability to forge messages from Alice or Bob.

Perfect Forward Secrecy



Perfect Forward Secrecy



Eve

Public info, e.g.
C1
C2
C3
…
Cn

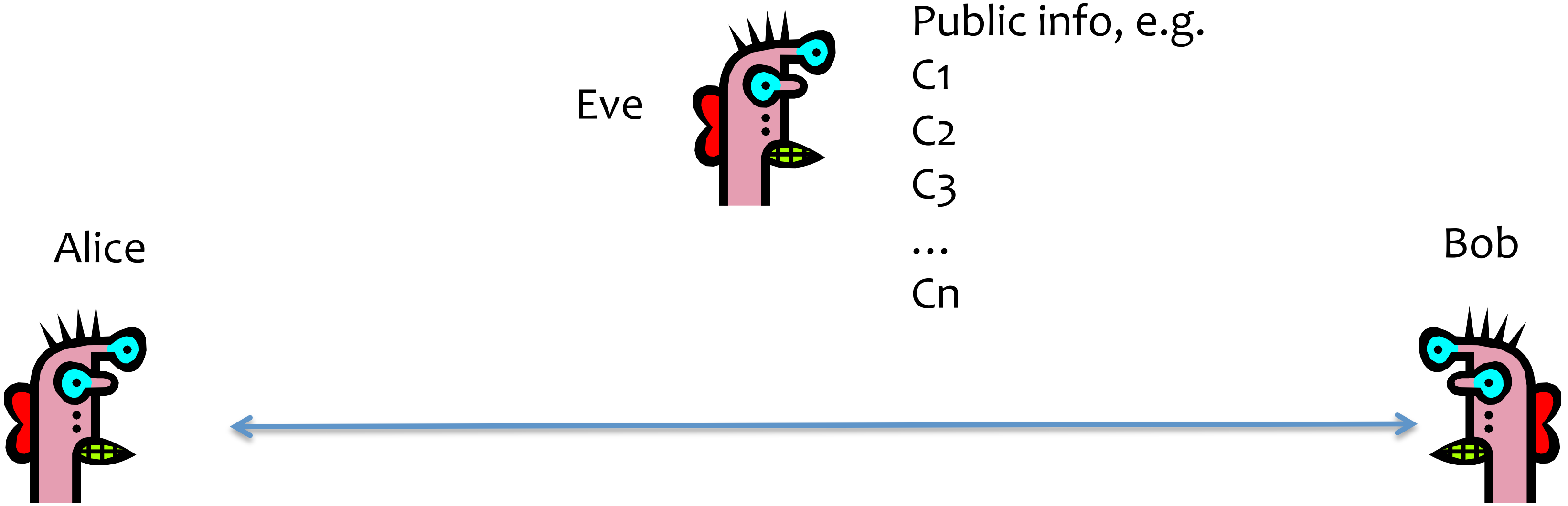
Alice

Bob

Secrets_A

Secrets_B

Perfect Forward Secrecy



Eve

Public info, e.g.
C1
C2
C3
…
Cn

Alice

Bob

Secrets_A

If Eve compromises Alice or Bob's computers at a later date, we would like to prevent her from being able to learn what M1, M2, M3, etc. correspond to C1, C2, C3, etc.

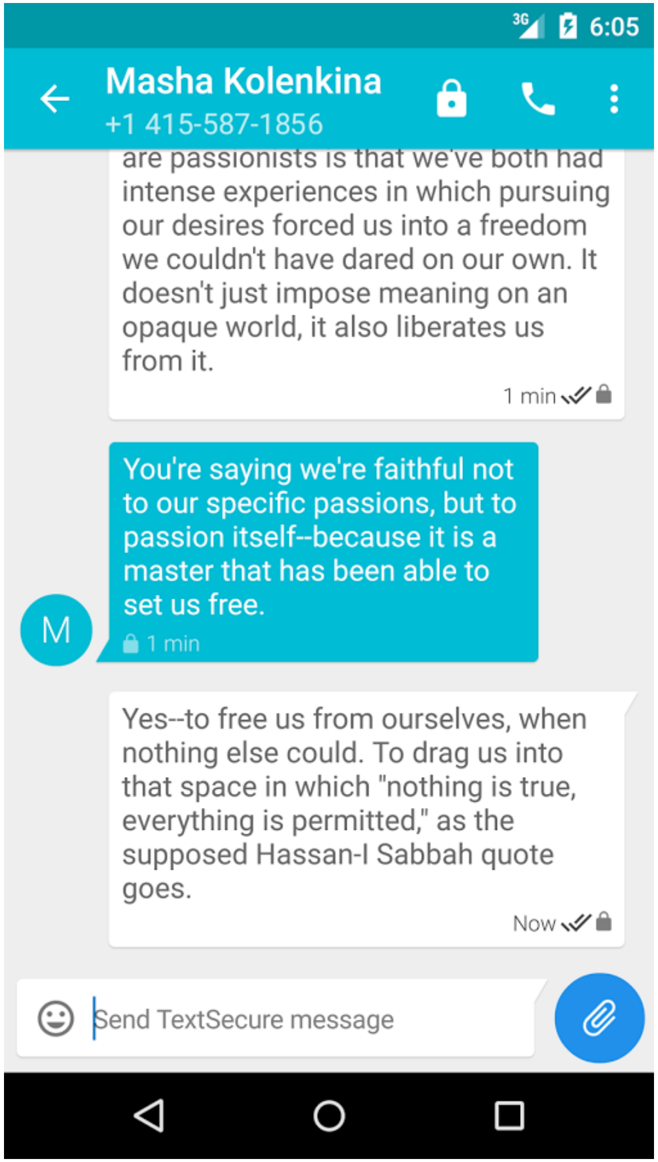
Secrets_B

OTR: Ratcheting

- Idea: Use a new key for every session/ message/time period.

Signal

- End-to-end encrypted chat/IM based on OTR
- Provides variations on ratcheting, deniability, etc.
- Widely used, public code, audited.



3G 6:05

Masha Kolenkina
+1 415-587-1856

are passionists is that we've both had intense experiences in which pursuing our desires forced us into a freedom we couldn't have dared on our own. It doesn't just impose meaning on an opaque world, it also liberates us from it.

1 min

You're saying we're faithful not to our specific passions, but to passion itself--because it is a master that has been able to set us free.

M

1 min

Yes--to free us from ourselves, when nothing else could. To drag us into that space in which "nothing is true, everything is permitted," as the supposed Hassan-I Sabbah quote goes.

Now