Announcements

This Wednesday: **No office hour from me**

- I am travelling to DC for CPS PI meeting

This Thursday:

- Guest Lecture on ***Machine Learning Security***
  - One of the hottest security research topics today
  - Some of you might find the speaker familiar ☺
- Attention: **It's within the scope of final exam**

CS 134
SPRING 2019

LECTURE 14
Privacy and Anonymity

[lecture slides are adapted from previous slides by Prof. Gene Tsudik]

Privacy

- Privacy and Society
  - Basic individual right & desire
  - Also relevant to corporations & government agencies
  - Recently increased awareness
    - But, public's perception of privacy is fickle
- Privacy and Technology in Recent Years
  - >>Information disclosed on the Internet
  - >>Handling and transfer of sensitive information
  - <<Privacy and accountability

(Image from geekologie.com)

Privacy on Public Networks

- The Internet is designed as a public network
  - Anyone on your LAN (wired or wireless) can see your traffic
  - Network routers see all traffic that passes through them
- Routing information is public
  - IP packet headers identify source and destination addresses
  - A passive observer can easily figure out who is talking to whom
- Encryption (e.g., SSL or IPSec) does not hide identities
  - Encryption hides payload, not routing+addressing information
  - Even IP-level encryption (tunnel-mode IPsec/ESP) reveals IP addresses of IPsec gateways

Applications of Anonymity (1)

- Privacy
  - Hide online transactions, web browsing, etc. from intrusive governments, marketers, archival/search entities (e.g., Google) as well as from criminals and snoops
- Untraceable Electronic Mail
  - Corporate whistle-blowers
  - Political dissidents in oppressive societies
  - Socially sensitive communications
  - Confidential business negotiations
- Law Enforcement and Intelligence
  - Sting operations
  - Secret communications on a public network
    - Informers, secret agents, etc.

Applications of Anonymity (2)

- Digital/Electronic Cash
  - Electronic currency with properties of paper money (online purchases unlinkable to buyer's identity)
- Anonymous Electronic Voting
- Censorship-Resistant Publishing
- Crypto-Anarchy
  - "Some people say that "anarchy won't work." That's not an argument against anarchy; that's an argument against work." – Bob Black

Applications of Anonymity (3)

- Porn
- Human Trafficking
- Libel
- Disinformation = Fake News / Propaganda
- Sale of Illegal Substances (e.g., SilkRoad)
- Tax Avoidance (via Untraceable Payments)
- Incitement to Criminal Activity (e.g., Murder, Rioting, Genocide, Terrorism)

What is **Anonymity**?

- **Anonymity**: inability to identify someone within a set of subjects (size varies)
  - Different from PRIVACY – right to be left alone
  - To be anonymous, need to hide your activities among similar activities by others
  - One cannot be anonymous alone!
    - Big difference between anonymity and confidentiality
- **Unlinkability**: separation of action and identity performing that action
  - For example, sender and his email are no more related after observing communication than they were before
- **Unobservability**: inability to tell whether a certain action took place
  - very hard to achieve

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 a packet flow
- Compromise of Network Nodes (such as Routers)
  - Not obvious which nodes have been compromised
    - Attacker may be passively logging traffic
  - Do not fully trust any individual node
    - Assume that some fraction of nodes is good, but do not know which

Chaum's Mix
(David Chaum, ca. 1980-1981)

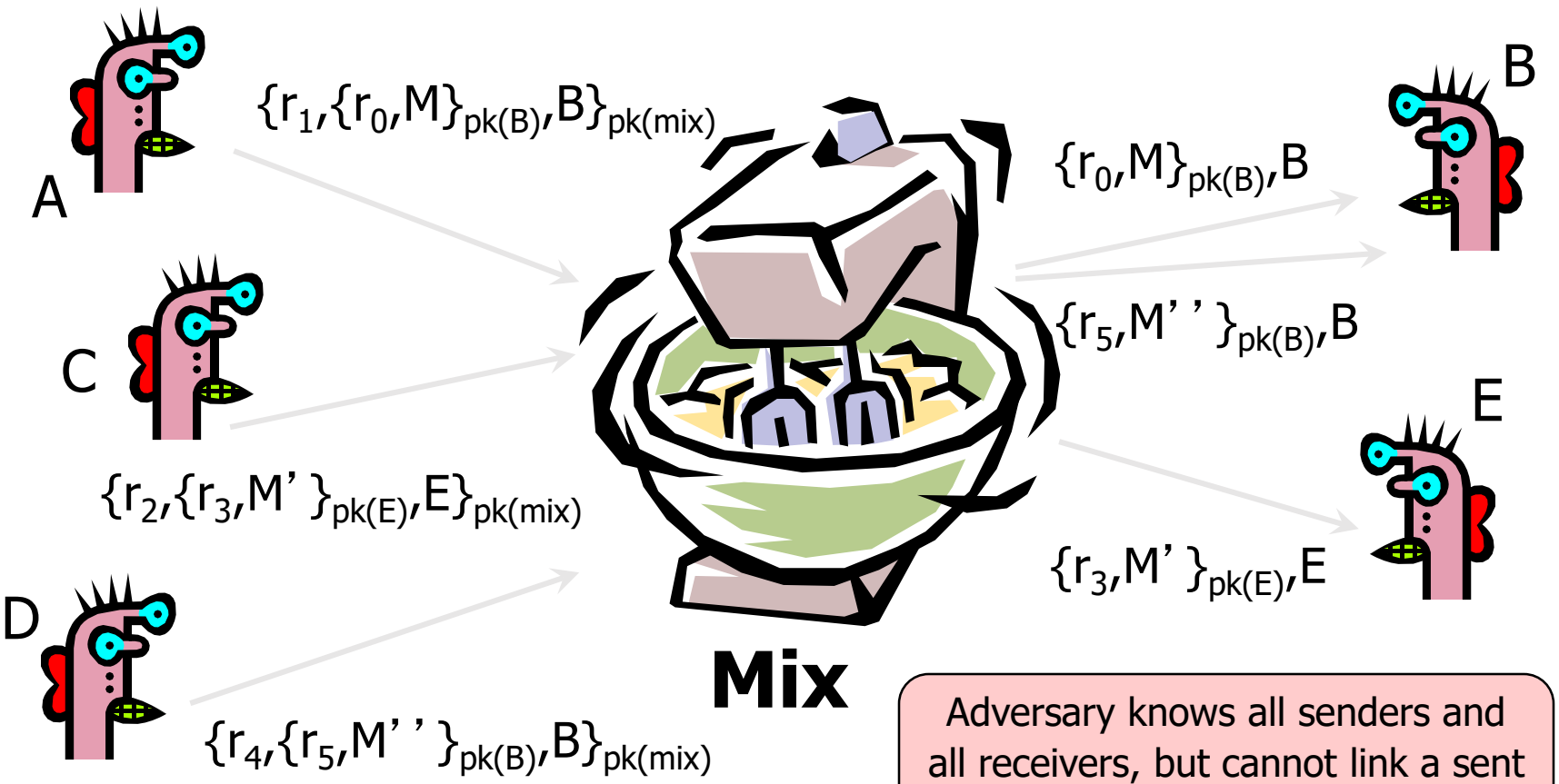
- Earliest proposal for anonymous email:
  - David Chaum, "Untraceable electronic mail, return addresses, and digital pseudonyms", Communications of the ACM, February 1981.

- Public-key crypto + trusted re-mailer (Mix)
  - Untrusted communication medium
  - Public-keys used as persistent pseudonyms

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

- Modern anonymity systems use Mix as the basic building block

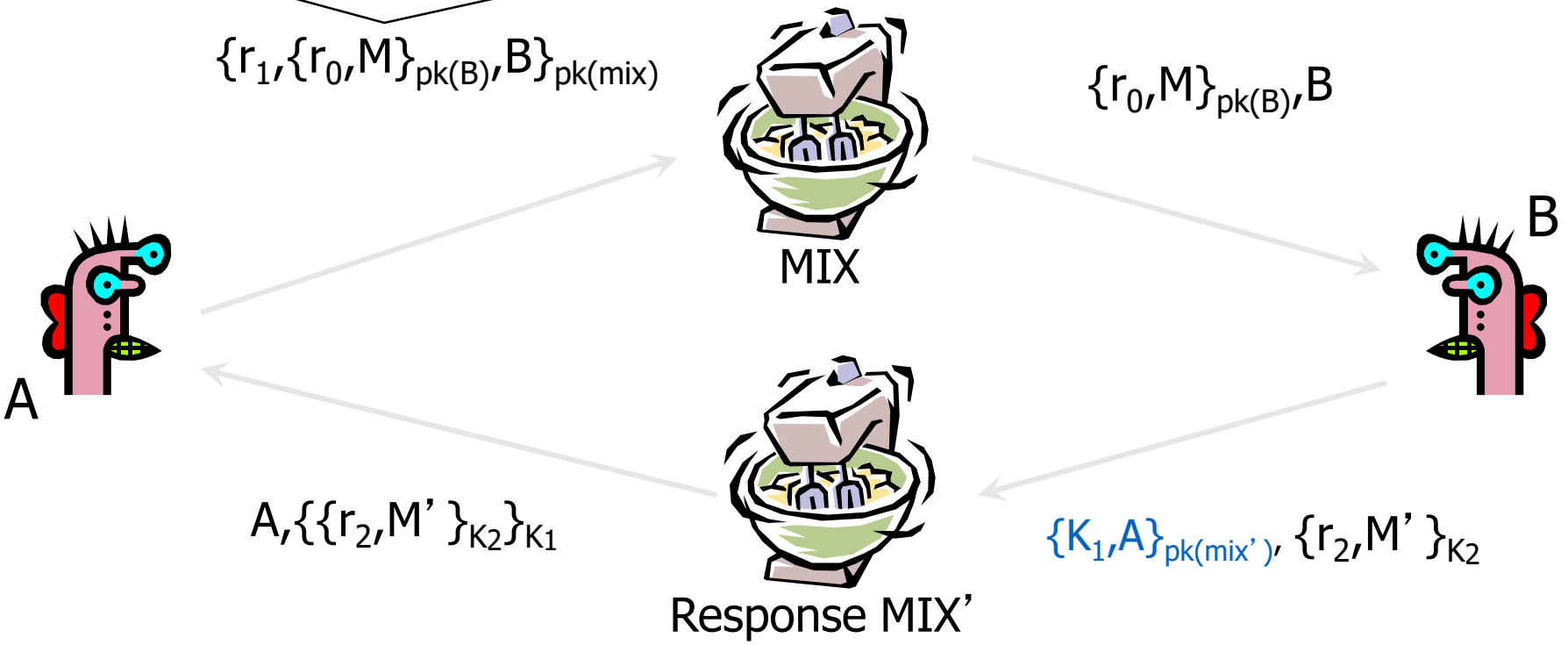
Basic Mix Design



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

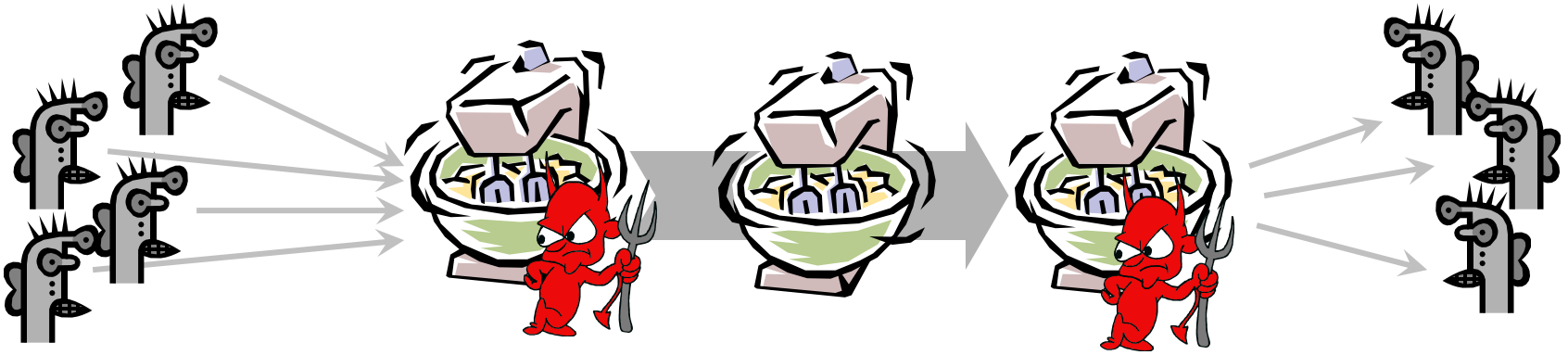
Anonymous Return Addresses

M includes $\{K_1,A\}_{pk(mix')}$, K_2 , where K_2 is a fresh public key and MIX' is possibly different from MIX



Secrecy without authentication
(good for an online confession service)

Mix Cascade

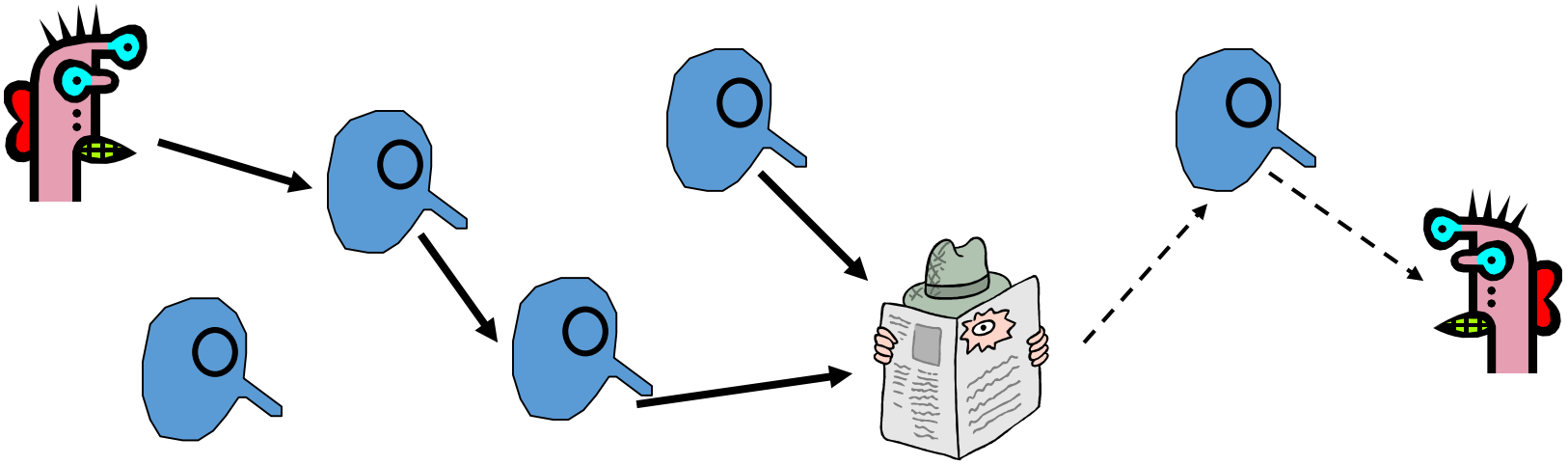


- Messages are sent through a sequence of mixes
  - Can also form an arbitrary network of mixes ("mixnet")
- Some mixes may be controlled by attacker, but even a single good mix guarantees some anonymity
- Pad and buffer traffic to foil correlation attacks

Disadvantages of Basic Mixnets

- Public-key encryption and decryption at each mix are computationally expensive
- Basic mixnets have high latency
  - Ok for email, but not for anonymous Web browsing
- Challenge: low-latency anonymity network
  - Use public-key cryptography to establish a "circuit" with pairwise symmetric keys between hops on the circuit
  - Then use symmetric decryption and re-encryption to move data messages along the established circuits
  - Each node behaves like a mix; anonymity is preserved even if some nodes are compromised

Another Idea: Randomized Routing

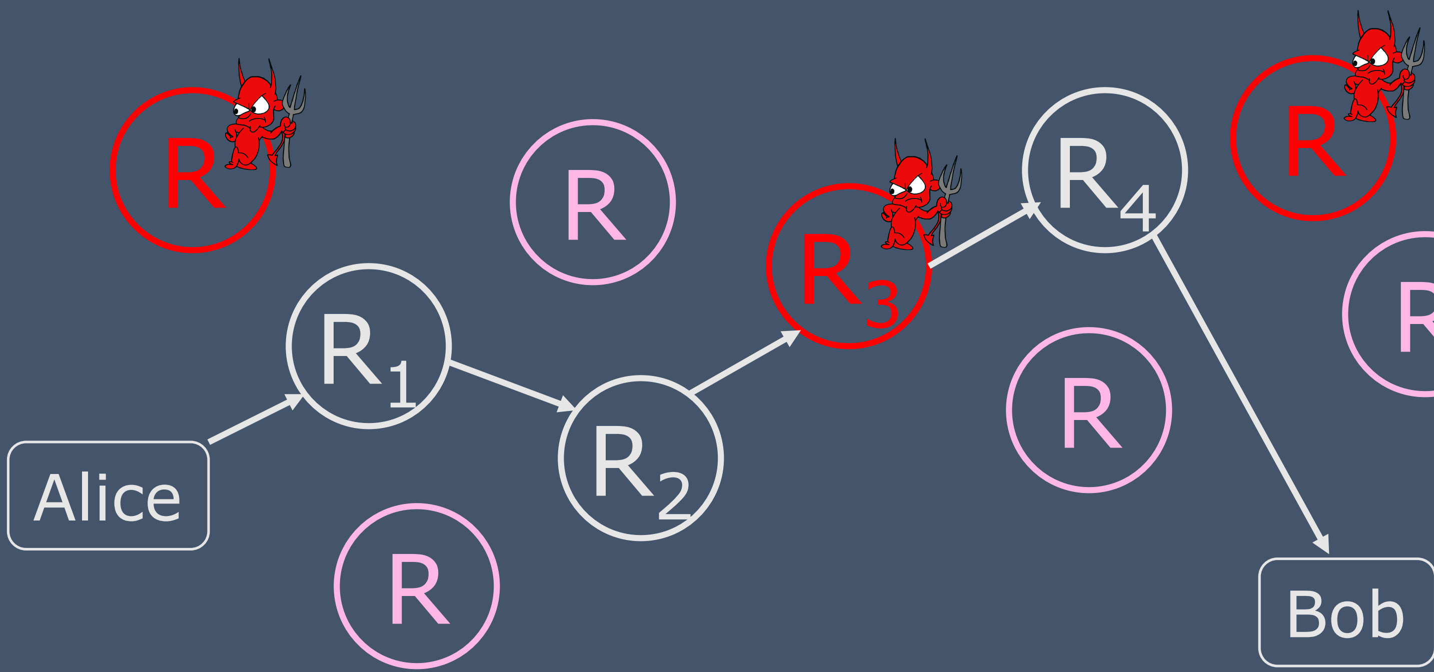


- Hide sources by routing messages randomly
  - Popular technique: Crowds, Freenet, Onion routing

- Routers do not know if the apparent source of a message is the true sender or another router

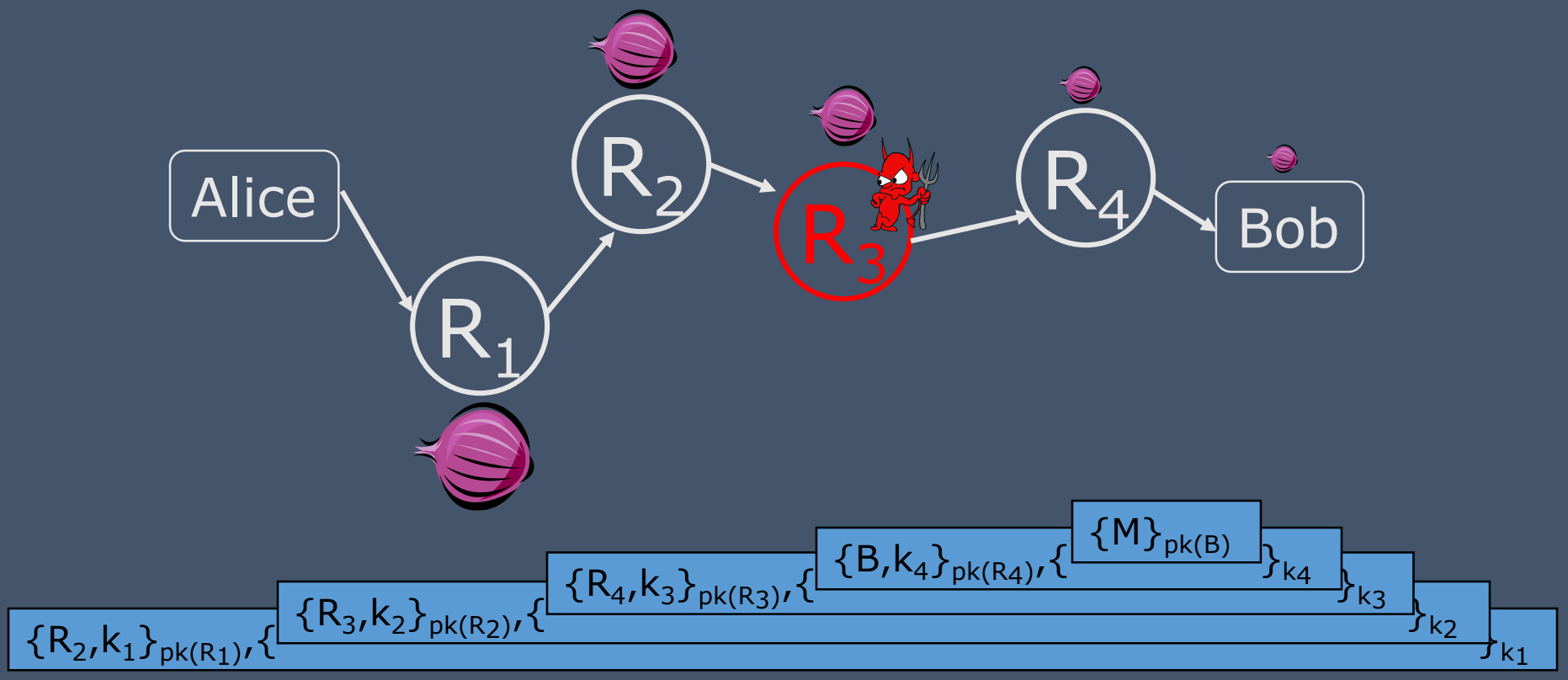
[Reed, Syverson, Goldschlag 1997]

Onion Routing (early Tor)



- Sender chooses a random sequence of routers
  - Some routers are honest, some are not
  - Sender controls path length

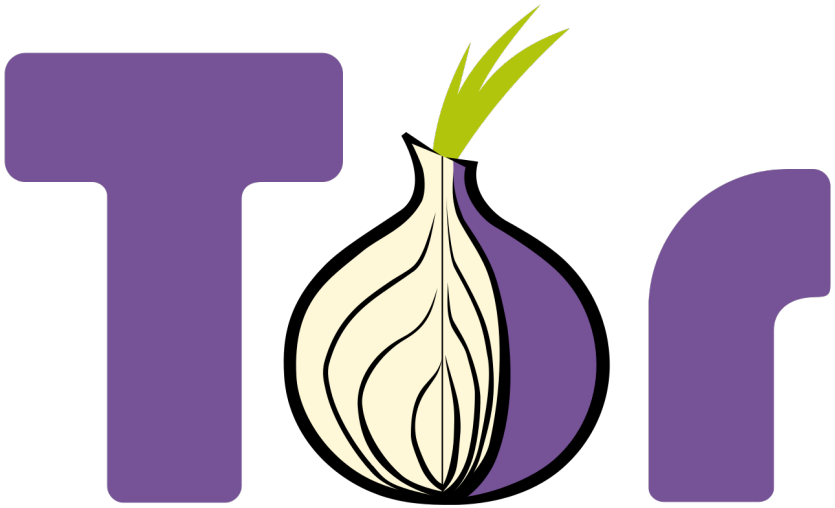
Route Establishment (Early Tor)



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

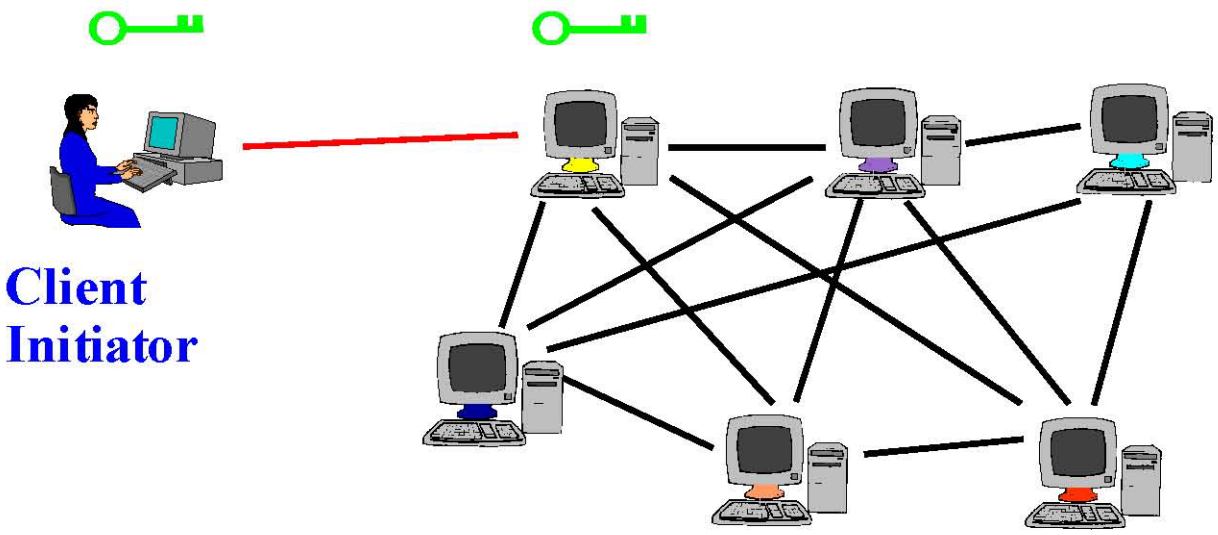
The Onion Router (current Tor)

- Second-generation onion routing network
  - http://tor.eff.org
  - Specifically designed for low-latency anonymous Internet communications (e.g., Web browsing)
  - Running since October 2003
- Hundreds of nodes on all continents
- 2.5 million users as of 05/2019
- "Easy-to-use" client proxy
  - Freely available, can use it for anonymous browsing
  - Available for smartphones and tablets too
  - TorBrowser – get it!!!



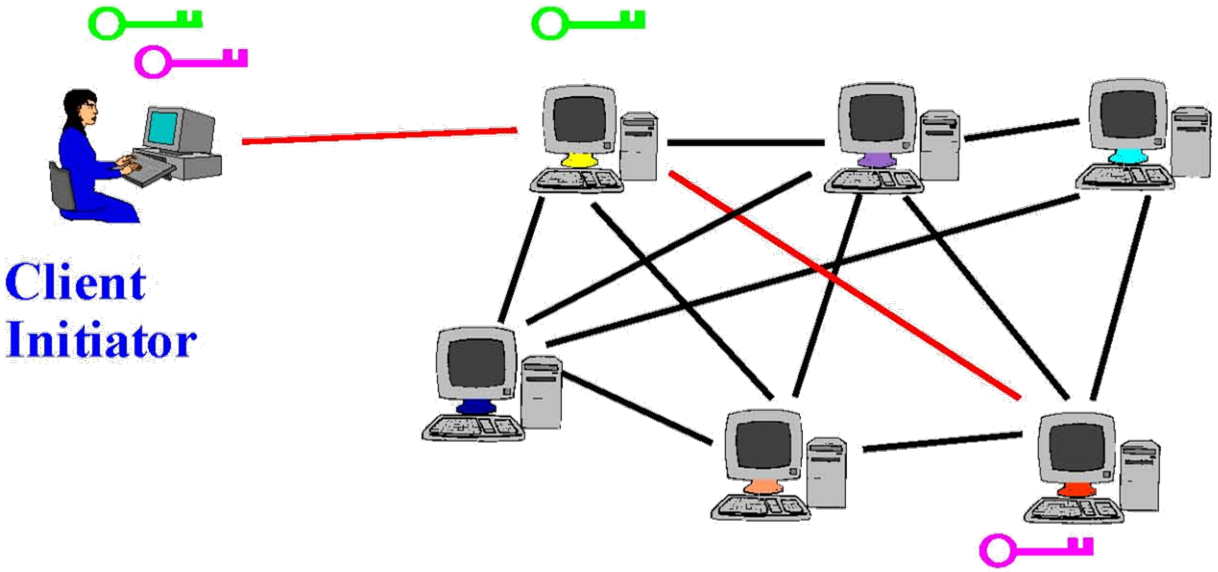
Tor Circuit Setup (1)

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



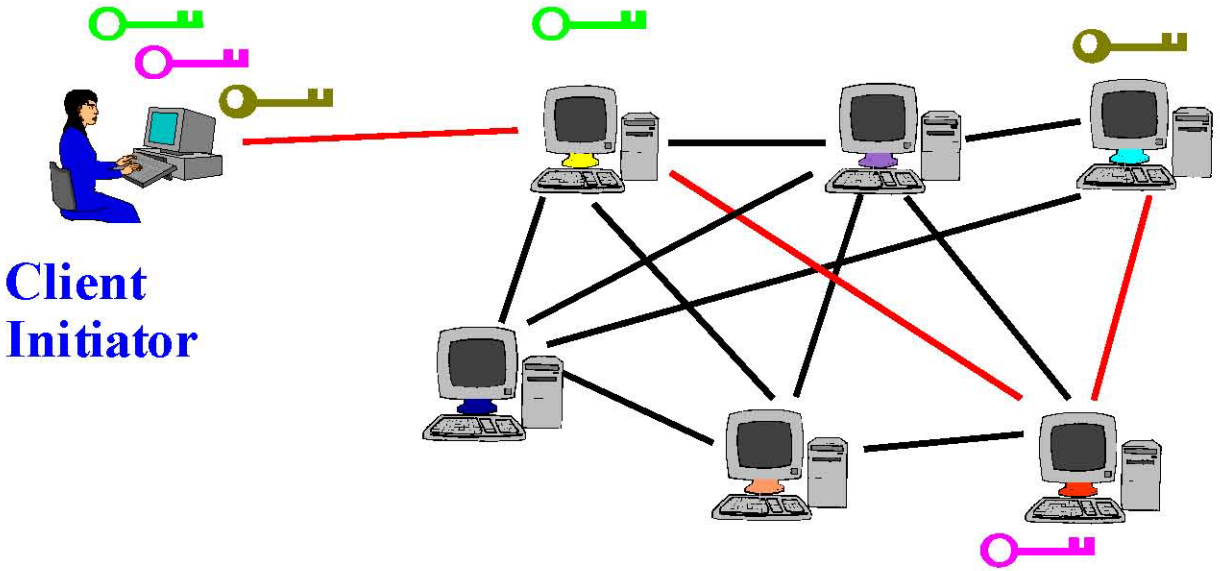
Tor Circuit Setup (2)

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



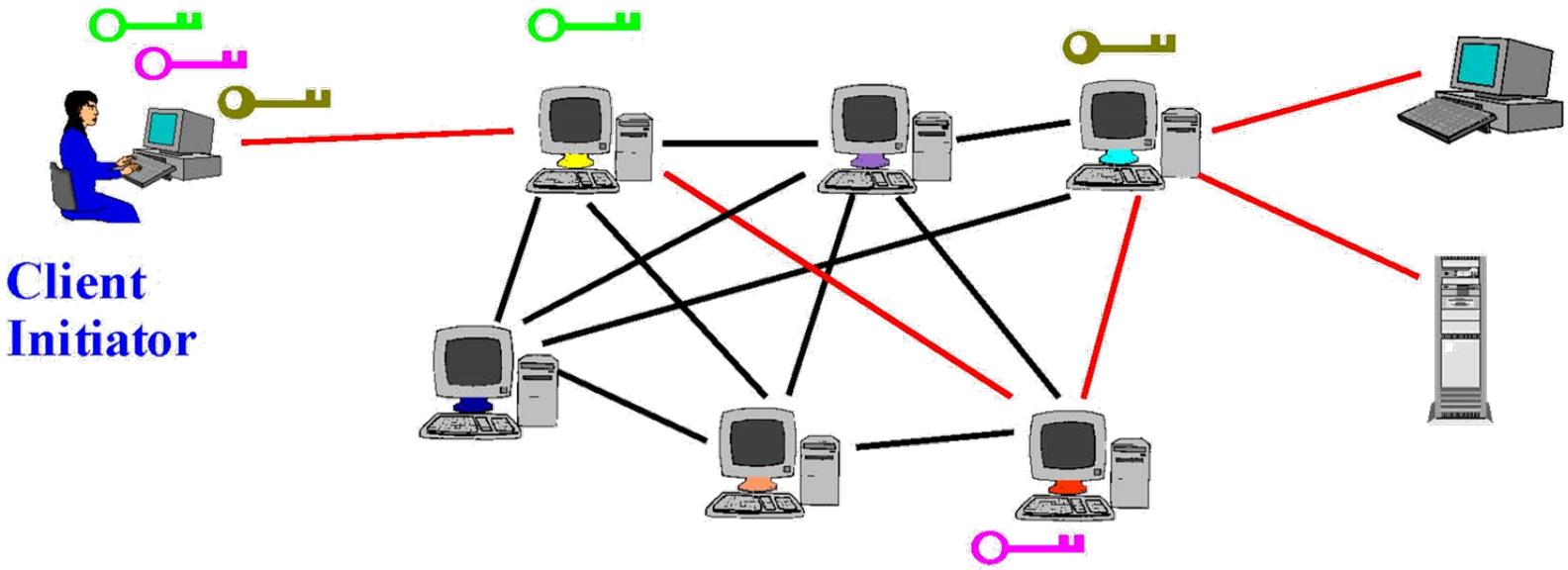
Tor Circuit Setup (3)

- 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 (also to multiple dst-s)
  - Datagrams are decrypted and re-encrypted at each link



Tor Management Issues

- Many applications can share one circuit
  - Multiple TCP streams over one anonymous connection
- Tor router do not 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

Deployed Anonymity Systems

- Free Haven project has an excellent bibliography on anonymity
  - http://www.freehaven.net/anonbib
- 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

Dining Cryptographers

- How to make a message public, but in a perfectly untraceable manner
  - David Chaum. "The dining cryptographers problem: unconditional sender and recipient untraceability." Journal of Cryptology, 1988.
- Guarantees information-theoretic anonymity for message senders
  - VERY strong form of anonymity: defeats adversary who has unlimited computational power
- Difficult to make practical
  - In group of size N, need N random bits to send 1 bit

Three-Person DC Protocol

- Three cryptographers are having dinner.
- Either NSA is paying for the dinner, or one of them is paying, but wishes to remain anonymous.
- The cryptographers respect each other's right to make an anonymous payment, but want to find out whether the NSA paid
- So they decide to execute a two-stage protocol:

1. Each diner flips a coin and shows it to his left neighbor.
   - Every diner sees two coins: his own and his right neighbor's
2. Each diner announces whether the two coins are the same. If he is the payer, he lies (says the opposite).
3. IF Number of "same"=1 or 3 $\Rightarrow$ NSA is paying

   IF Number of "same"=0 or 2 $\Rightarrow$ one of them is paying
   - But a non-payer cannot tell which of the other two is paying!

Let's see why it is correct

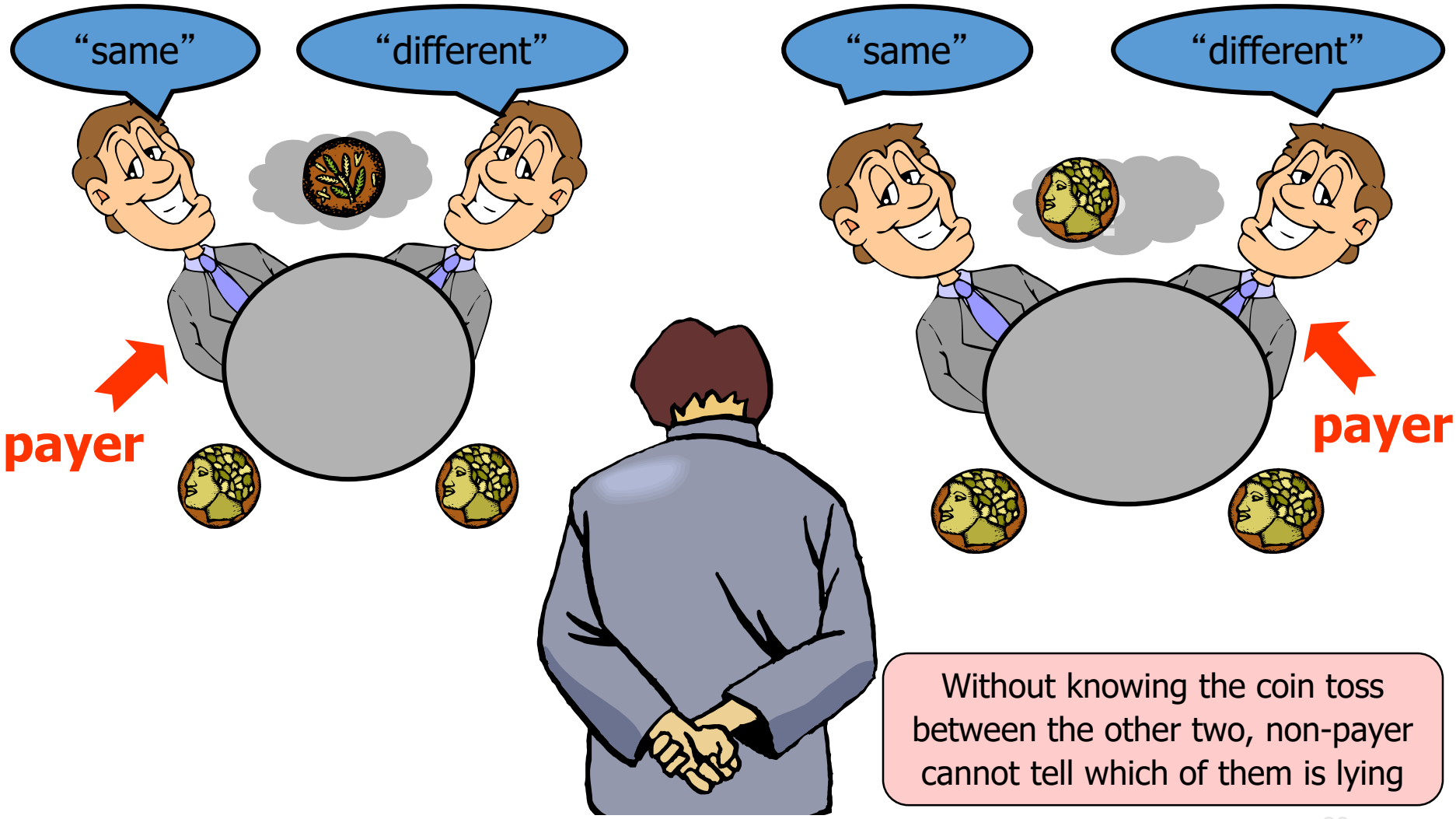
- 8 Possibilities

Coin 1	Coin 2	Coin 3
0	0	0
0	0	1
0	1	0
0	1	1
1	0	0
1	0	1
1	1	0
1	1	1

Coin 1, 2	Coin 2, 3	Coin 3, 1
Same	Same	Same
Same	Different	Different
Different	Different	Same
Different	Same	Different
Different	Same	Different
Different	Different	Same
Same	Different	Different
Same	Same	Same

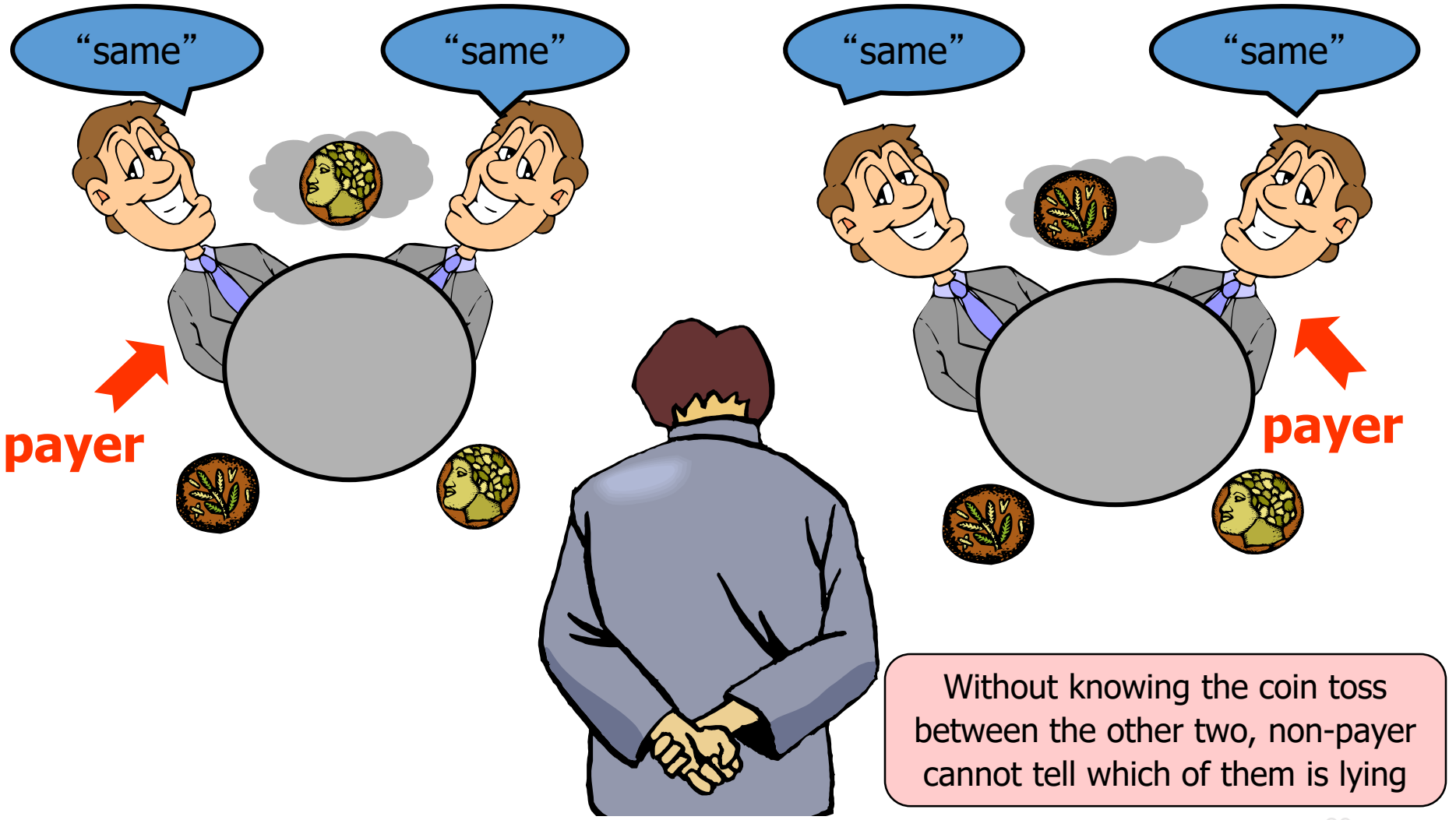
If nobody is lying, # of "same" **can only be 1 or 3** -> NSA paid!
Otherwise, one of the cryptographers must have paid, but **impossible to tell who**!

Non-Payer's View: Same Coins



Without knowing the coin toss between the other two, non-payer cannot tell which of them is lying

Non-Payer's View: Different Coins



Without knowing the coin toss between the other two, non-payer cannot tell which of them is lying

Super-posed Sending

- This idea generalizes to any group of size N
- For each bit of the message, every user generates 1 random bit and sends it to ONE neighbor
  - Every user learns 2 bits (his own and his neighbor's)
- Each user announces own bit XOR neighbor's bit
- Sender announces own bit XOR neighbor's bit XOR message bit
- XOR all announcements = message bit
  - Every randomly generated bit occurs in this sum twice (and is canceled by XOR), message bit occurs once