CSE 410 Spring 2025
Privacy-Enhancing Technologies

Marina Blanton

Department of Computer Science and Engineering
University at Buffalo

Lecture 14: Anonymous Communication

Anonymity on the Internet

We previously discussed (a lack of) privacy when surfing the web

Now we look beyond the web

How do we retain privacy and anonymity on the internet in general?

Anonymity on the Internet

- If we don't specify our name or other personal information, our communication can seem anonymous
- Normally, however, this is not the case:
  - if we connect to a chat channel, the server knows what address we are coming from
  - if we connect to any service, the server knows what address we are coming from
  - if you send an encrypted email, the endpoints still can be recovered
- But does it matter?

Internet Surveillance

- Internet surveillance techniques are known as traffic analysis
  - it can be used to infer who is talking to whom over a public network
- Knowing the source and destination of our traffic allows others to track your behavior and interests

Internet Surveillance

- Internet surveillance techniques are known as traffic analysis
  - it can be used to infer who is talking to whom over a public network
- Knowing the source and destination of our traffic allows others to track your behavior and interests

> "We kill people based on metadata."
>
> — General Michael Hayden, former director of the NSA and CIA

Internet Surveillance

Traffic analysis can lead to various consequences:

- an e-commerce website can use price discrimination based on your country or institution of origin
- this can even threaten your job and physical safety by revealing who and where you are
  - e.g., you are traveling abroad and connect to your employer's computers to check mail
- when abroad, you can inadvertently reveal your national origin and professional affiliation to anyone observing the network
- this holds even if the connection is encrypted

Traffic Analysis

How does traffic analysis work?

- Internet data packets have two parts: data payload and header used for routing
- the payload is what is being sent
  - e.g., an email message, a web page, an audio file
- even if the payload is encrypted, traffic analysis still reveals a lot about what you are going (and possibly what you are saying)

Traffic analysis focuses on the header that discloses source, destination, size, timing, etc.

Traffic Analysis

- The basic problem is that the recipient of your communications can see that you sent it
  - so can authorized intermediaries, i.e., Internet service providers, and sometimes unauthorized intermediaries
- A simple form of traffic analysis might involve someone sitting between the sender and recipient on the network looking at headers
- More powerful types include:
  - spying on multiple parts of the Internet
  - using sophisticated statistical techniques to track the communication patterns

Benefits of Anonymous Communication

Suppose we can build anonymous communication channels, what does it enable us to do?

- the basic line is that it allows organizations and individuals to share information over public networks without compromising privacy
- individuals can keep websites from tracking them
- individuals can connect to news sites, instant messaging services, and the like when these are blocked by their local Internet providers
- individuals can publish websites and other services without needing to reveal the location of the site
- individuals can conduct socially sensitive communication
  - e.g., chat rooms and web forums for rape and abuse survivors or people with illnesses

Benefits of Anonymous Communication

What else do anonymous channels enable us to do?

- journalists can communicate more safely with whistleblowers and dissidents
- organizations can enable their workers to connect to their home websites while in foreign countries without letting others know for whom they are working
- activist groups recommend anonymous communication as a mechanism for maintaining civil liberties online
- corporations can perform competitive analysis and protect sensitive procurement patterns from eavesdroppers
- law enforcement can visit and surveil websites without leaving government IP addresses in their logs

Anonymous Communication

Anonymity likes company

- you cannot be anonymous by yourself
  - but you can have confidentiality by yourself
- a network that protects only Department of Defense (DoD) network users won't hide that connections from that network are from DoD
- you can be anonymous by hiding in the crowd

There are several technical approaches to achieve anonymity

The most popular are mixes and proxies

Recall Encryption Types

Recall the differences between symmetric and public-key encryption

symmetric encryption	public-key encryption
generate a secret key k	generate a pair of public-private keys (pk, sk) and announce pk
encryption and decryption use the same key k	everyone can encrypt using pk, the key owner can decrypt using sk
main advantage is speed	main advantage is the ability to realize more functionalities
e.g., Advanced Encryption Standard (AES) (128-bit key)	e.g., RSA (3072-bit key)

Recall Encryption Types

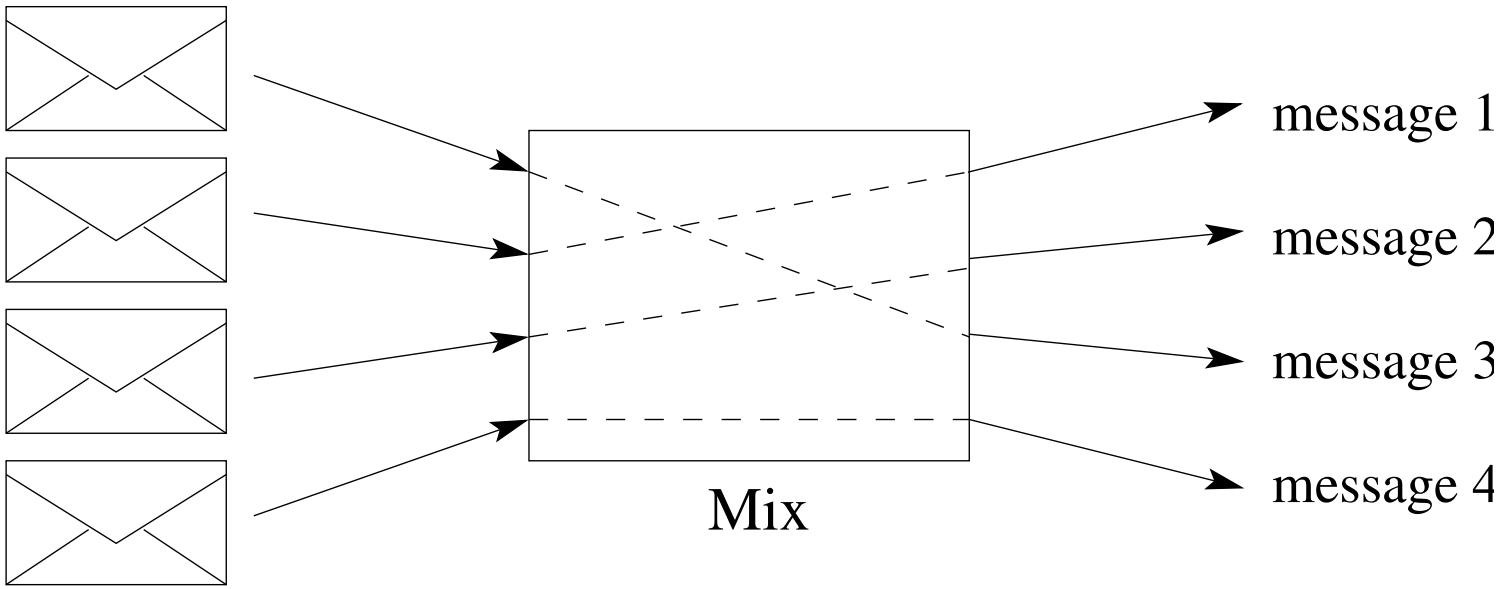
Recall the differences between symmetric and public-key encryption

symmetric encryption	public-key encryption
generate a secret key k	generate a pair of public-private keys (pk, sk) and announce pk
encryption and decryption use the same key k	everyone can encrypt using pk, the key owner can decrypt using sk
main advantage is speed	main advantage is the ability to realize more functionalities
e.g., Advanced Encryption Standard (AES) (128-bit key)	e.g., RSA (3072-bit key)

see https://keylength.com for key sizes

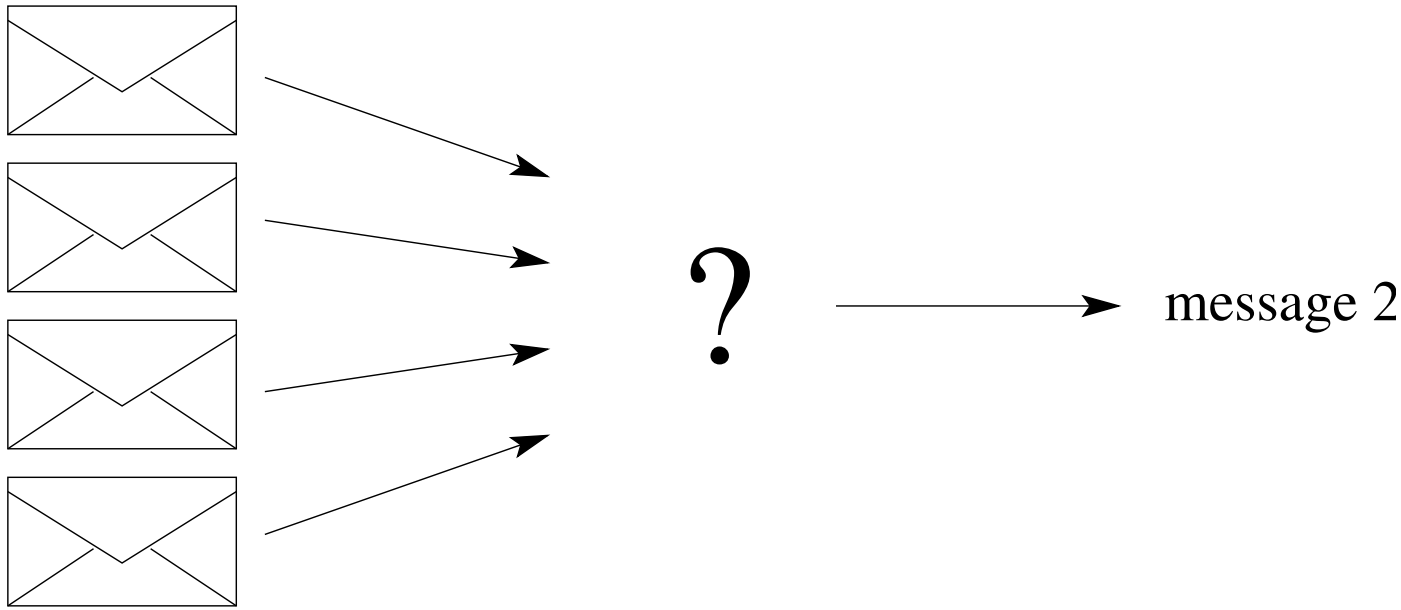
Mixes

- What does a mix do?
  - it receives encrypted messages
  - it then randomly permutes and decrypts inputs



Mixes

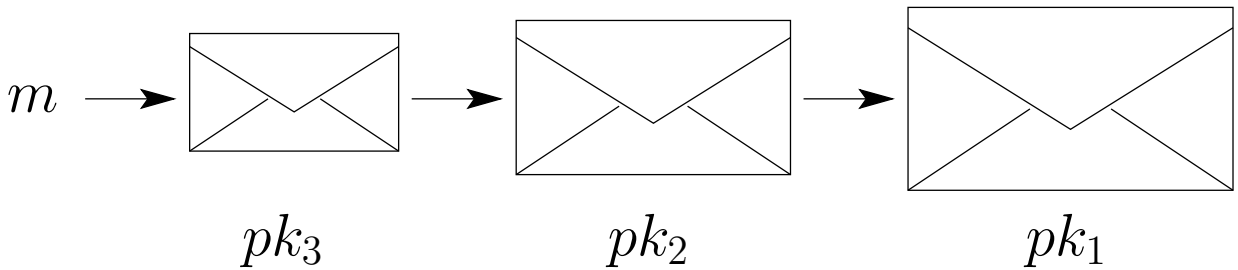
- The key property is that an adversary cannot tell which ciphertext corresponds to a given message



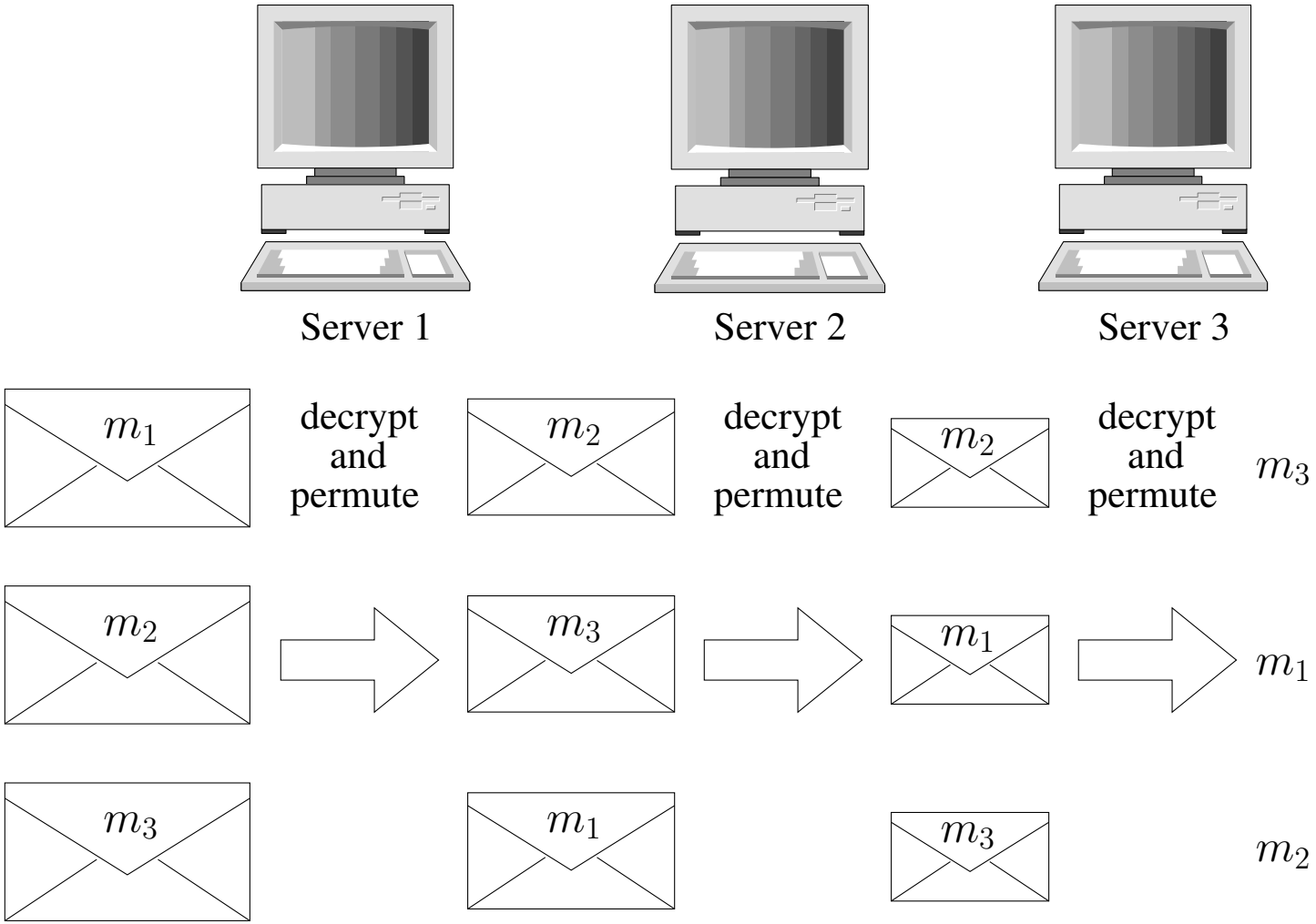
Mixes

- The basic mix was introduced by Chaum in 1981
  - there is a number of servers each with its own public key pk_i
  - to send a message m through servers 1, 2, and 3, envelope it using all of the servers' keys

$$c = \mathsf{Enc}_{pk_1}(\mathsf{Enc}_{pk_2}(\mathsf{Enc}_{pk_3}(m)))$$



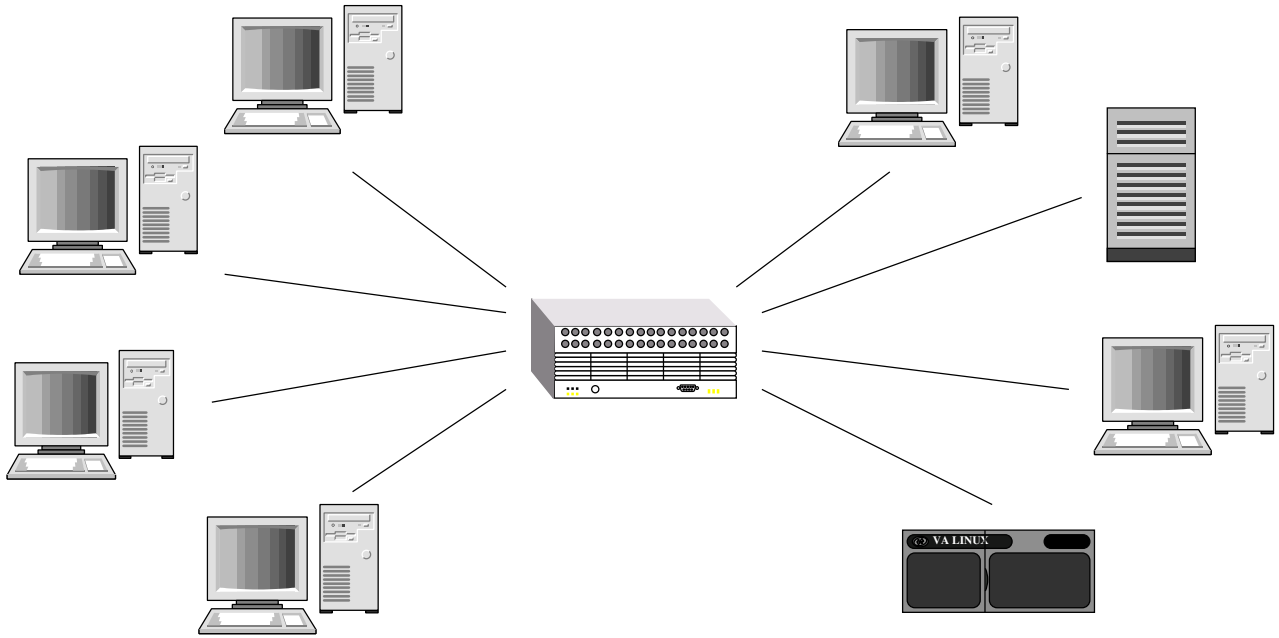
Mixes


Server 1
Server 2
Server 3
m_1
decrypt and permute
m_2
decrypt and permute
m_2
decrypt and permute
m_3
m_2
m_3
m_1
m_1
m_3
m_1
m_3
m_2

Mixes

- Each server on the way knows only which server gave it data and which server it is giving data to
- No individual server ever knows the complete path that a data packet has taken
- One honest server preserves privacy
- Mixnets were introduced for email and other high latency applications
  - each layer of message requires expensive public-key cryptography
- But what if you need quick interaction?
  - web browsing, remote login, chat, etc.

Anonymizing Proxy

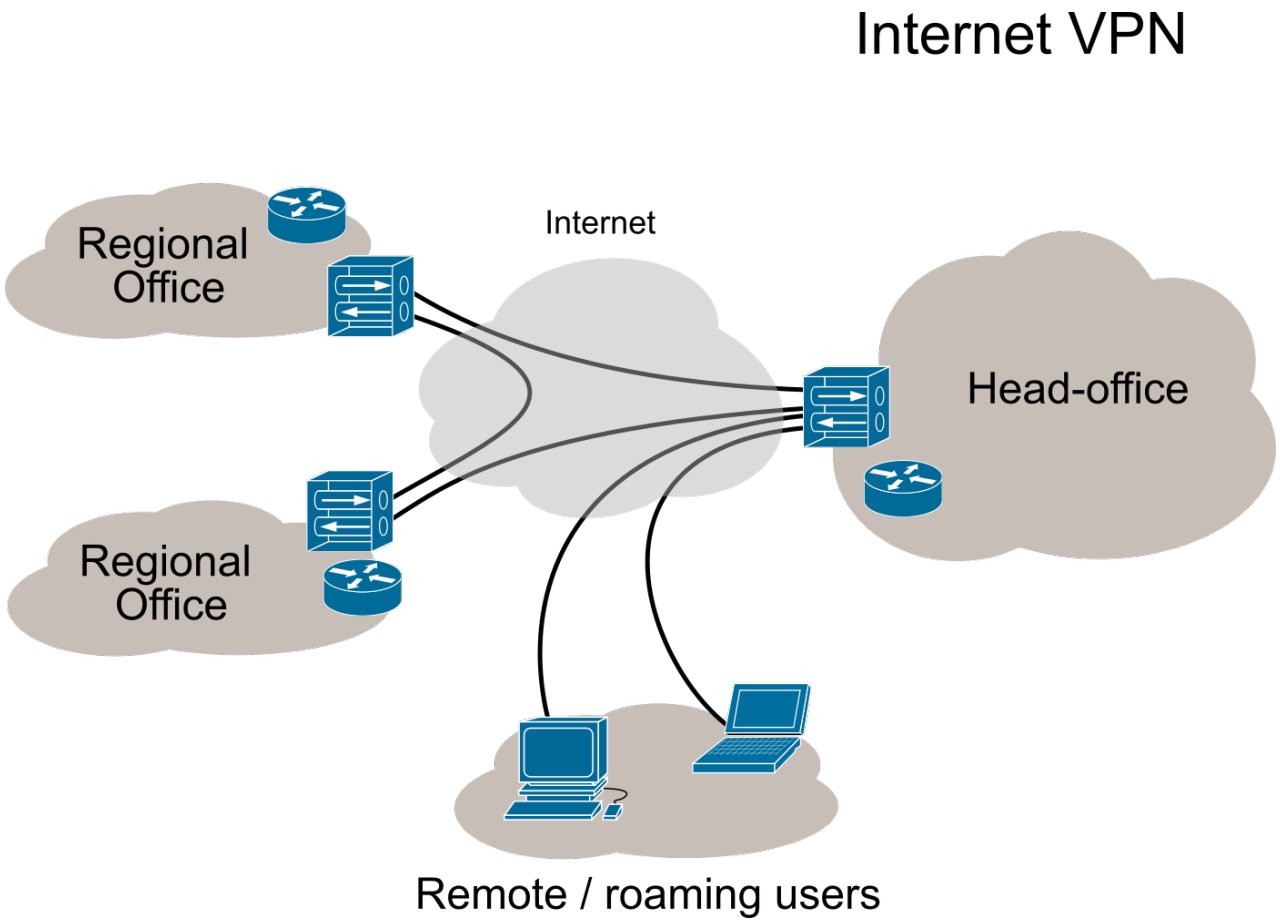


- communications appear to come from the proxy, not true senders
- it can use low-cost symmetric encryption (or no encryption)
- it thus is appropriate for web connections, SSL/TLS, ssh, etc.

Anonymizing Proxy

- Advantages: simple, focuses a lot of traffic for more anonymity
- Disadvantages: a single point of failure, compromise, attack
- Risks of using anonymizing HTTP proxies
  - all data you send to the service must first go through the proxy
  - a malicious proxy server can record everything you send to it, including unencrypted logins and passwords
  - thus, don't use proxy servers of unknown integrity
  - if there is no choice, do not pass any sensitive information through the proxy unencrypted

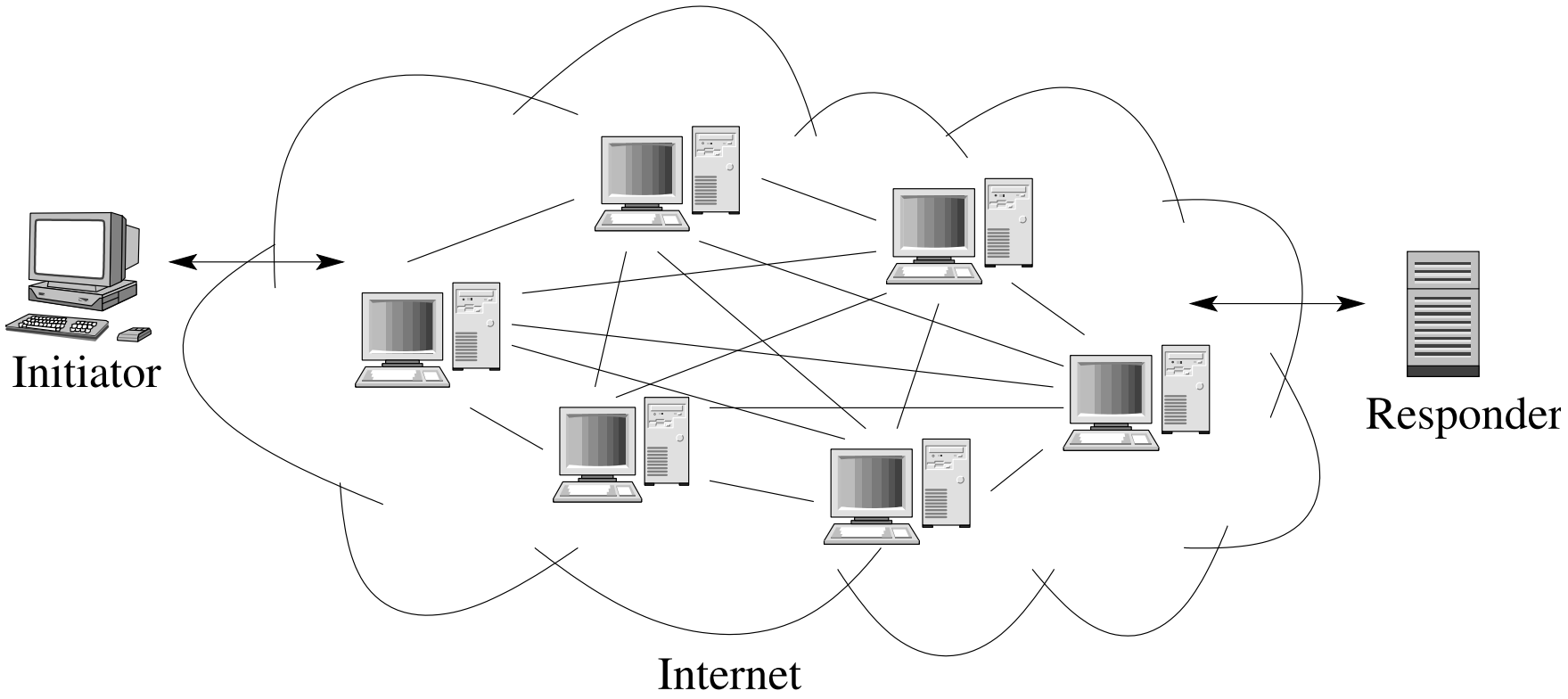
Using VPN as a Proxy



Onion Routing

- Onion Routing can be used to build traffic analysis resistant infrastructure
- The main idea is to combine advantages of mixes and proxies
  - use (expensive) public-key crypto to establish circuits
  - use (cheaper) symmetric-key crypto to move data
- Trust is distributed like in mixes
- Actual communication is much faster than in mixes

The Onion Routing (TOR) Network



TOR

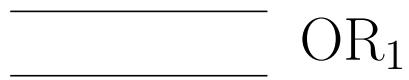
TOR establishes routing connections called circuits

- during circuit setup session keys for symmetric cryptography are negotiated using servers' public keys
- after some time session keys used in a circuit are refreshed to limit the impact of key compromise
- session keys are independent of long-term public keys
- this means that compromise of a router and its long-term keys does not lead to compromised communication

TOR Circuits

During TOR circuit setup:

- the client chooses a set of onion routers to tunnel packets through
- the client's software establishes a session key and a circuit with the first onion router on the list



- it then tunnels through that circuit to extend to the second router on the list, etc.
  - this means that the second router can know only the previous router and nothing beyond that

TOR

Client applications connect and communicate over the Tor circuit

- many applications can share it to communicate with various destinations

Directory servers maintain a list of onion routers, their status, location, current keys, etc.

- they also control which nodes can join the networks (helps prevent certain attacks and abuse)

TOR Details

Tor setup in more detail

- each user runs local software called an onion proxy to fetch directories, establish circuits, and handle connections from user applications
- each onion router maintains a long-term identity key and a short-term onion key
  - the identity key is used to sign TLS certificates, router descriptor information (address, bandwidth, etc.), and directories
  - the onion key is used to decrypt requests from users to setup a circuit and negotiate session keys
- the TLS protocol establishes a short-term link key when communicating between onion routers
  - these keys are rotated periodically and independently

TOR Circuits

Tor circuit setup

- the client's onion proxy (OP) chooses routers OR_1, OR_2, ...
- OP engages in a Diffie-Hellman key establishment with OR_1:
  - OP sends g^{a_1} encrypted under OR_1's key
  - OR_1 responds with g^{b_1} and a hash of $k_1 = g^{a_1 b_1}$

$$\text{OP} \xrightarrow{\quad \mathsf{Enc}_{pk_1}(g^{a_1}) \quad} \xleftarrow{\quad g^{b_1},\ h(k_1 || \text{“handshake”}) \quad} \text{OR}_1$$

  - the hash tells OP that OR_1 indeed computed $g^{a_1 b_1}$

TOR Circuits

Tor circuit setup

- OP then uses OR_1 to extend the circuit to OR_2:
  - OP tunnels through OR_1 key exchange negotiation for OR_2
  - OR_1 relays the request to OR_2 and forwards OR_2's reply to OP

$$\text{OP} \xrightarrow{\mathsf{Enc}_{pk_2}(g^{a_2})} \text{OR}_1 \xrightarrow{\mathsf{Enc}_{pk_2}(g^{a_2})} \text{OR}_2$$
$$\text{OP} \xleftarrow{g^{b_2},\ h(k_2||\text{“handshake”})} \text{OR}_1 \xleftarrow{g^{b_2},\ h(k_2||\text{“handshake”})} \text{OR}_2$$

  - here $k_2 = g^{a_2 b_2}$ is a session key shared between OP and OR_2
- the process continues until session keys with all of the routers on the path are established

TOR Circuits

- Client applications connect and communicate over the Tor circuit
  - many applications can share it to communicate with various destinations
- Established circuits use layered encryption as in mixes, but now decryption is much faster
- As before, each router randomly permutes the packets
- Session keys are re-negotiated after a short period of time (e.g., one minute)

TOR

TOR properties

- replay attacks are not effective
  - replayed circuit setup will result in a new session key at an honest onion router
- recording all traffic sent to a node and later breaking its public key will not reveal encrypted content
  - this is called perfect forward secrecy
- it can adapt to network dynamics
  - if one router becomes unusable, building a whole new circuit is not required

TOR Hidden Services

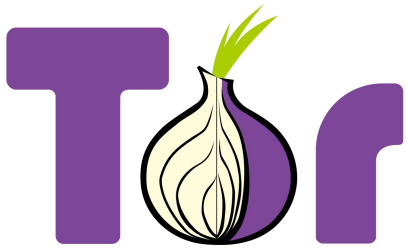
- TOR makes it possible for users to hide their locations while offering services
  - such services include web publishing, instant messaging servers, etc.
  - for example, a TOR user can setup a website where people publish material without worrying about censorship
  - nobody is able to determine who is offering the site and nobody knows who is posting to it
- These services are called hidden services, and setting up a hidden service includes
  - selecting a few onion routers as introduction points
  - advertising these points on the lookup service
  - building a circuit from each introduction point to the service

TOR Today

TOR underwent a lot of research and implementation efforts to defeat various traffic analysis attacks

It is currently being offered as a TOR browser

- it offers protection against traffic analysis
- it offers protection against web tracking
- see http://www.torproject.org



Being slower than other browsers is unavoidable

Summary

- Anonymous communication has many motivations for use by individuals, organizations, and the government
- Early proposals include mixes and proxies
- The onion routing (Tor) project provides a real-life system for achieving anonymous communications
  - http://www.torproject.org