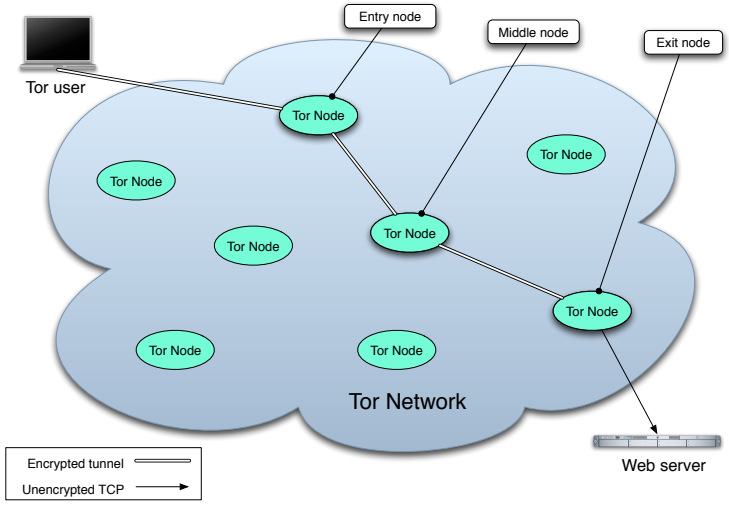
Anonymity Systems Requirements and Architecture



Steven J. Murdoch

http://www.cl.cam.ac.uk/users/sjm217/

Anonymous communications

- Anonymous communication systems allow people to communicate without giving away their identity (in practice, IP address)
- Sometimes used in their own right: e.g. for browsing the Internet
- Sometimes used with other privacy enhancing technologies: e.g. Private Information Retrieval, Anonymous credentials
- One growing use is censorship resistance

Construction of the Internet

- Internet Service Providers (ISPs) give people access to the Internet, and allow people to host services such as websites
- ISPs connect to other ISPs so the customers of one ISP can communicate with those of others
- Some ISPs have international connections, others use those of the bigger ISPs
- ISPs are typically based in one country and governed by the laws of that country
- Often governments control international Internet connections, either directly or via regulation

Construction of the Internet

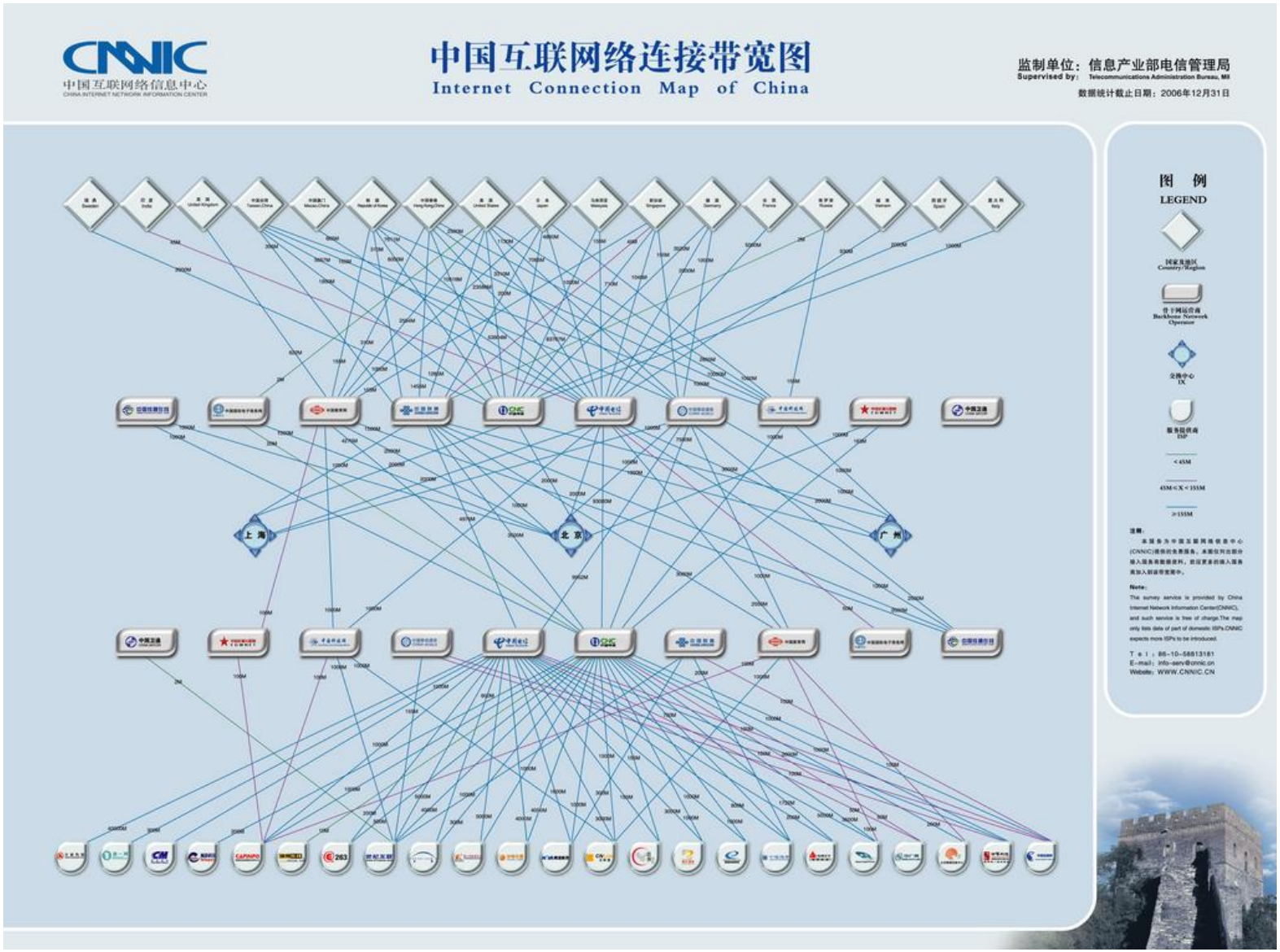
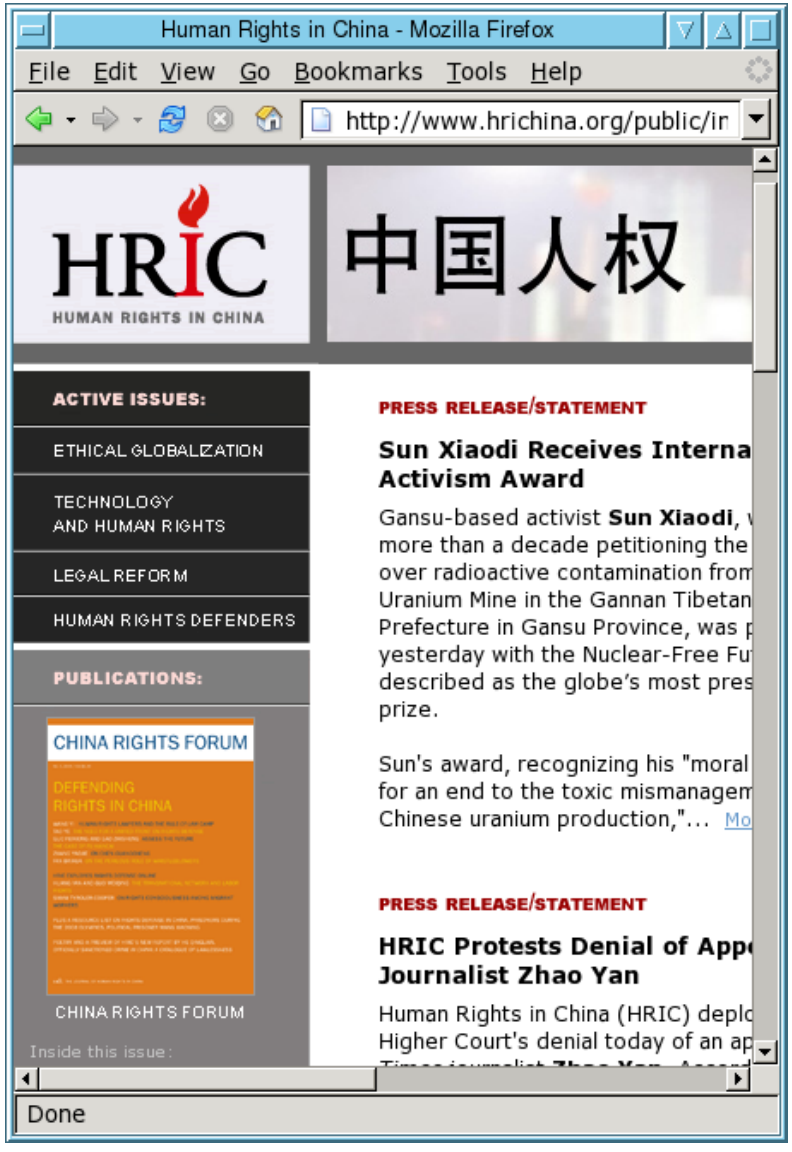


Diagram: China Internet Network Information Center

What is being blocked, and why

- Out of the 40 countries studied by the OpenNet Initiative in 2006, 26 censored the Internet in some way
- The types of material censored varied depending on country, e.g.:
  - Human Rights (blocked in China)
  - Religion (blocked in Saudi Arabia, UAE, Iran, Bahrain)
  - Pornography (blocked in Saudi Arabia, UAE, Iran, Bahrain, Singapore, Burma, . . .)
- Other issues censored include: military and militant websites; sex education, alcohol/drugs, music; gay and lesbian websites; news



What is being blocked, and why

- Out of the 40 countries studied by the OpenNet Initiative in 2006, 26 censored the Internet in some way
- The types of material censored varied depending on country, e.g.:
  - Human Rights (blocked in China)
  - Religion (blocked in Saudi Arabia, UAE, Iran, Bahrain)
  - Pornography (blocked in Saudi Arabia, UAE, Iran, Bahrain, Singapore, Burma, . . .)
- Other issues censored include: military and militant websites; sex education, alcohol/drugs, music; gay and lesbian websites; news

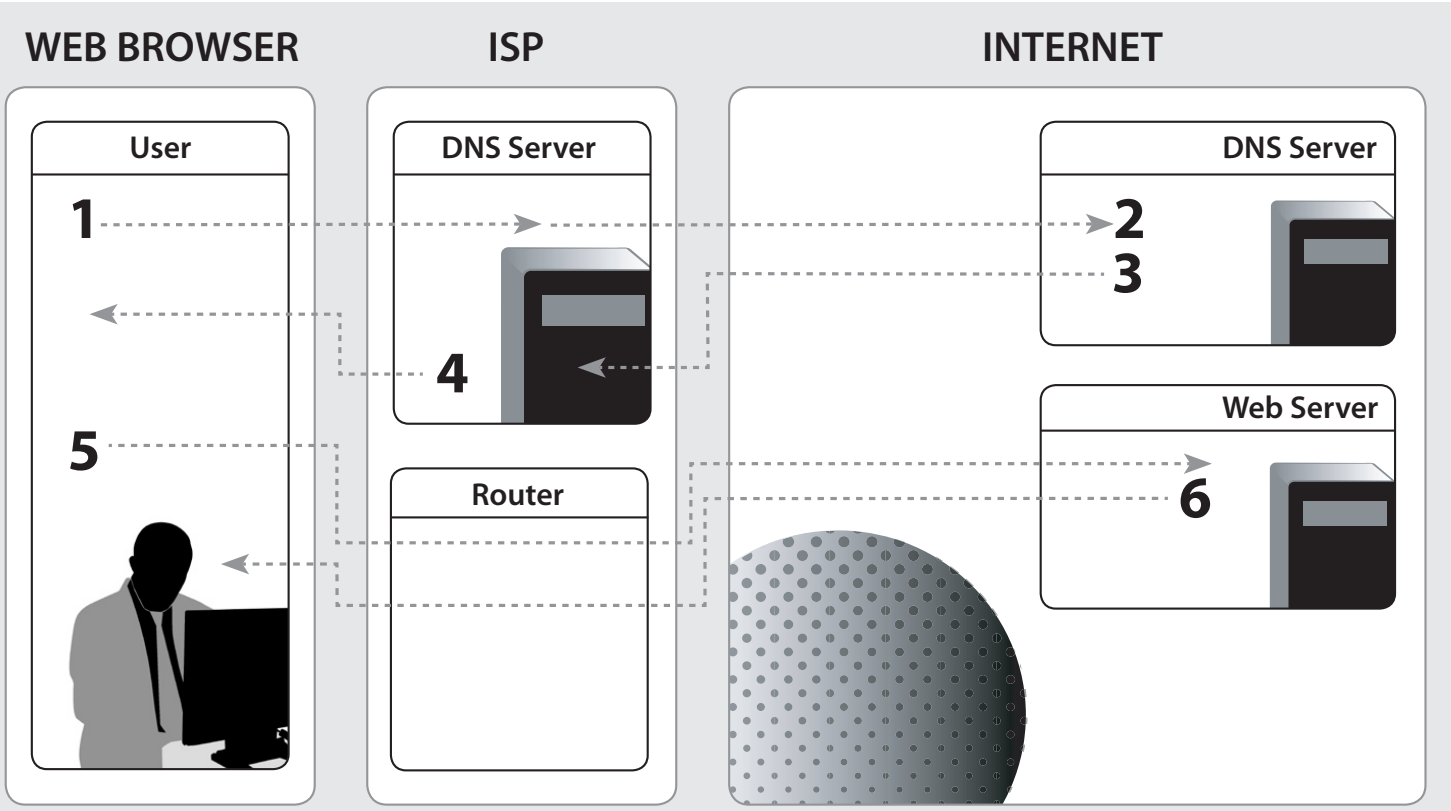
What is being blocked, and why

- Out of the 40 countries studied by the OpenNet Initiative in 2006, 26 censored the Internet in some way
- The types of material censored varied depending on country, e.g.:
  - Human Rights (blocked in China)
  - Religion (blocked in Saudi Arabia, UAE, Iran, Bahrain)
  - Pornography (blocked in Saudi Arabia, UAE, Iran, Bahrain, Singapore, Burma, . . .)
- Other issues censored include: military and militant websites; sex education, alcohol/drugs, music; gay and lesbian websites; news

Blocking with technology

- When a country's government controls international connectivity, they can block requests for banned websites
- There are a number of different approaches (DNS blocking, IP address blocking, etc.)
- Software may be produced in-country, but often is an adapted commercial product
- These companies not only make the software, but provide a continuously updated list of websites to be blocked

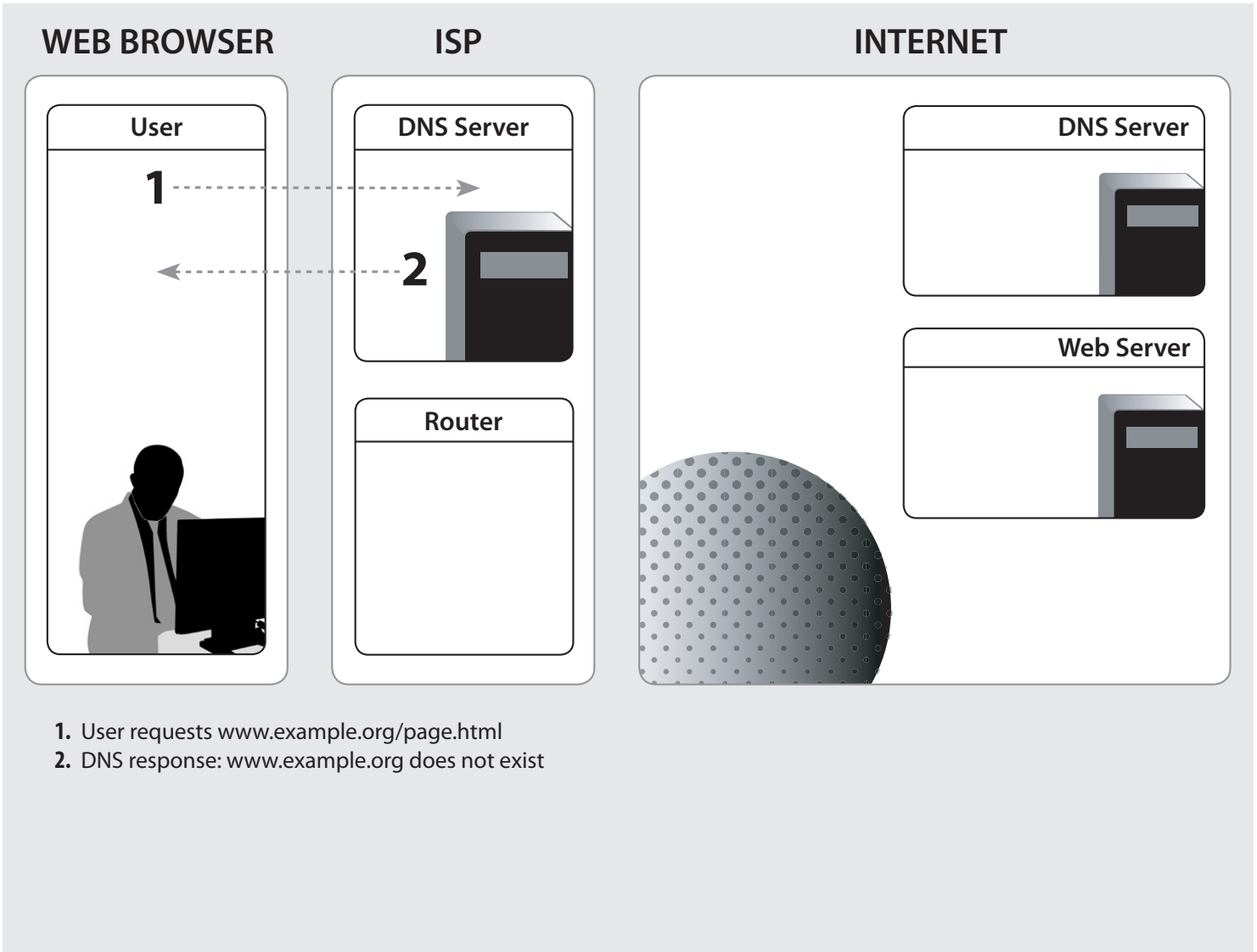
Normal web browsing



1. User requests www.example.org/page.html
2. DNS lookup for www.example.org
3. Lookup response: www.example.org is 192.0.2.166
4. www.example.org is 192.0.2.166
5. Get web page: www.example.org/page.html at 192.0.2.166
6. Here is www.example.org/page.html

Diagram: Jane Gowan

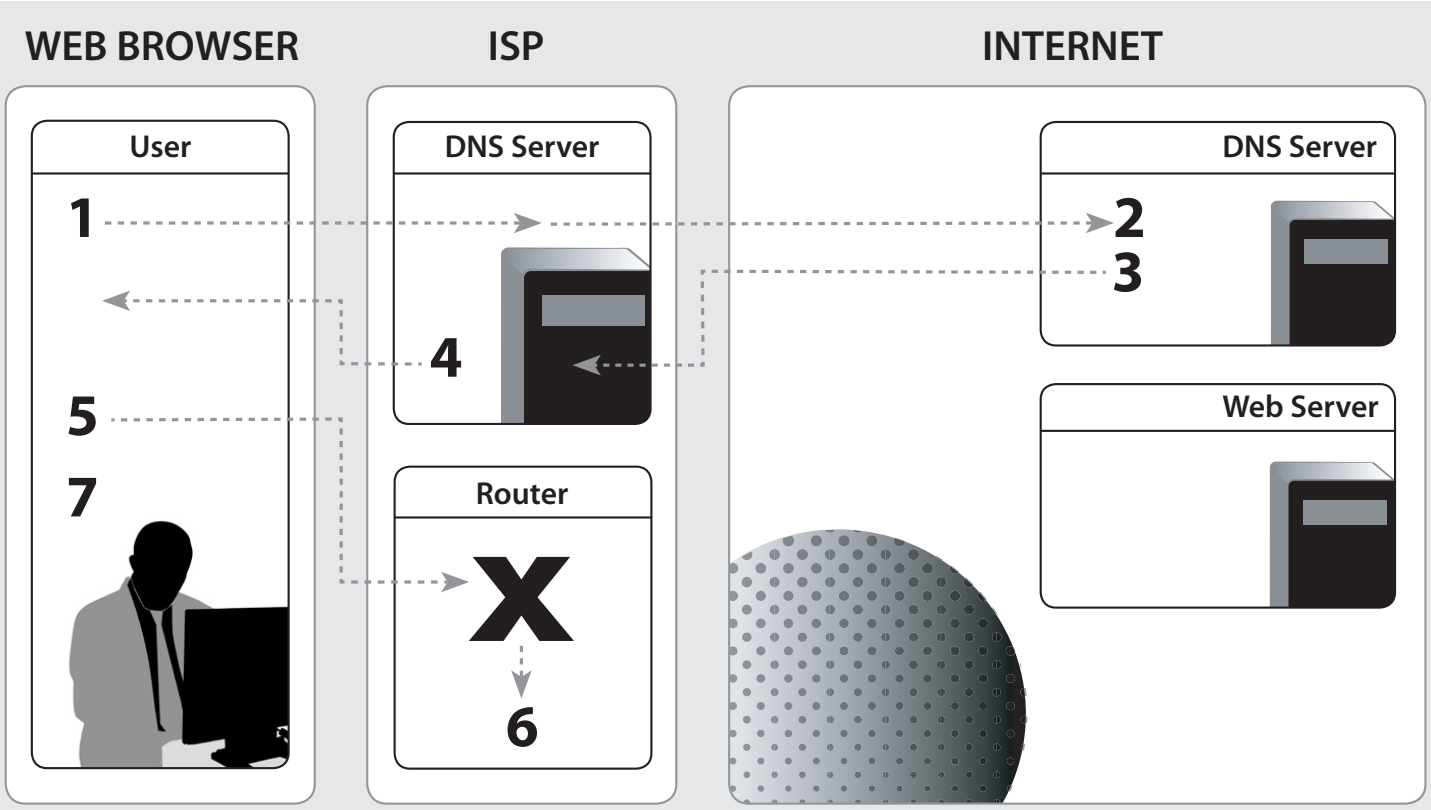
DNS tampering



1. User requests www.example.org/page.html
2. DNS response: www.example.org does not exist

Diagram: Jane Gowan

IP blocking



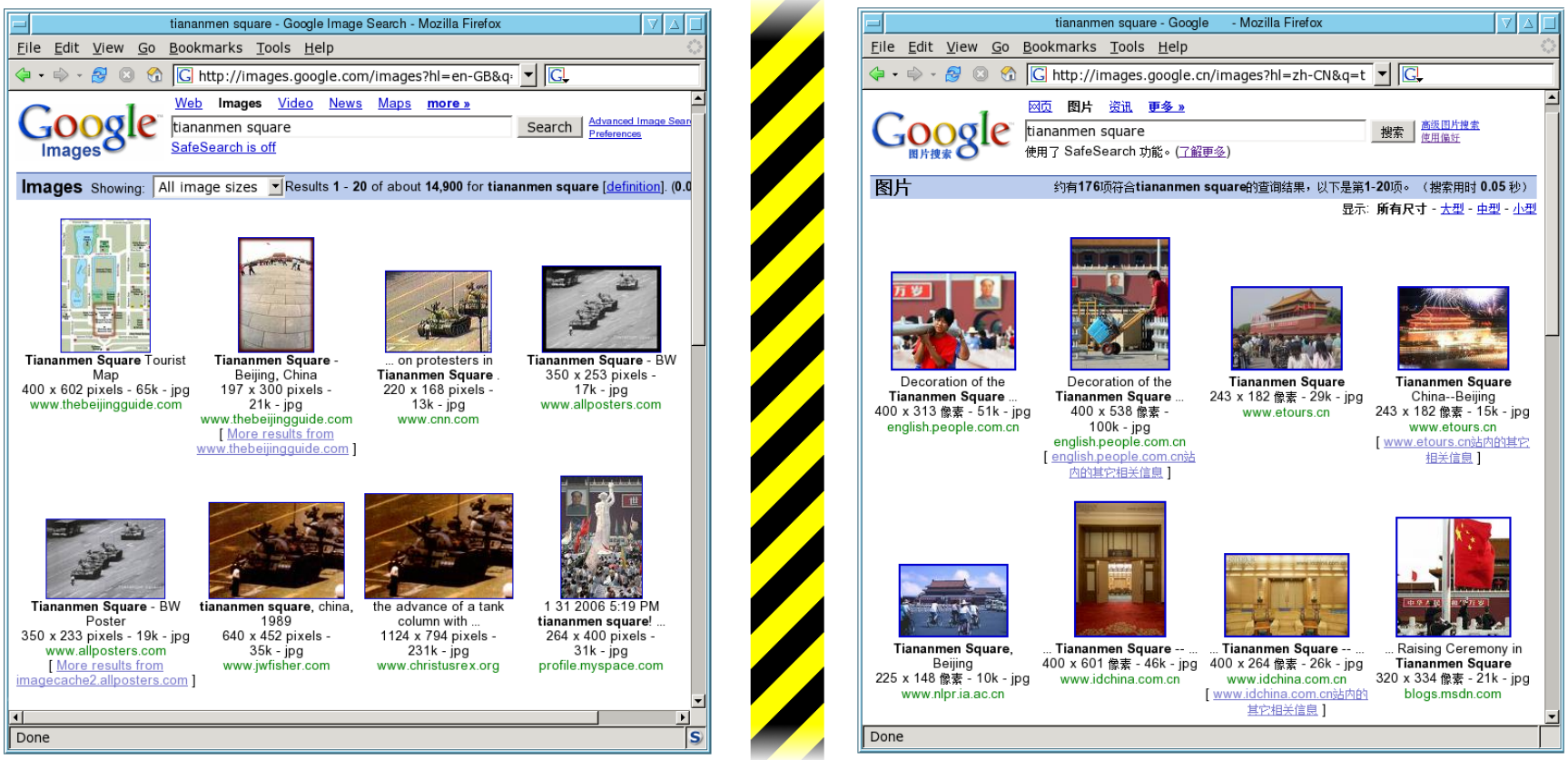
1. User requests www.example.org/page.html
2. DNS lookup for www.example.org
3. Lookup response: www.example.org is 192.0.2.166
4. www.example.org is 192.0.2.166
5. Get web page: www.example.org/page.html at 192.0.2.166
6. Router drops all packets to 192.0.2.166
7. Browser concludes that www.example.org is inaccessible

Diagram: Jane Gowan

Tradeoffs in blocking systems

- DNS blocking
  - Easy and cheap to implement
  - Blocks at domain name granularity – overblocks protocols, webpages
  - Trivial to bypass
- IP blocking
  - Easy and cheap to implement
  - Blocks at IP address (perhaps port) – overblocks virtual hosting
- Proxy blocking
  - Expensive to implement
  - Blocks at webpage level – low overblocking
- Hybrid blocking – IP based redirection to proxy
  - Tricky to get right, but cheap
  - Has some vulnerabilities
  - Blocks at webpage level – low overblocking

Even if a site is accessible, it may be removed from search engine results



Searching for “Tiananmen Square” on Google.com and Google.cn

Limitations of blocking

- Censorship systems block legitimate content and fail to block banned content
- It is fairly easy for readers and publishers to circumvent the technical measures
- Building and maintaining censorship systems is expensive
- Blocking one type of content encourages other types to be blocked
- Often the process of censorship is not transparent

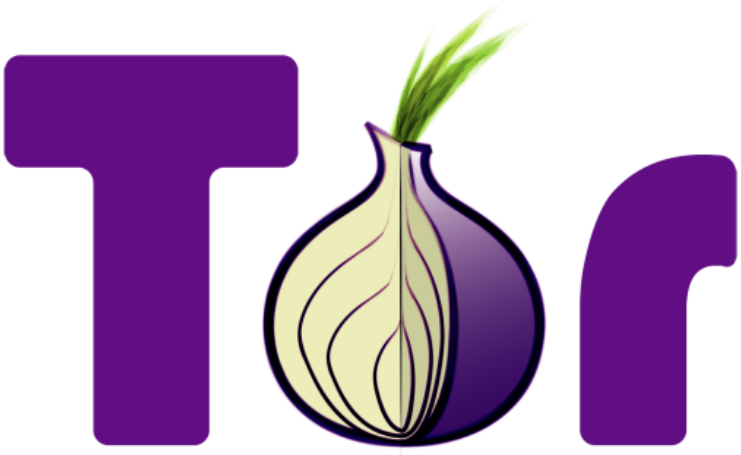
Photograph: David Gaya

Blocking through laws, fear, and intimidation

- ISPs may be forced to block sites themselves, or implement self-regulation
- People can be intimidated into not testing rules through fear of detection and retribution
- These may be through laws, social pressure or extra-legal punishment
- All these approaches may be used at the same time, and complement each other

Censorship resistance system requirements

- Software to resist censorship should
  - Hide where user is visiting (to prevent blocking)
  - Hide who the user is (to protect them from intimidation)
- These properties should be maintained even if the censorship resistance system is partially compromised

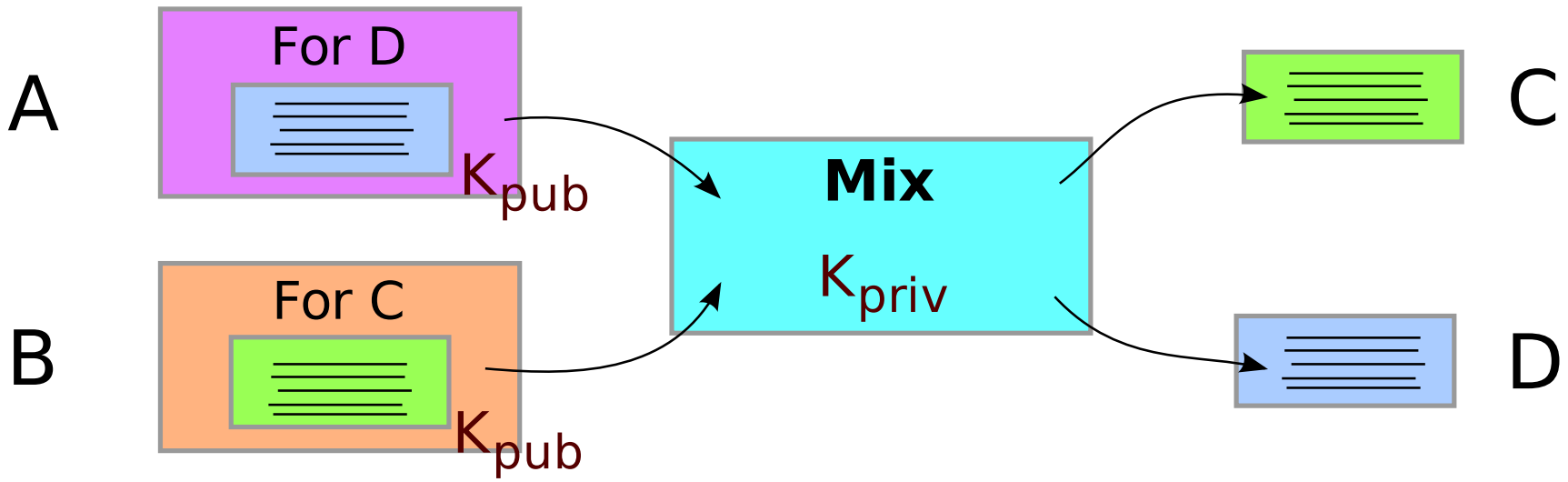


There are many other reasons why people might want privacy

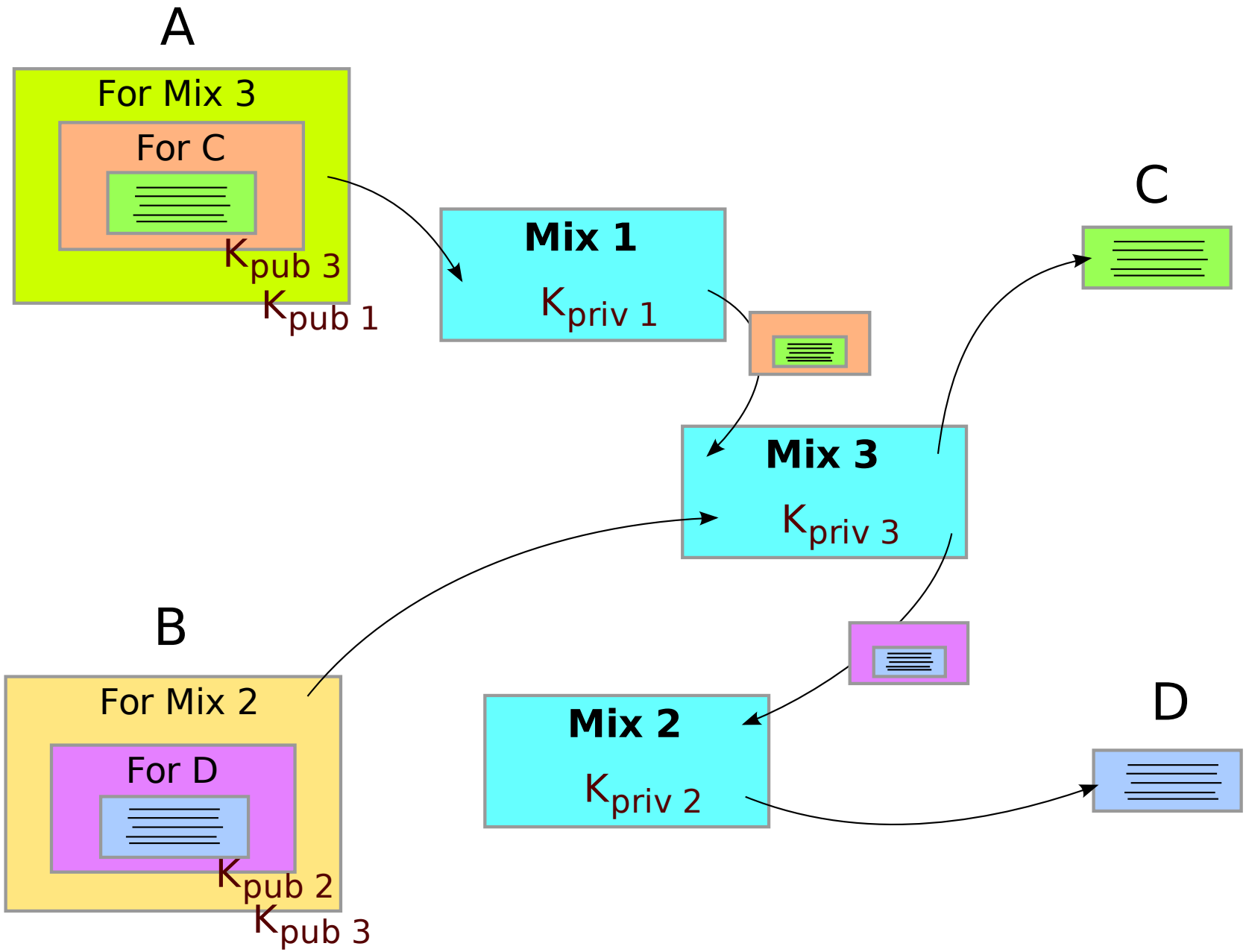
- Ordinary people
  - To avoid personal information being sold to marketers
  - Protect themselves when researching sensitive topics
- Militaries and law enforcement
  - To carry out intelligence gathering
  - Protect undercover field agents
  - Offer anonymous tip lines
- Journalists
  - To protect sources, such as whistle blowers
- Human rights workers
  - To publicise abuses and protect themselves from surveillance
  - Blogging about controversial subjects
- Businesses
  - To observe their competition and build anonymous collaborations

Anonymous communication

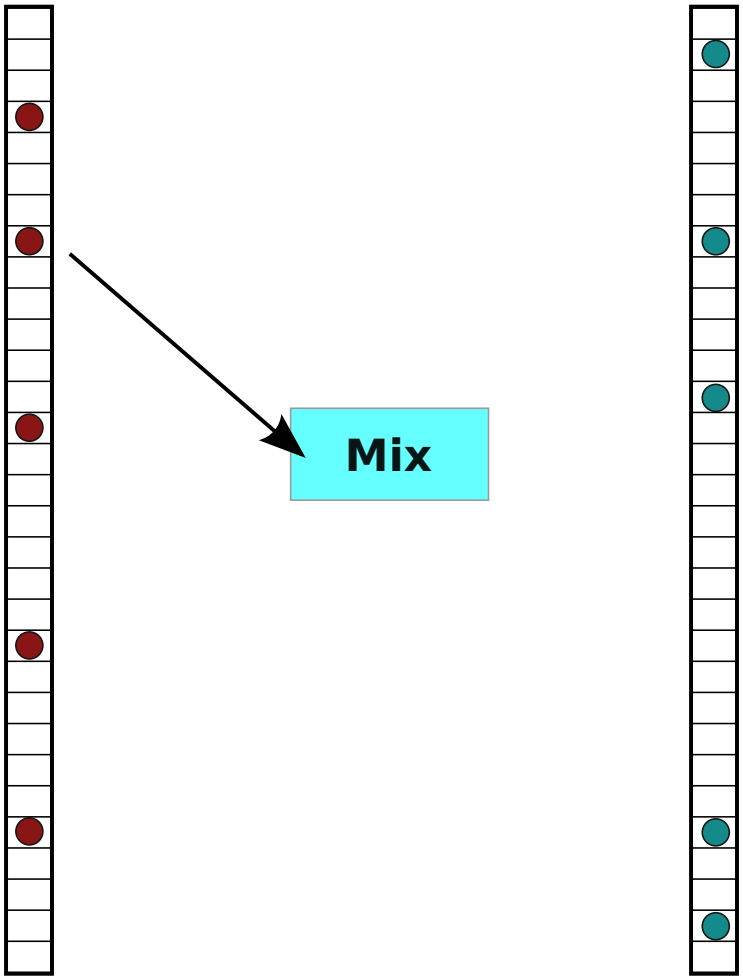
- People have to hide in a crowd of other people (“anonymity loves company”)
- The goal of the system is to make all users look as similar as possible, to give a bigger crowd
- Hide who is communicating with whom
- Layered encryption and random delays hide correlation between input traffic and output traffic



Remailers

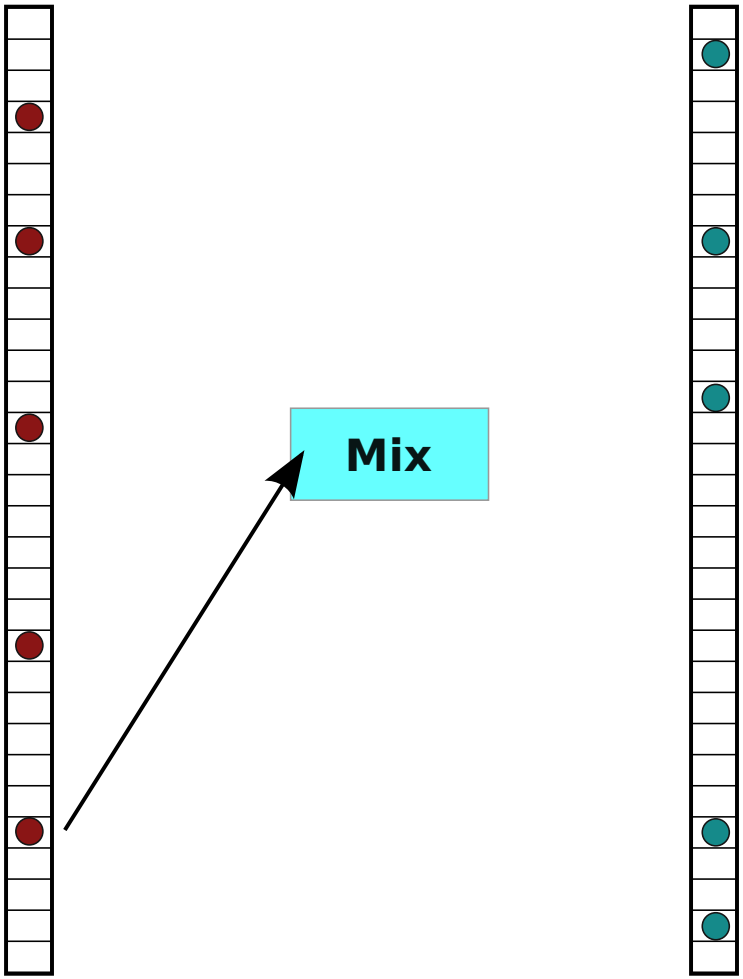

A
For Mix 3
For C
$K_{pub\ 3}$
$K_{pub\ 1}$
Mix 1
$K_{priv\ 1}$
C
Mix 3
$K_{priv\ 3}$
B
For Mix 2
For D
$K_{pub\ 2}$
$K_{pub\ 3}$
Mix 2
$K_{priv\ 2}$
D

Threshold mix



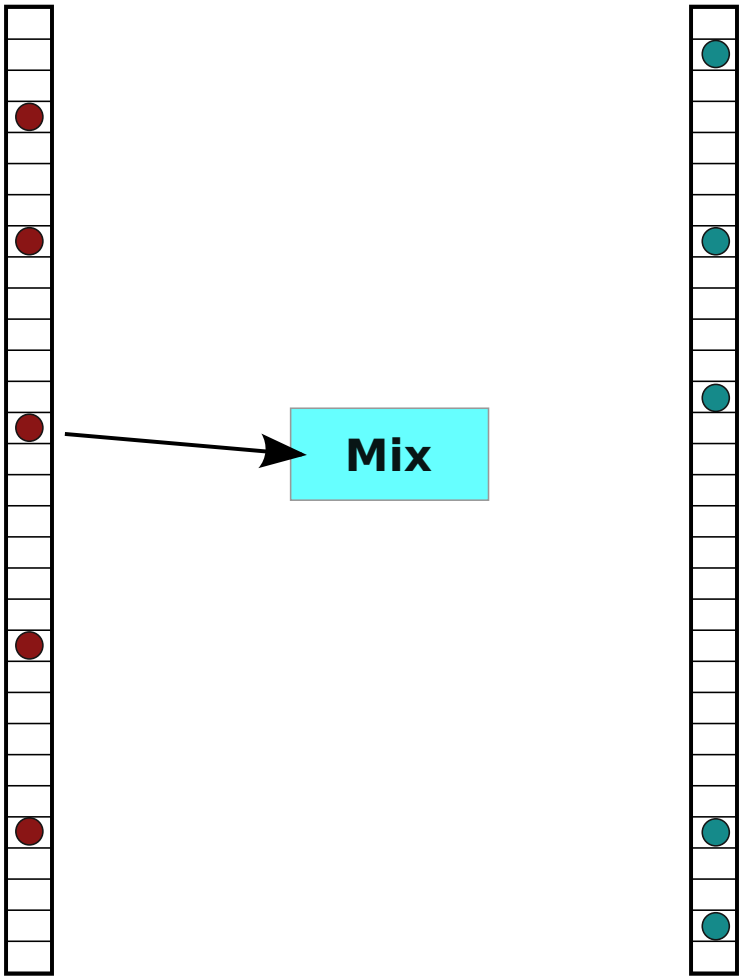
- In each round, the “threshold mix” accepts a fixed number of messages
- Once the number of messages reaches the “batch size” the mix flushes and sends them all, in a random order
- Other strategies are possible, but this is the type of mix we will examine in the exercise
- After observing one round, the attacker knows the set of senders and receivers, but not who sent each message

Threshold mix



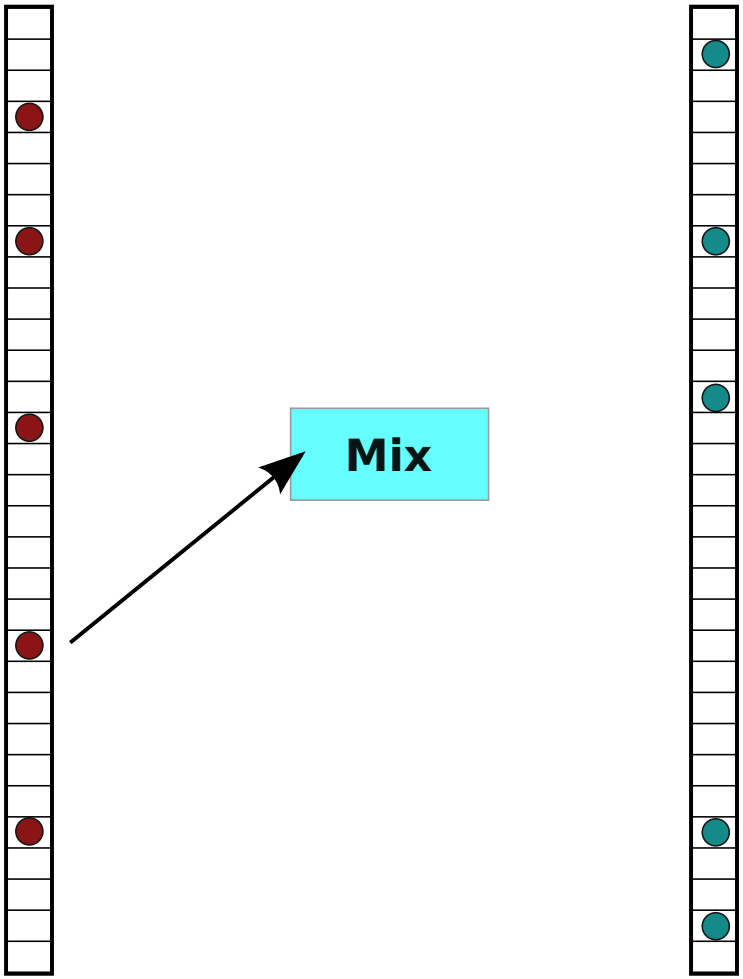
- In each round, the “threshold mix” accepts a fixed number of messages
- Once the number of messages reaches the “batch size” the mix flushes and sends them all, in a random order
- Other strategies are possible, but this is the type of mix we will examine in the exercise
- After observing one round, the attacker knows the set of senders and receivers, but not who sent each message

Threshold mix



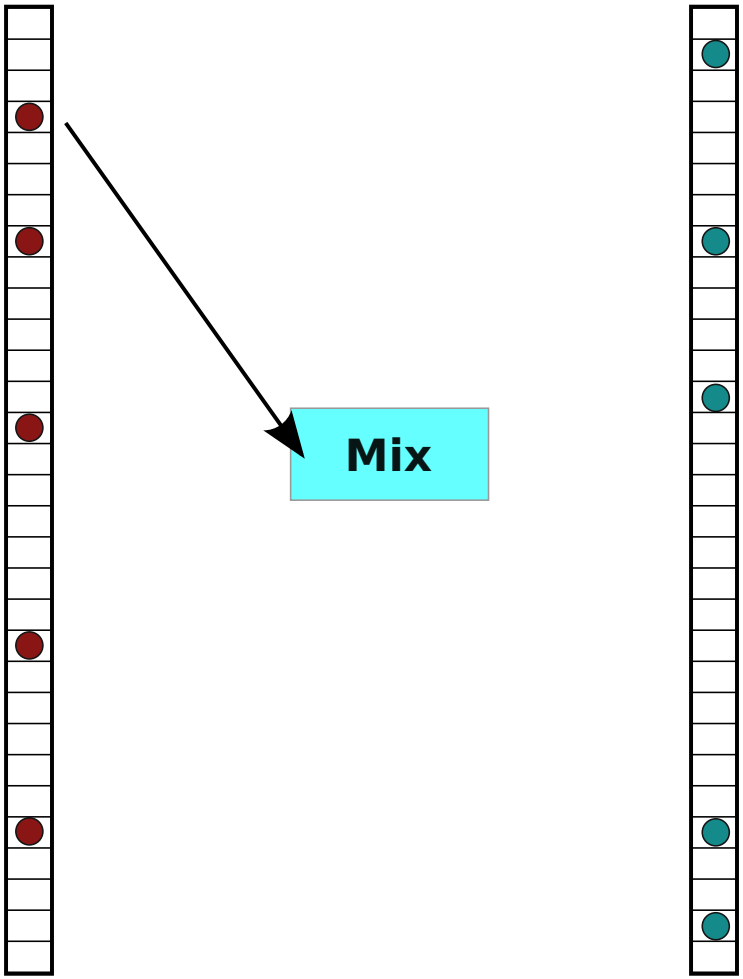
- In each round, the “threshold mix” accepts a fixed number of messages
- Once the number of messages reaches the “batch size” the mix flushes and sends them all, in a random order
- Other strategies are possible, but this is the type of mix we will examine in the exercise
- After observing one round, the attacker knows the set of senders and receivers, but not who sent each message

Threshold mix



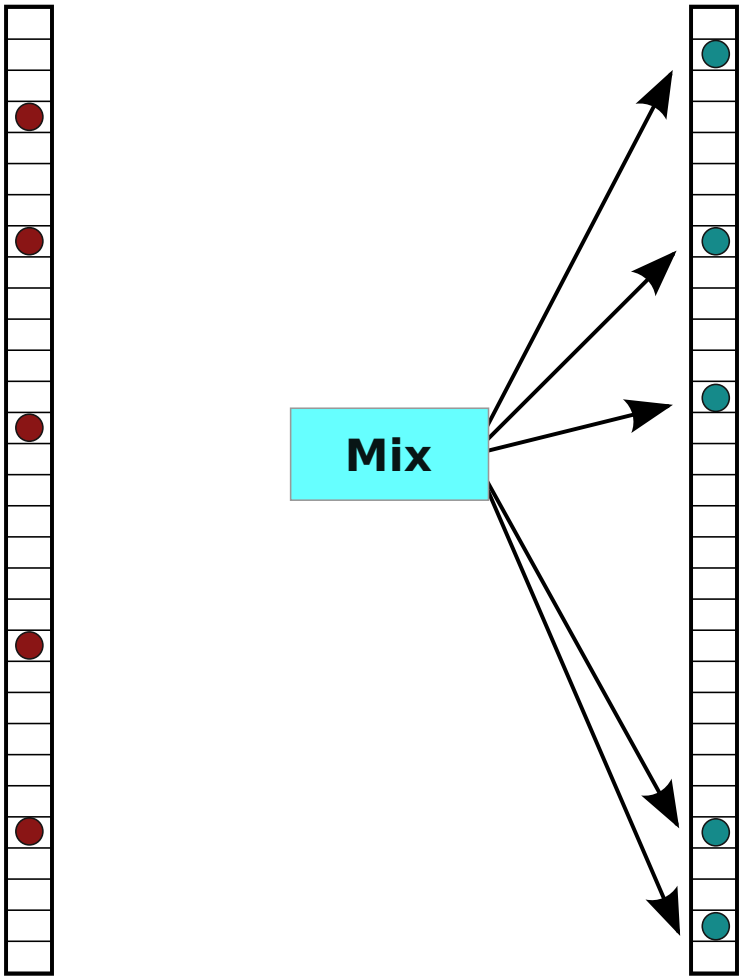
- In each round, the “threshold mix” accepts a fixed number of messages
- Once the number of messages reaches the “batch size” the mix flushes and sends them all, in a random order
- Other strategies are possible, but this is the type of mix we will examine in the exercise
- After observing one round, the attacker knows the set of senders and receivers, but not who sent each message

Threshold mix



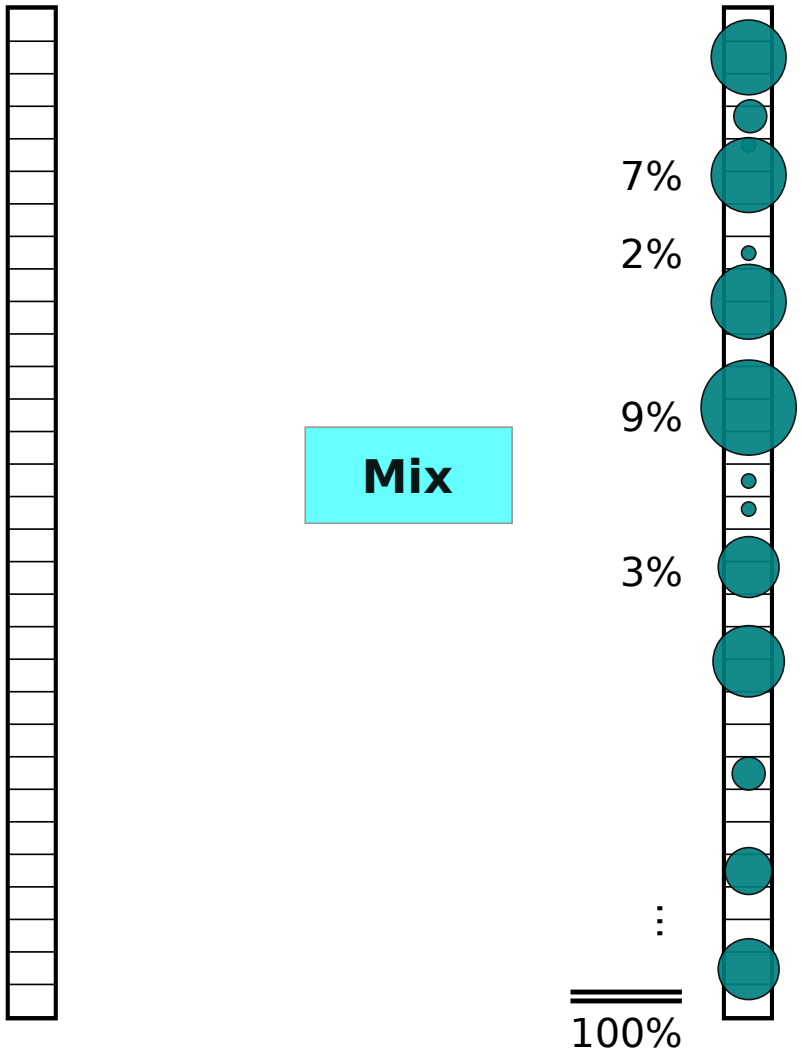
- In each round, the “threshold mix” accepts a fixed number of messages
- Once the number of messages reaches the “batch size” the mix flushes and sends them all, in a random order
- Other strategies are possible, but this is the type of mix we will examine in the exercise
- After observing one round, the attacker knows the set of senders and receivers, but not who sent each message

Threshold mix



- In each round, the “threshold mix” accepts a fixed number of messages
- Once the number of messages reaches the “batch size” the mix flushes and sends them all, in a random order
- Other strategies are possible, but this is the type of mix we will examine in the exercise
- After observing one round, the attacker knows the set of senders and receivers, but not who sent each message

Traffic Analysis

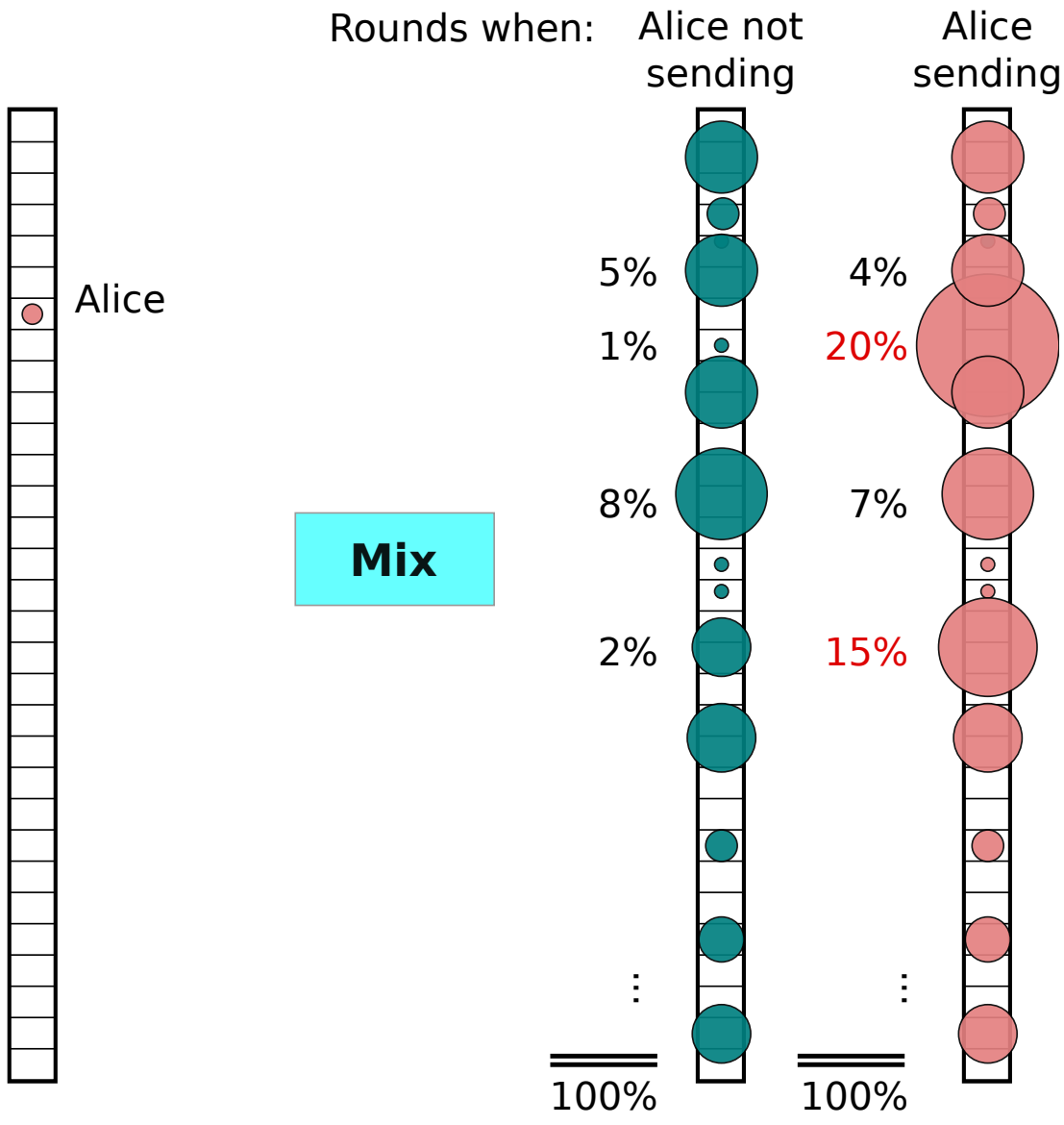


- By observing traffic over many rounds, the adversary can count each recipient's share of the messages received
- Some users will receive more messages than others
- These users may be of interest, so the target of further investigation
- e.g. Bob's share is:

$$\frac{\text{messages received by Bob}}{\text{messages received in total}}$$

over all rounds

Tracking Alice's Contacts



- Can observe each Bob's share in both rounds where Alice was sending, and rounds where she was not
- Recipients whose share jumps when Alice is sending are likely Alice's friends
- Score = (Bob's share in rounds where Alice is sending) − (Bob's share in rounds where Alice **not** sending)

Anonymity systems exist for hiding both email and web traffic

- Hiding web traffic is a fundamentally more difficult problem than hiding email
- Anonymity is achieved by making all traffic look the same (padding) and hiding timing correlations (delays)
- Web traffic is very variable (few kB to few GB): so padding doesn't work well
- Long latencies would be intolerable for interactive traffic: so adding delays don't work well
- However it is not all bad: anonymity needs other users to hide in
- There is much more web traffic than there is email, so this partially makes up for the lower security

Tor is a low-latency anonymity system

- Based on technology developed in the Onion Routing project
- Commonly used for web browsing (works for any TCP traffic)
- Originally built as a pure anonymity system (hides who is talking to whom)
- Now designed to resist censorship too (hides whether someone is using the system at all)
- Centralised directory authorities publish a list of all servers

Tor hides communication patterns by relaying data through volunteer servers

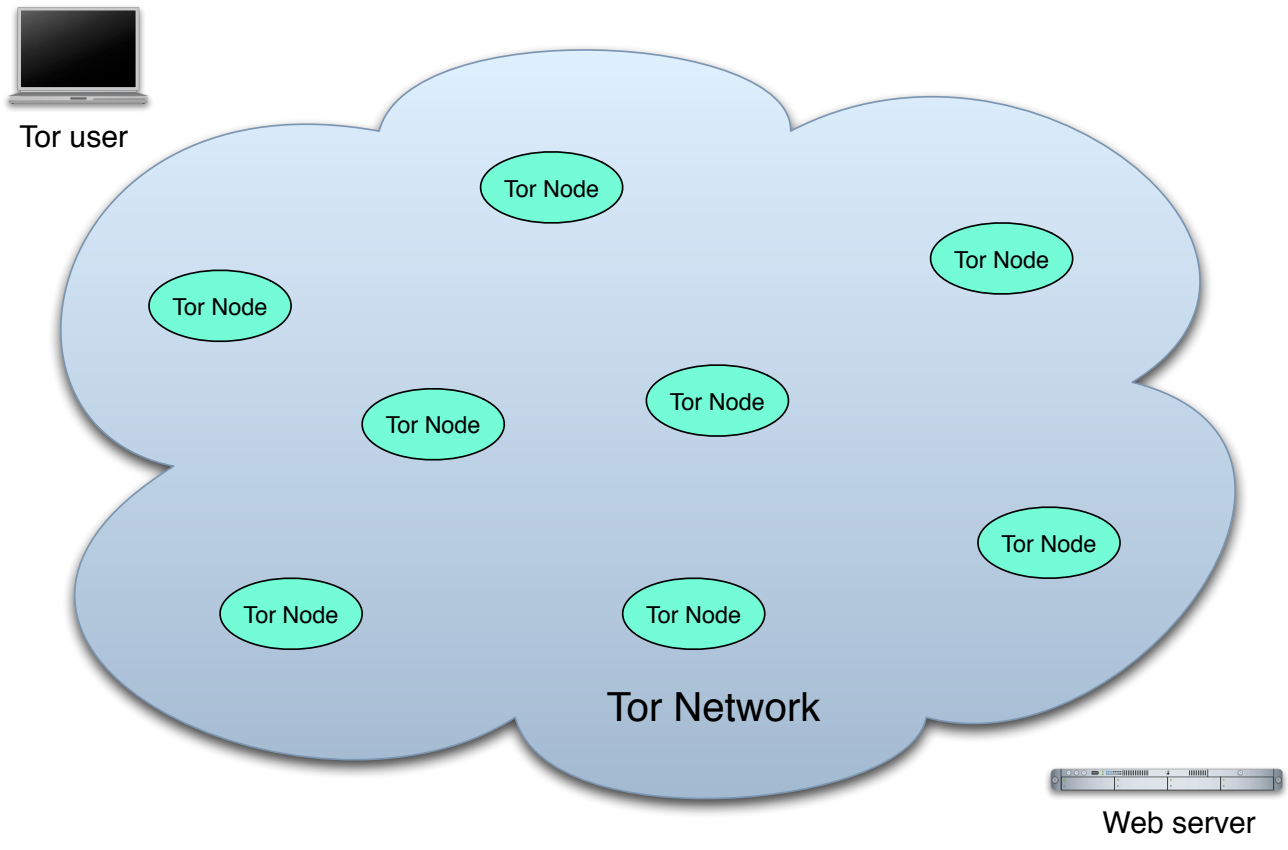


Diagram: Robert Watson

Tor hides communication patterns by relaying data through volunteer servers

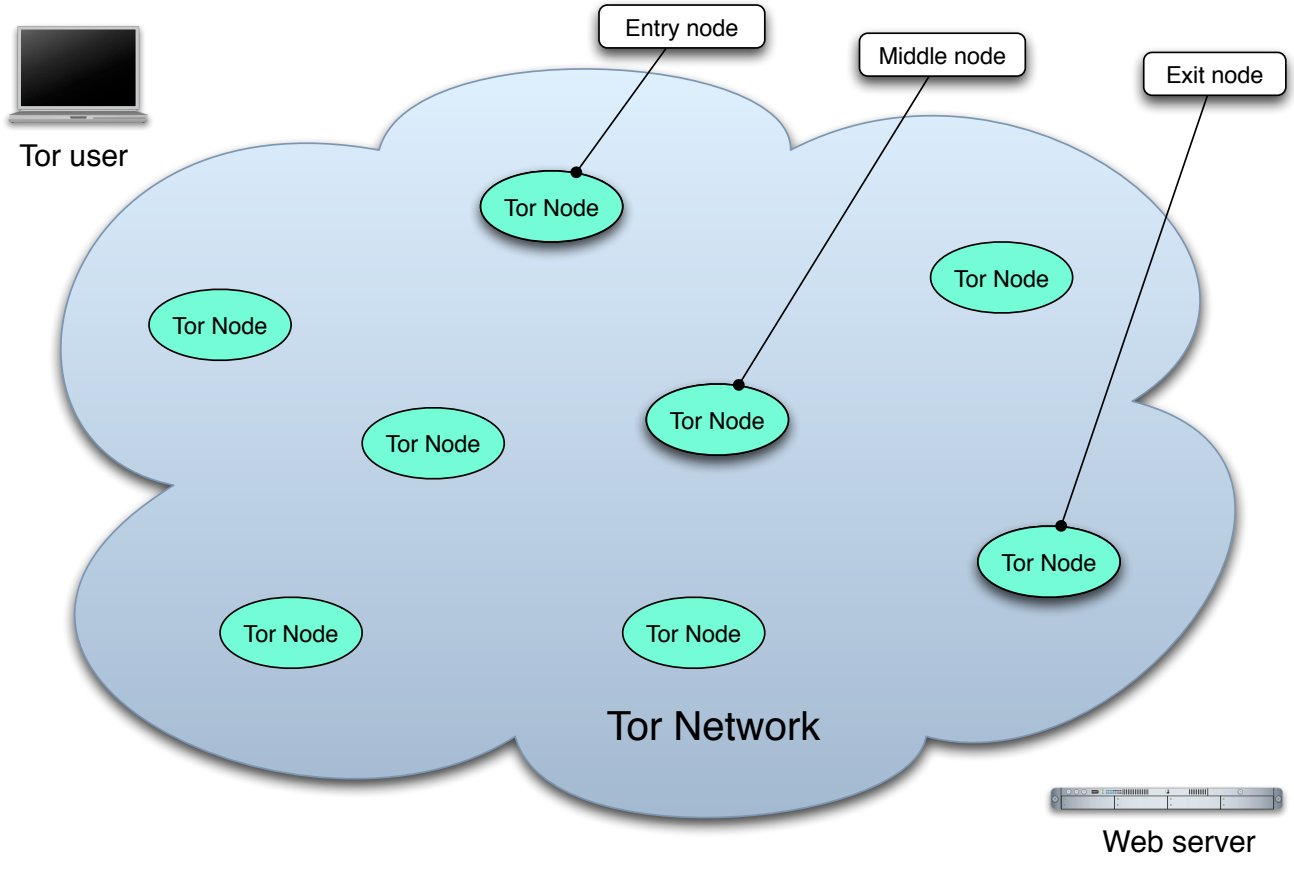


Diagram: Robert Watson

Tor hides communication patterns by relaying data through volunteer servers

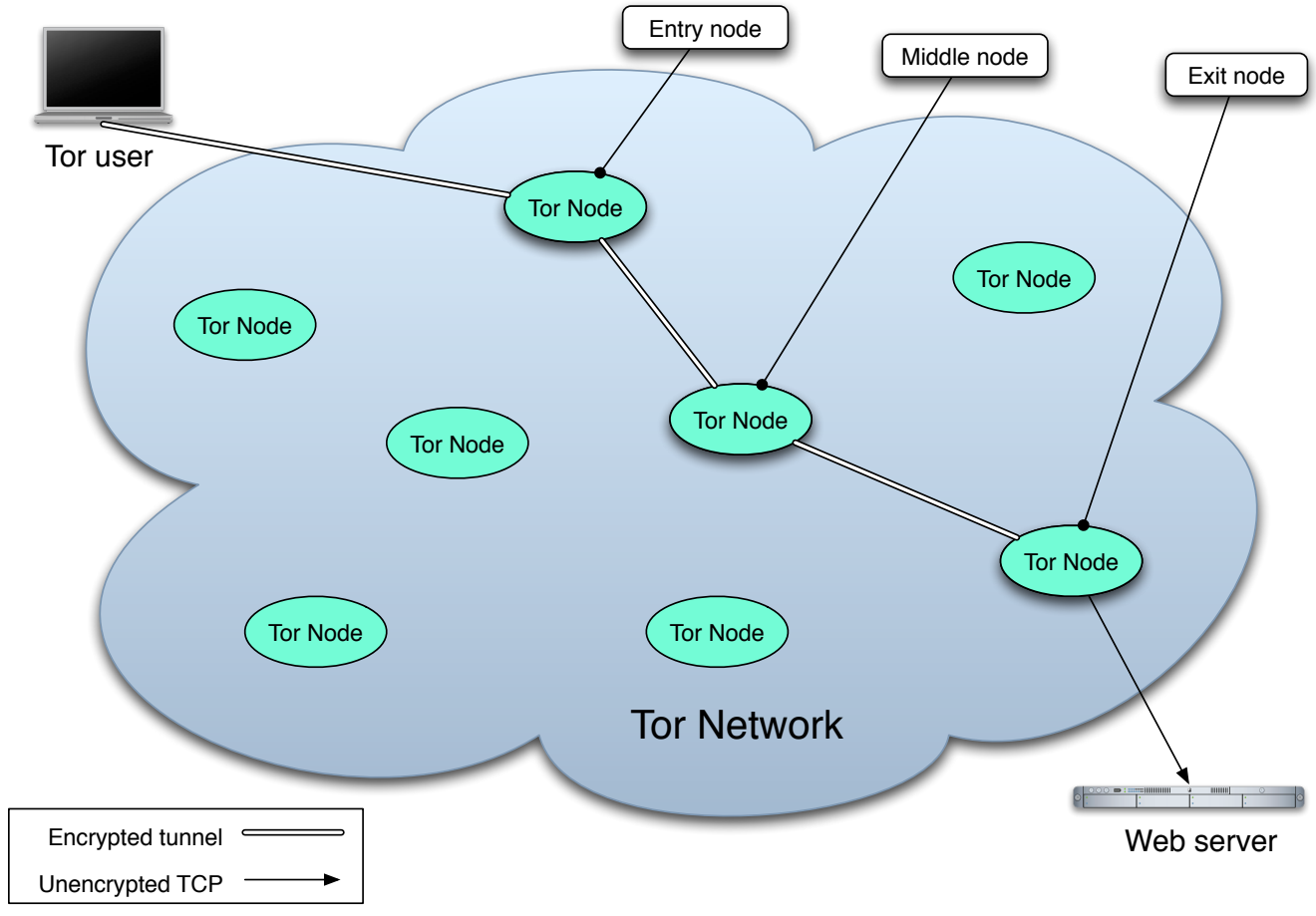
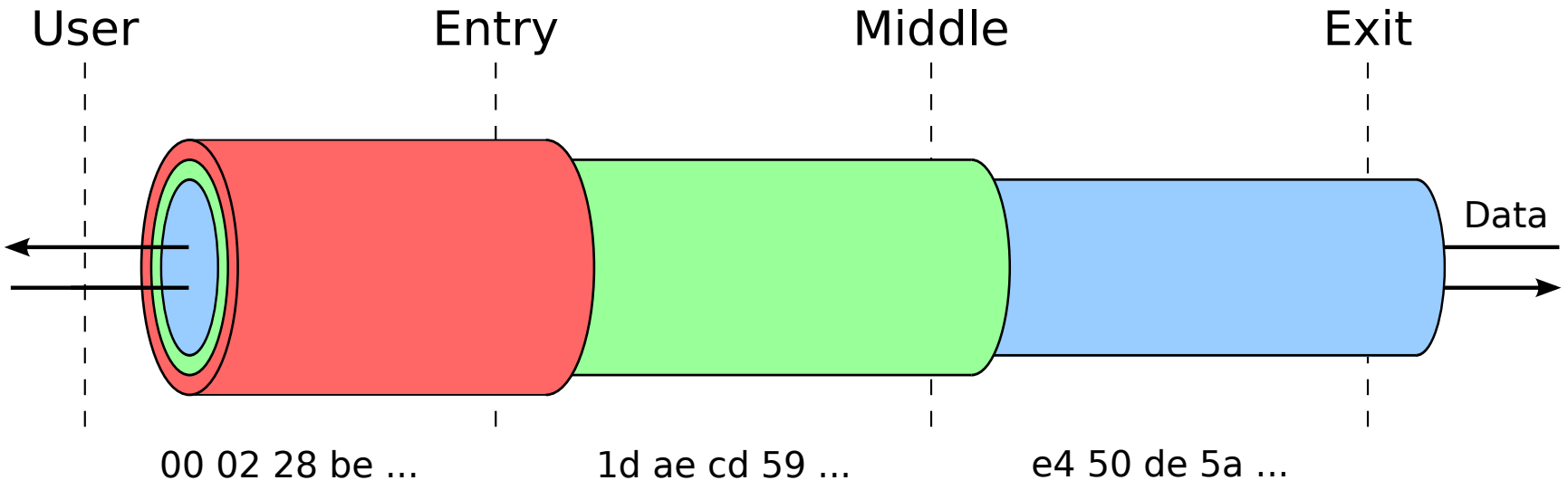


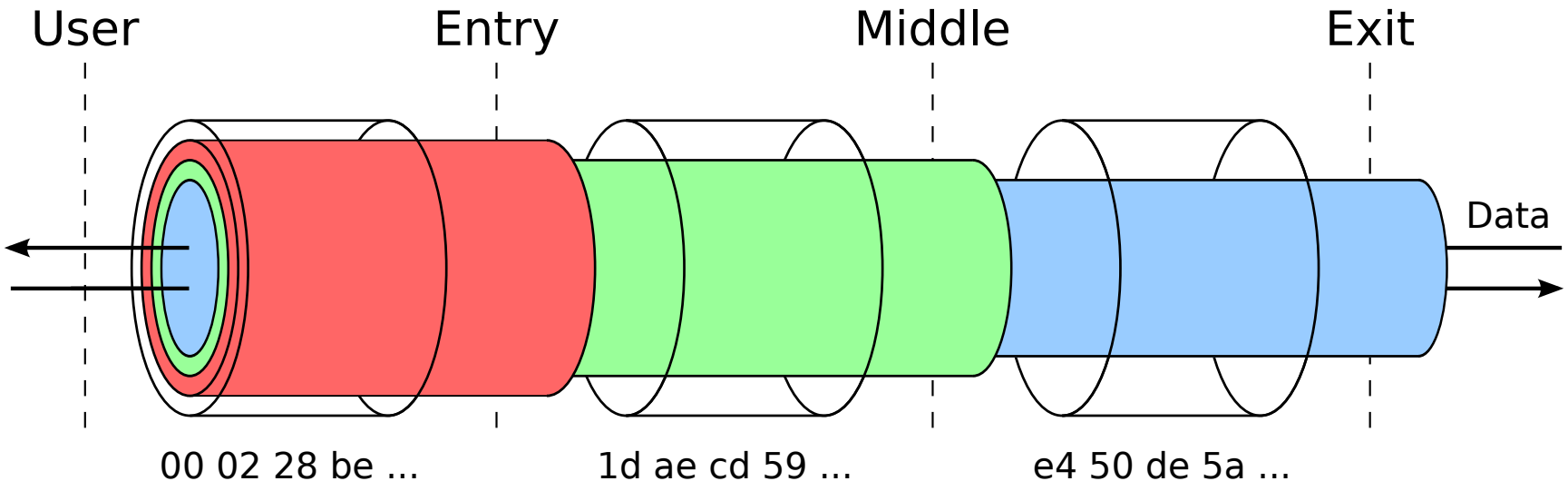
Diagram: Robert Watson

Tor uses two types of encryption



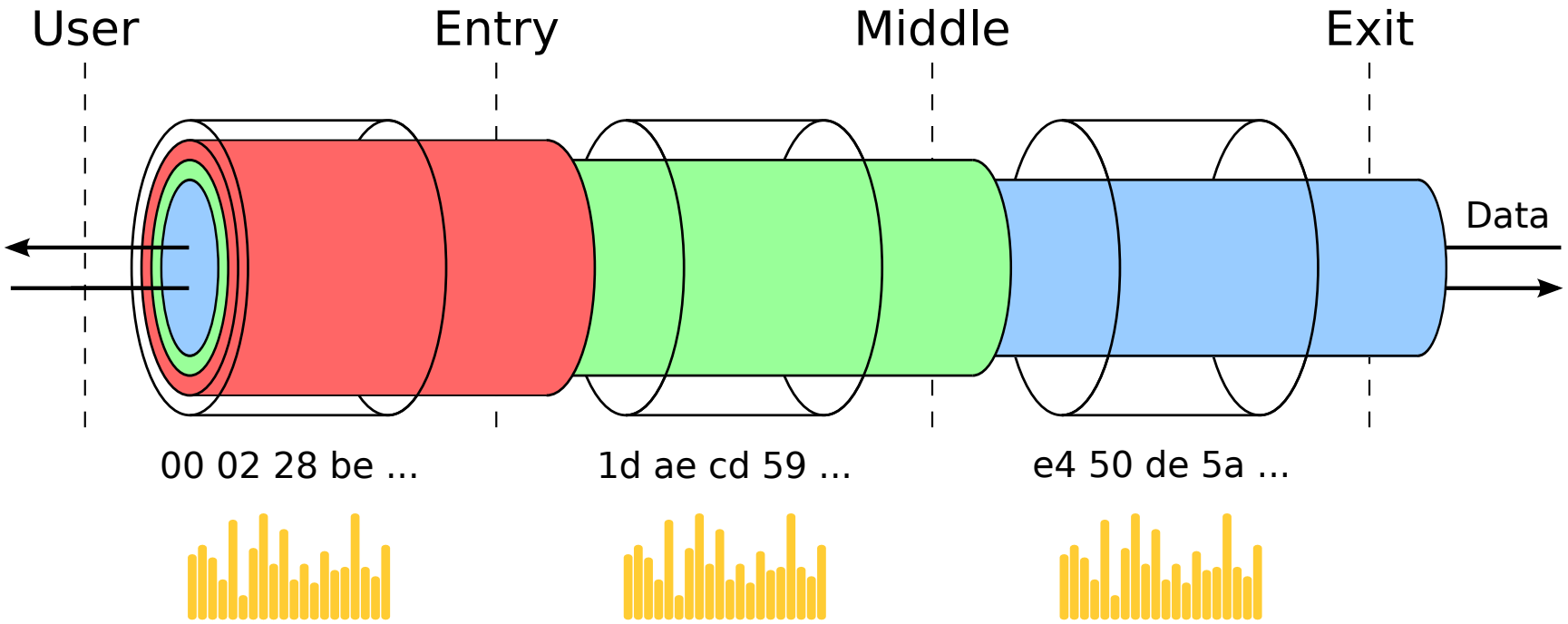
Circuit encryption unlinks data entering and leaving a server

Tor uses two types of encryption



Circuit encryption unlinks data entering and leaving a server
Link encryption (TLS) disguises individual circuits

Tor uses two types of encryption



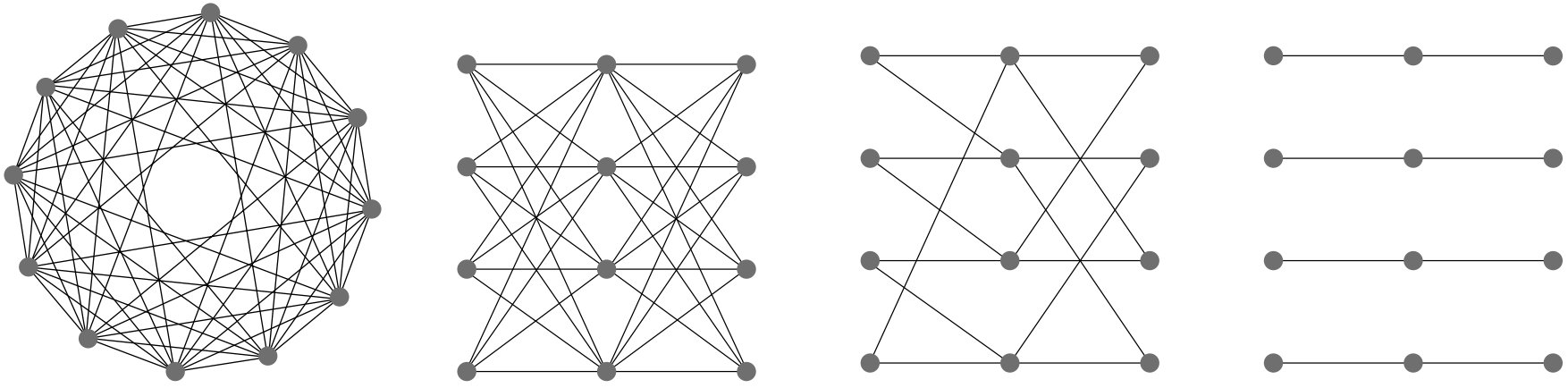
Circuit encryption unlinks data entering and leaving a server
Link encryption (TLS) disguises individual circuits
But data rate is unchanged so traffic analysis can correlate flows

Architectural options: traffic analysis resistance

Resisting traffic analysis comes with a cost

- It requires either long delays, lots of padding, dropping messages, or some combination of these
- Even performing these steps is not sufficient against an adversary who can infiltrate the network
- Delays are used for email anonymous communication systems (e.g. MixMinion and MixMaster)
- Padding and dropping are not used in any widely deployed system
- Tor gains resistance by being well used and widely distributed

Architectural options: topology



Choosing how nodes can connect to other nodes affects:

- Performance
- Security
- Scalability

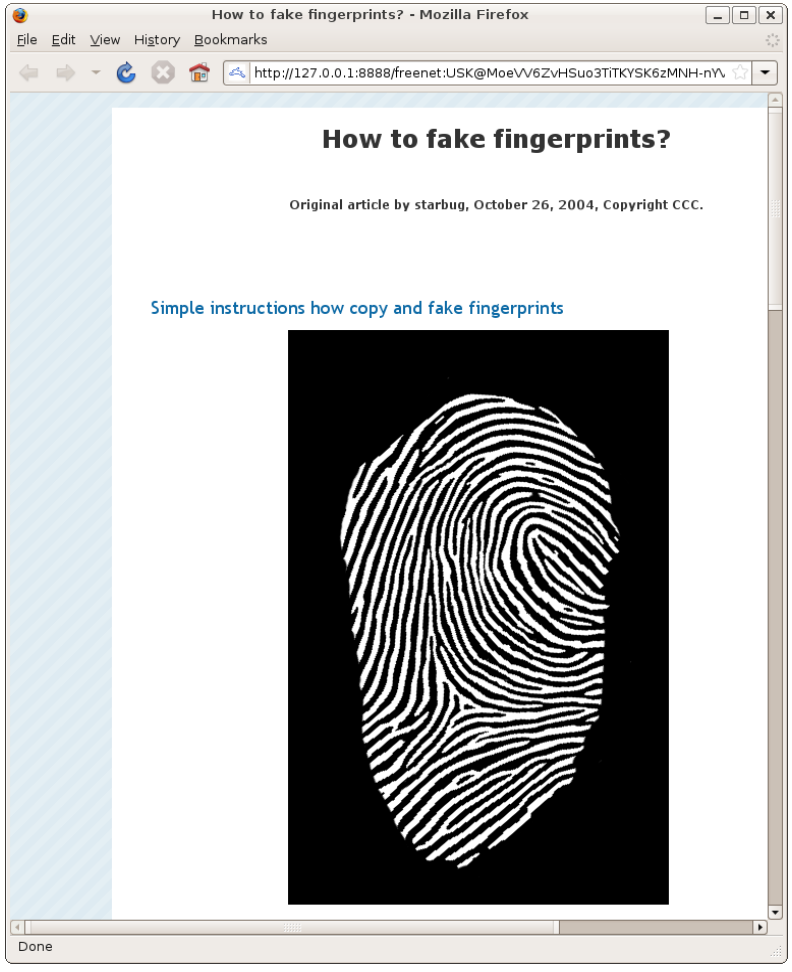
Architectural options: path selection

Different anonymity systems take different approaches to path selection

- Tor and remailers have a central database and source routing
  - Implementation is easier
  - Central database is a point of vulnerability
- P2P systems let nodes choose next hop on path
  - Reduces resistance to compromised nodes
  - Other attacks become possibl

Freenet is an anonymous content distribution network

- While Tor allows access to the Internet, Freenet creates a private network
- Users can create websites, share files and send/receive emails between other members of the network
- Content is hosted by sharing it amongst users of the network
- Users cannot select what content they host, and it is stored in an encrypted form



Psiphon a is censorship resistance system with different tradeoffs to Tor

- There is no centralized control, so it is hard to block but also hard for user to find a server
- Users do not have to download software, but this limits the strength of protection
- If the user cannot modify browser settings or install software, Psiphon is still usable
- Users within a censored country can ask someone they trust outside of the country to install the Psiphon server

Further information

"Tools and Technology of Internet Filtering", a chapter in "Access Denied".
`http://opennet.net/accessdenied`

"Security Engineering", 2nd Edition (Chapter 23).
`http://www.cl.cam.ac.uk/~rja14/book.html`

The anonymity bibliography
`http://www.freehaven.net/anonbib/`

The Tor Project website
`https://www.torproject.org/`

A copy of these slides will be available
`http://www.cl.cam.ac.uk/~sjm217/`