Theory and Design of Low-latency Anonymity Systems (Lecture 1)

Paul Syverson
U.S. Naval Research Laboratory
syverson@itd.nrl.navy.mil
http://www.syverson.org

Course Outline

Lecture 1:

- Usage examples, basic notions of anonymity, types of anonymous comms systems
- Crowds: Probabilistic anonymity, predecessor attacks

Lecture 2:

- Onion routing basics: simple demo of using Tor, network discovery, circuit construction, crypto, node types and exit policies
- Economics, incentives, usability, network effects

Course Outline

Lecture 3:

- Formalization and analysis, possibilistic and probabilistic definitions of anonymity
- Hidden services: responder anonymity, predecessor attacks revisited, guard nodes

Lecture 4:

- Link attacks
- Trust

Preliminaries

Lots of collaborators in what I am presenting. Some of the main ones, alphabetically:

George Danezis, Roger Dingledine, Matt Edman, Joan Feigenbaum, Aaron Johnson, Nick Mathewson, Lasse Øverlier

I try to remember to cite work of others as I go. Full citations should be in....

Preliminaries

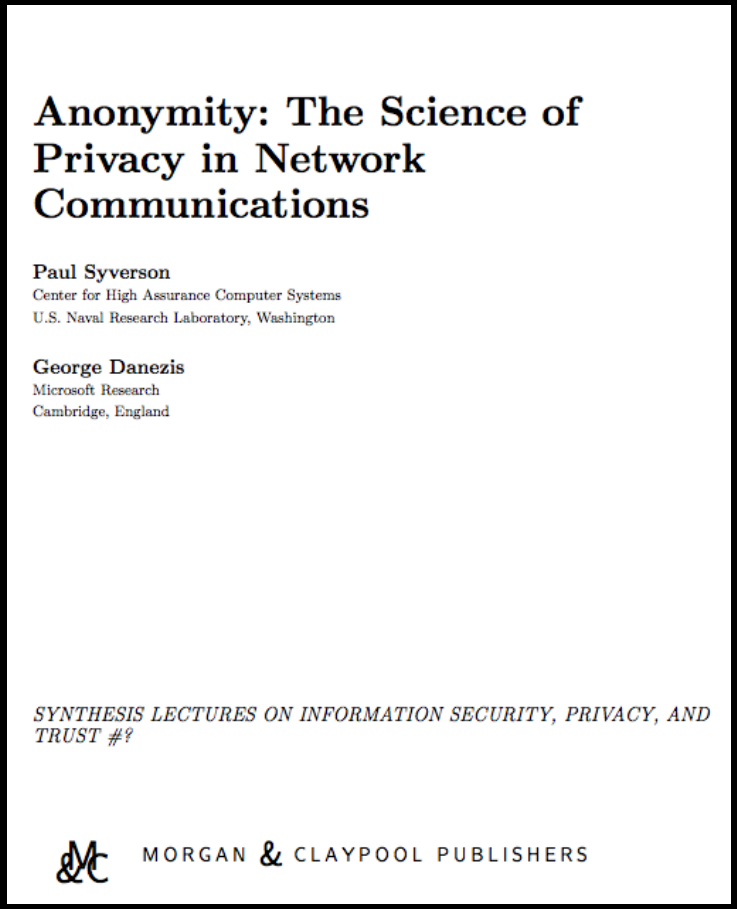
Book forthcoming in 2007.

Full draft in 1-3 months.

We would be happy to give a draft to any attendee of these lectures.

Especially we would like to get your comments.

Contact George or me if you want a copy.



Preliminaries

Please interrupt if you have questions, want clarification, etc.

Preliminaries

Please interrupt if you have questions, want clarification, etc.

In bocca al lupo.

Anonymous communications

Technical	Governmental/Social
1. What is it?	
	2. Why does it matter?
3. How do we build it?	

1.
What is anonymity anyway?

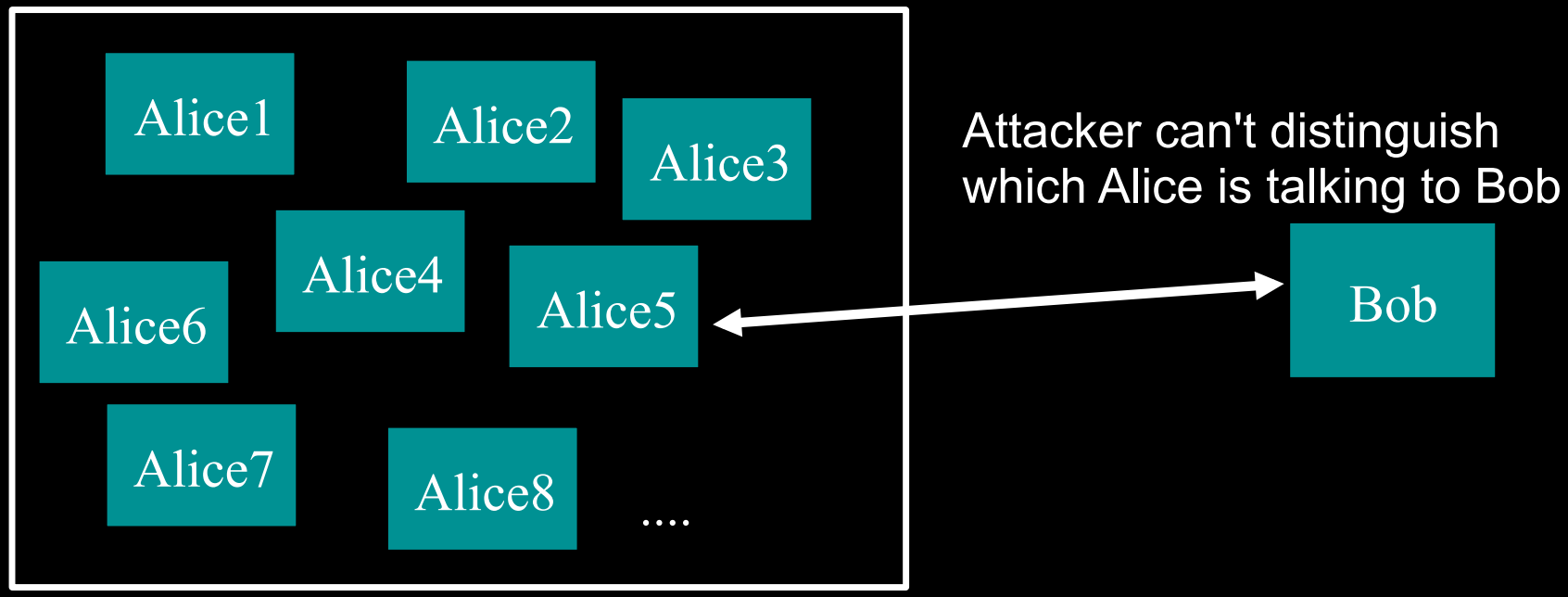
Informally: anonymity means you can't tell who did what

"Who wrote this blog post?"

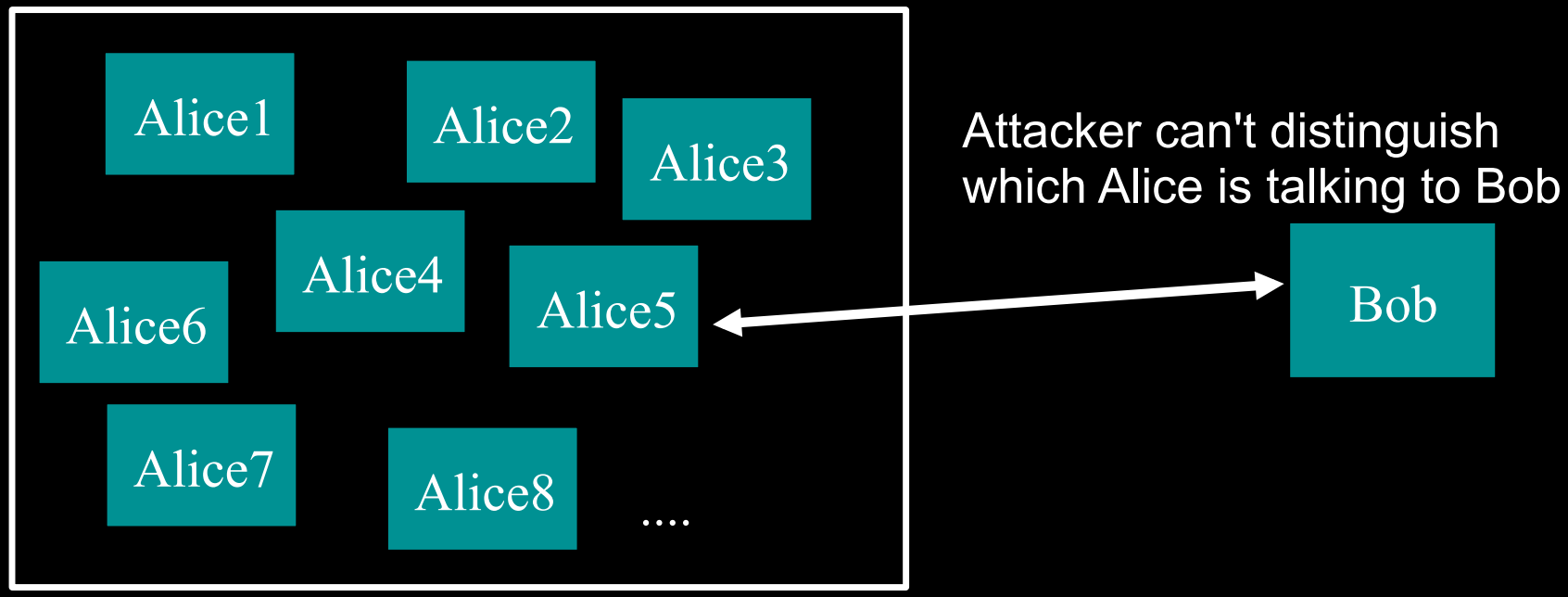
"Who's been viewing my webpages?"

"Who's been emailing patent attorneys?"

Formally: anonymity means indistinguishability within an “anonymity set”

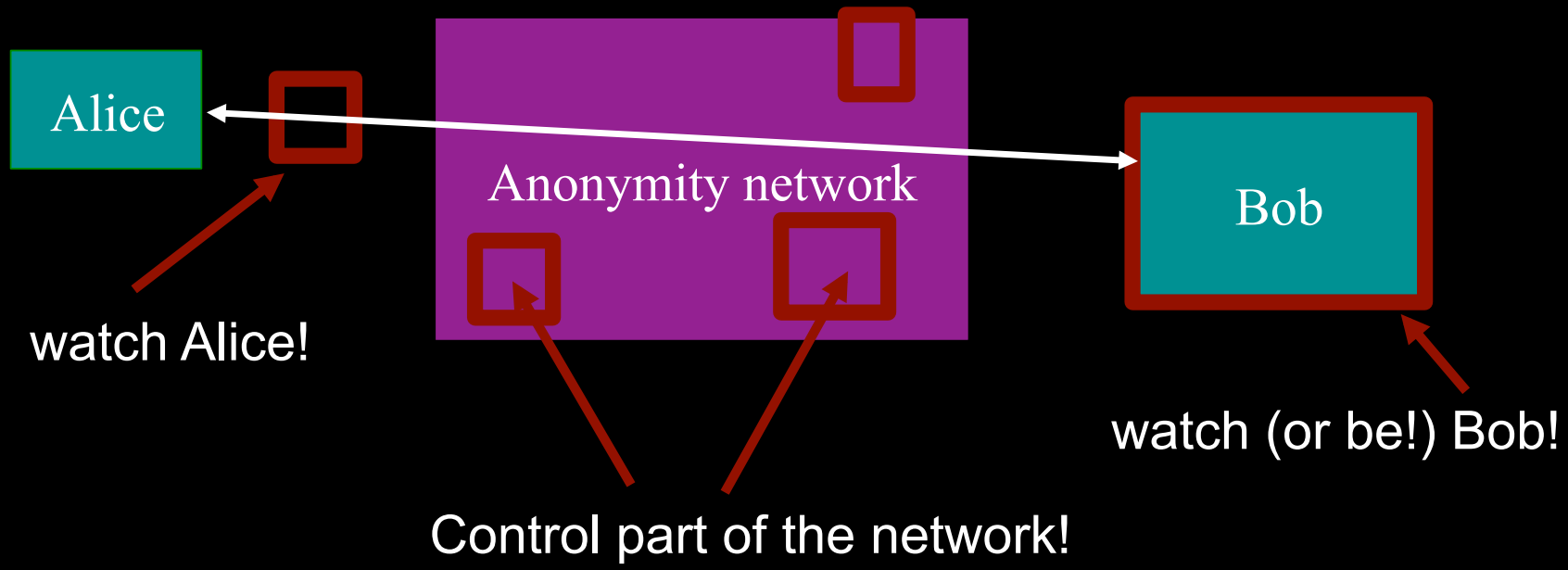


Formally: anonymity means indistinguishability within an “anonymity set”



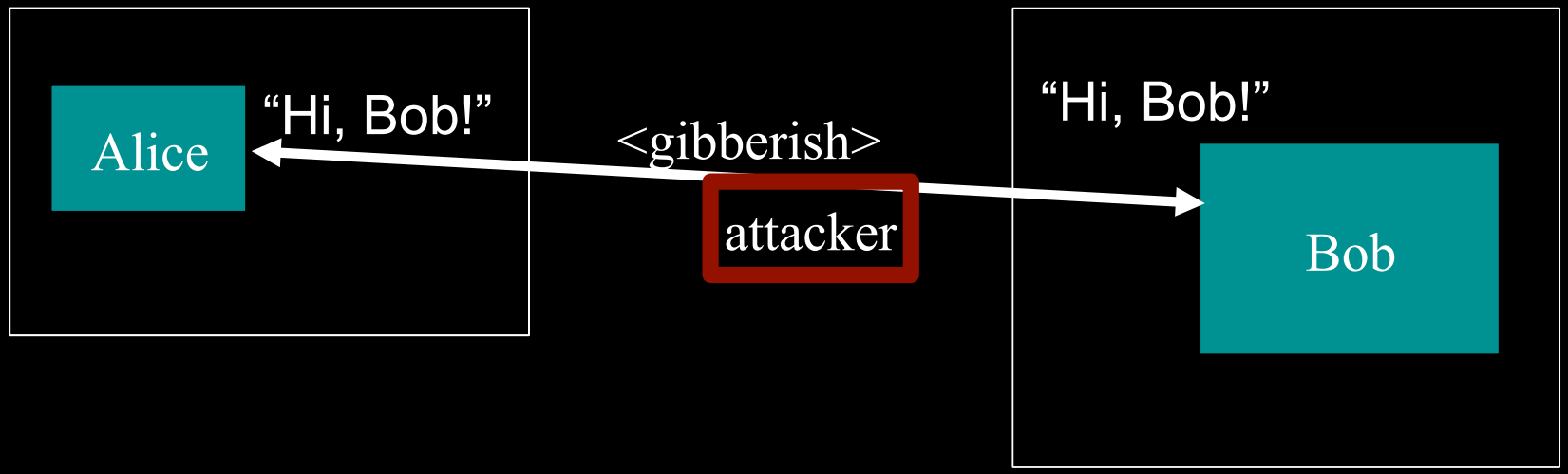
- Can't distinguish?
  - Basic anonymity set size
  - Probability distribution within anonymity set
  - ....

We have to make some assumptions about what the attacker can do.

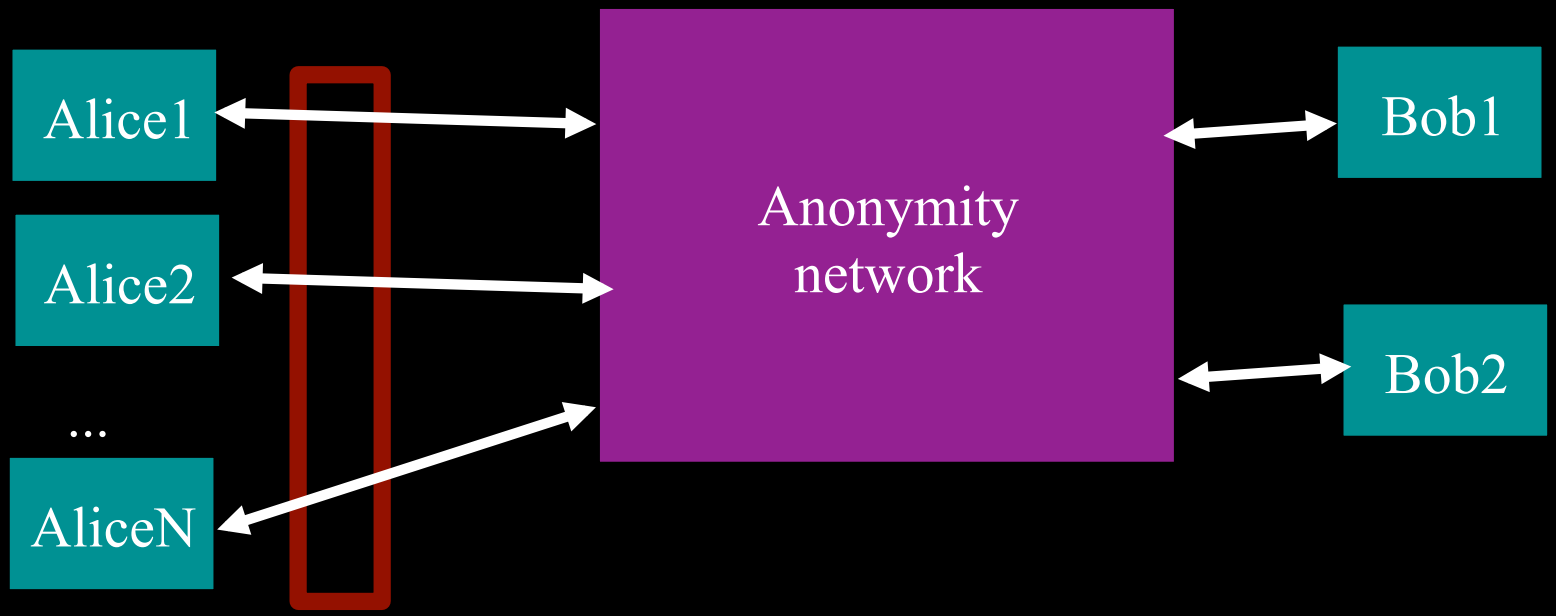


Etc, etc.

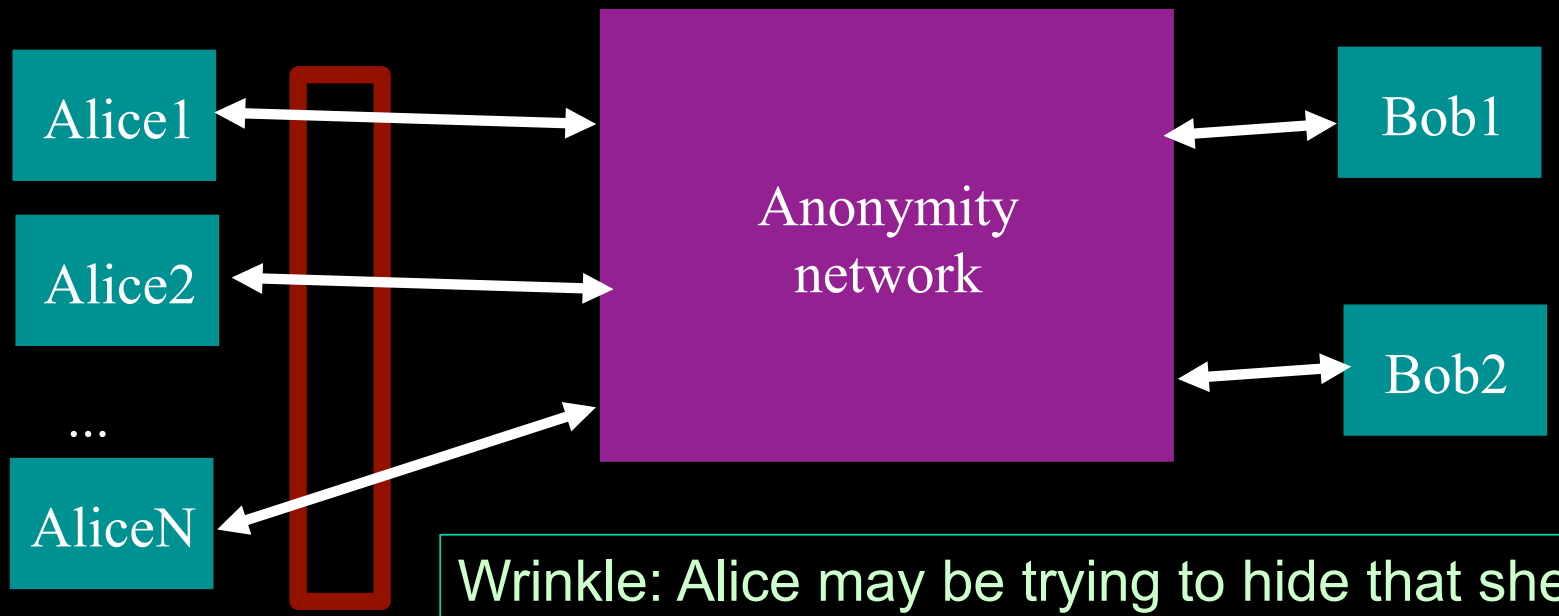
Anonymity isn't confidentiality: Encryption just protects contents.



Anonymity isn't steganography: Attacker can tell that Alice is talking; just not to whom.



Anonymity isn't steganography: Attacker can tell that Alice is talking; just not to whom.



Wrinkle: Alice may be trying to hide that she is talking to the anonymity network.

Anonymity isn't just wishful thinking

“You can't prove it was me!”

“Promise you won't look!”
“Promise you won't remember!”
“Promise you won't tell!”

“I didn't write my name on it!”

“Isn't the Internet already anonymous?”

Anonymity isn't just wishful thinking

"You can't prove it was me!"

Often statistical likelyhood matters more than legal proof.

Will others have incentives & ability to keep promises? Our goal is technical protections without reliance on policy promises.

"Promise you won't look!"

"Promise you won't remember!"

"Promise you won't tell!"

"I didn't write my name on it!"

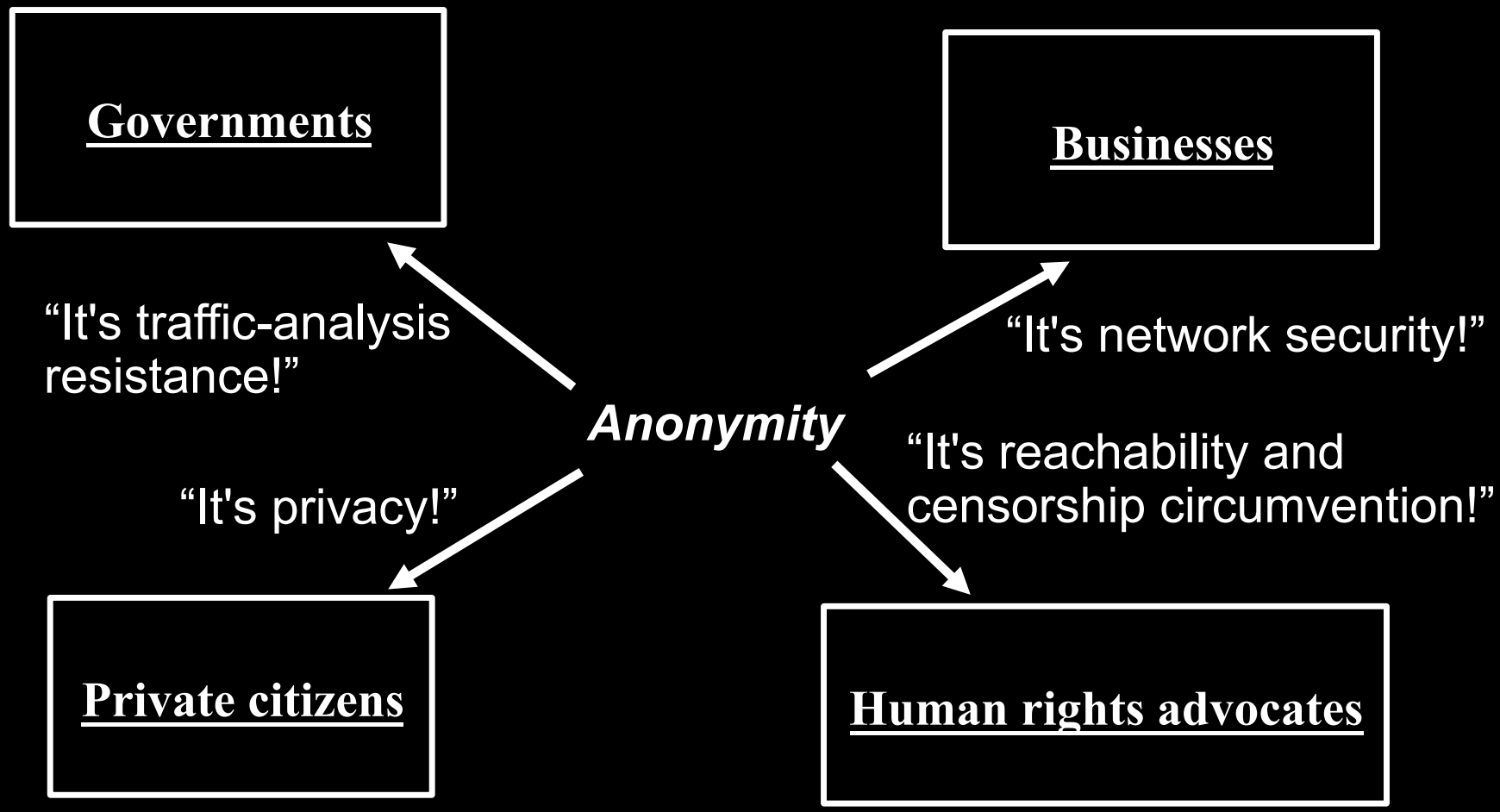
Not what we're talking about.

No!

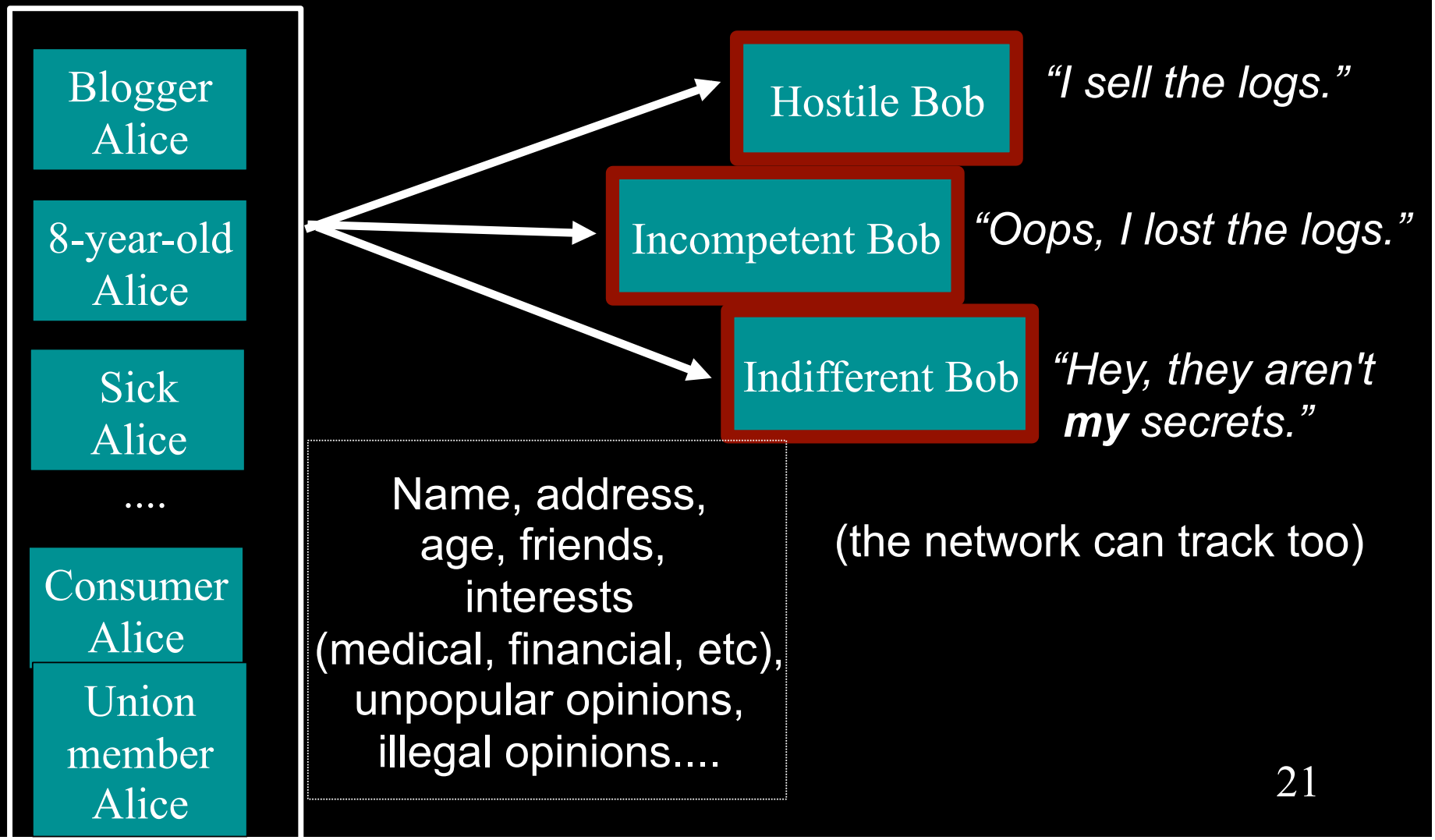
"Isn't the Internet already anonymous?"

2. **Why does anonymity matter?**

Anonymity serves different interests for different user groups.

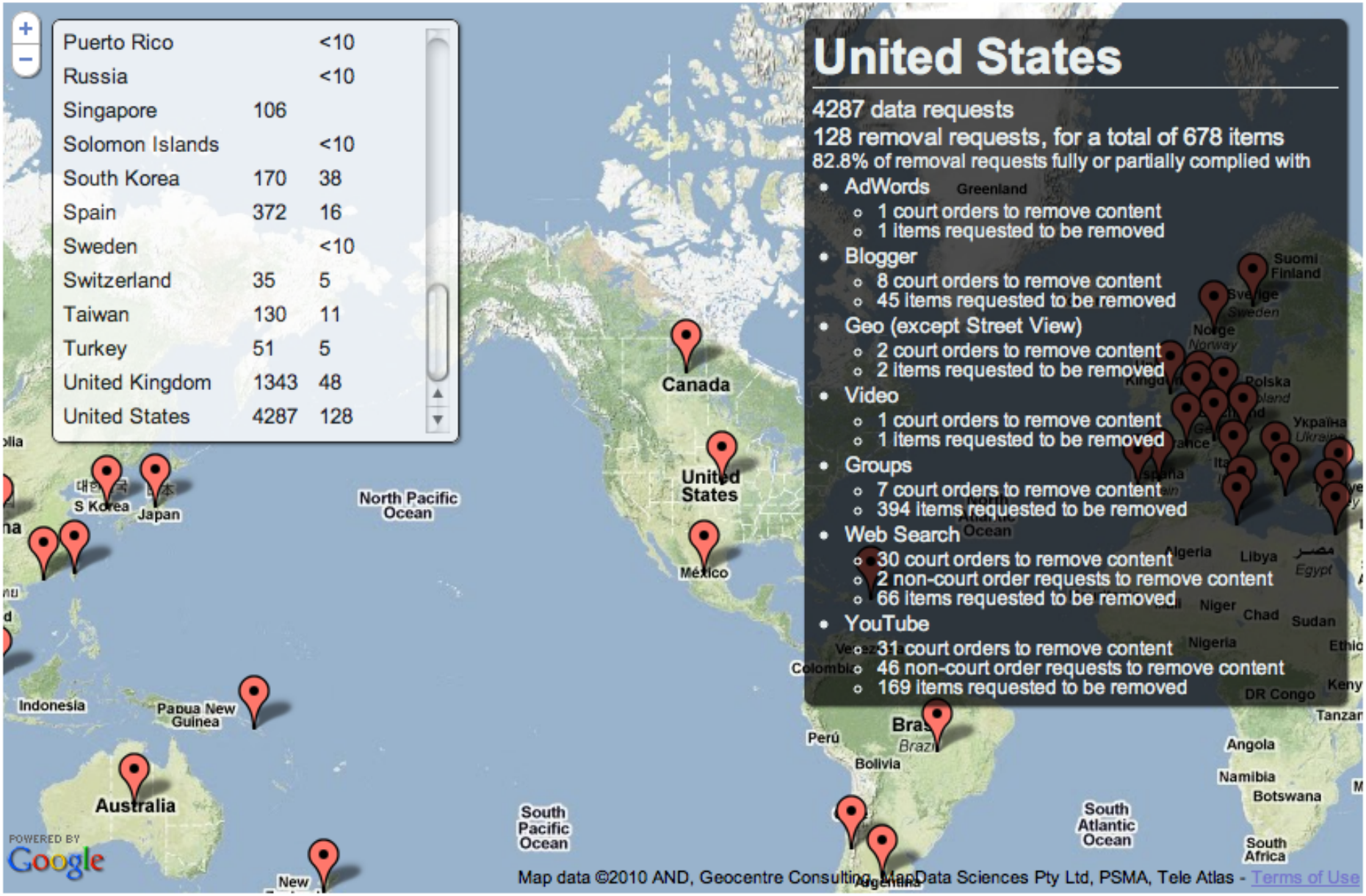


Regular citizens don't want to be watched and tracked.



Many people don't get to see the internet that you can see...

January 2010 to June 2010




January 2010 to June 2010
Austria 2 <10
Belgium 71 <10
Brazil 2435 398
Canada <10
Chile 115
China ?
Cyprus <10
France 1017 25
Germany 668 124
Greece <10
Hong Kong 50 <10
India 1430 30
Brazil
2435 data requests
398 removal requests, for a total of 19806 items
67.6% of removal requests fully or partially complied with
Blogger
15 court orders to remove content
98 items requested to be removed
Video
1 court orders to remove content
1 items requested to be removed
Gmail
1 court orders to remove content
1 items requested to be removed
Geo (except Street View)
1 court orders to remove content
1 items requested to be removed
Picasa
1 court orders to remove content
18318 items requested to be removed
orkut
99 court orders to remove content
220 non-court order requests to remove content
1045 items requested to be removed
Web Search
13 court orders to remove content
275 items requested to be removed
YouTube
46 court orders to remove content
1 non-court order requests to remove content
67 items requested to be removed
Canada
United States
México
North Pacific Ocean
North Atlantic Ocean
South Pacific Ocean
South Atlantic Ocean
Venezuela
Colombia
Perú
Bolivia
Argentina
Finland
Sverige
Norway
Kazakhstan
Iran
Afghanistan
Pakistan
Algeria
Libya
Egypt
Saudi Arabia
Nigeria
Ethiopia
Tanzania
Botswana
South Africa
New Zealand
Map data ©2010 AND, Geocentre Consulting, MapData Sciences Pty Ltd, PSMA, Tele Atlas - Terms of Use


January 2010 to June 2010
China
Chinese officials consider censorship demands to be state secrets, so we cannot disclose that information at this time.
Austria 2
Belgium 71 <10
Brazil 2435 398
Canada <10
Chile 115
China ?
Cyprus <10
France 1017 25
Germany 668 124
Greece <10
Hong Kong 50 <10
India 1430 30
Suomi Finland
Sverige Sweden
Norge Norway
Iceland
Россия Russia
Polska
Україна Ukraine
Kazakhstan
Mongolia
China
S Korea
Japan
Iraq
Iran
Afghanistan
Pakistan
India
Saudi Arabia
Algeria
Libya
مصر Egypt
Mauritania
Mali
Niger
Chad
Sudan
Nigeria
Ethiopia
Kenya
DR Congo
Tanzania
Angola
Namibia
Botswana
Madagascar
South Africa
South Atlantic Ocean
Indian Ocean
Thailand
Indonesia
Papua New Guinea
Australia
México
Venezuela
Colombia
Perú
Bolivia
Argentina
United States
North Atlantic Ocean
Greenland
POWERED BY Google
Map data ©2010 AND, Geocentre Consulting, MapData Sciences Pty Ltd, PSMA, Tele Atlas - Terms of Use

Google Transparency Report: Traffic

Home

Government Requests

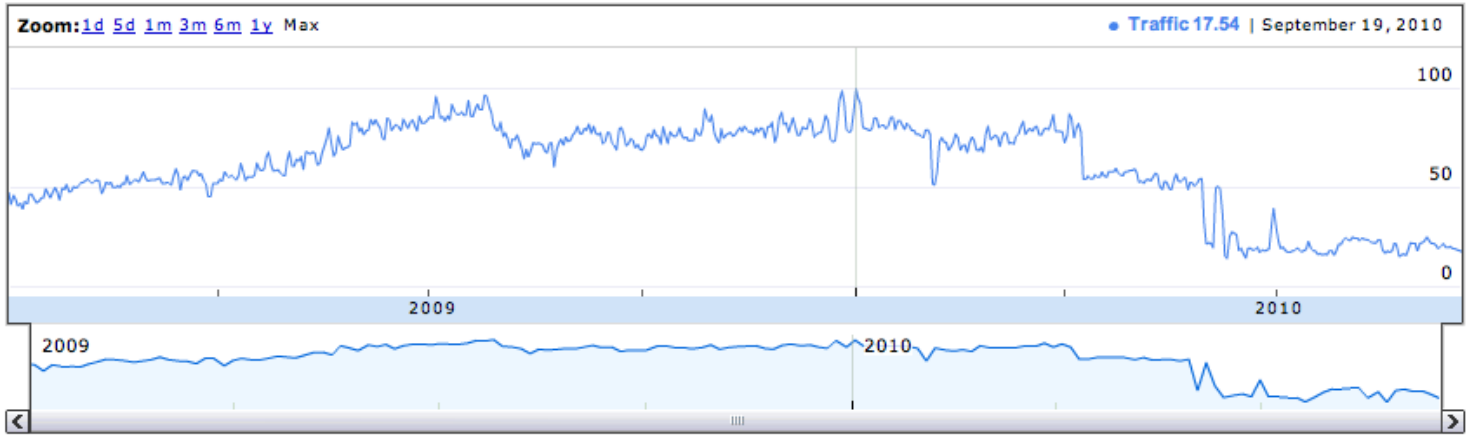
Traffic

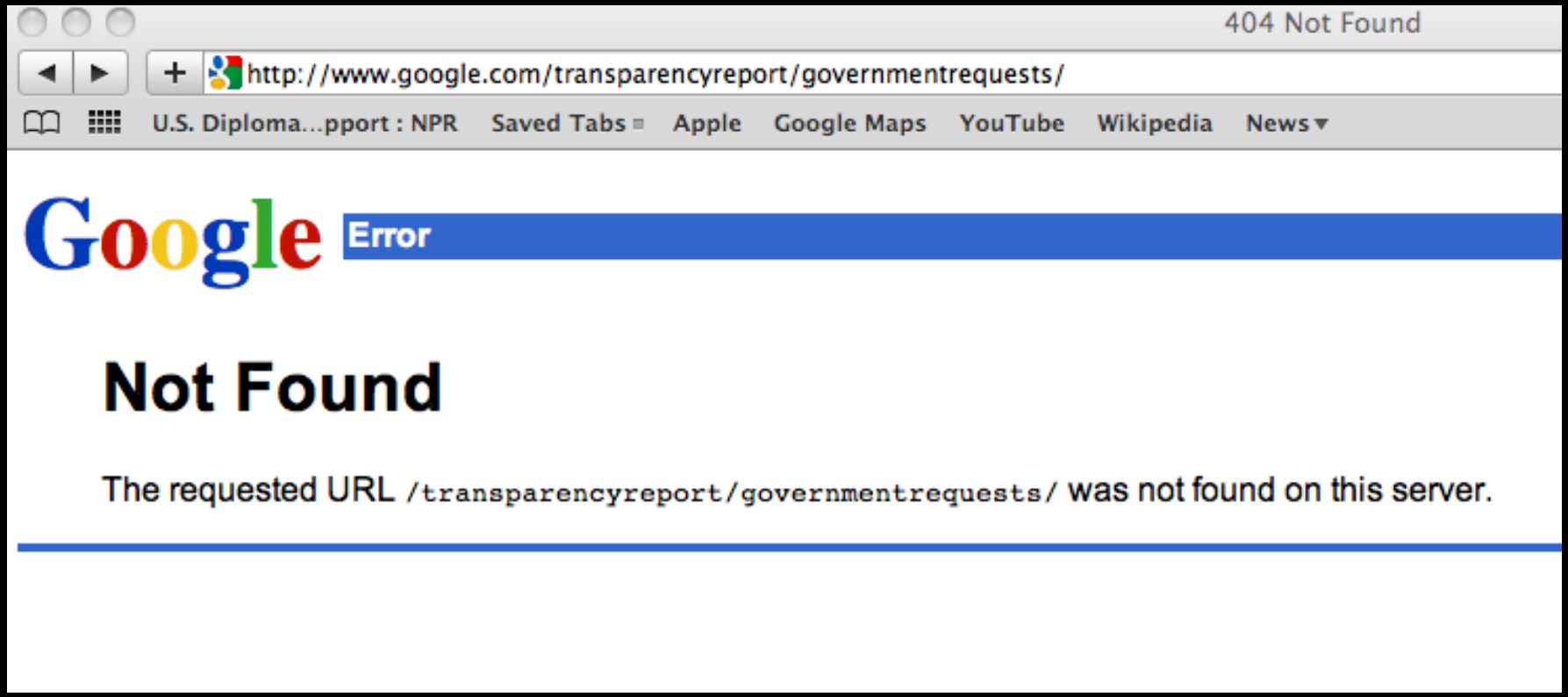
FAQ

This tool provides information about traffic to our services around the world. Each graph shows historic traffic patterns for a given country/region and service. You may select a country/region and then choose a service to view each respective graph. Graphs are updated as data are collected, normalized, and scaled in units of 0 to 100. By showing outages, this tool visualizes disruptions in the free flow of information, whether it's a government blocking information or a cable being cut.

Turkey | YouTube

Turkey YouTube Traffic

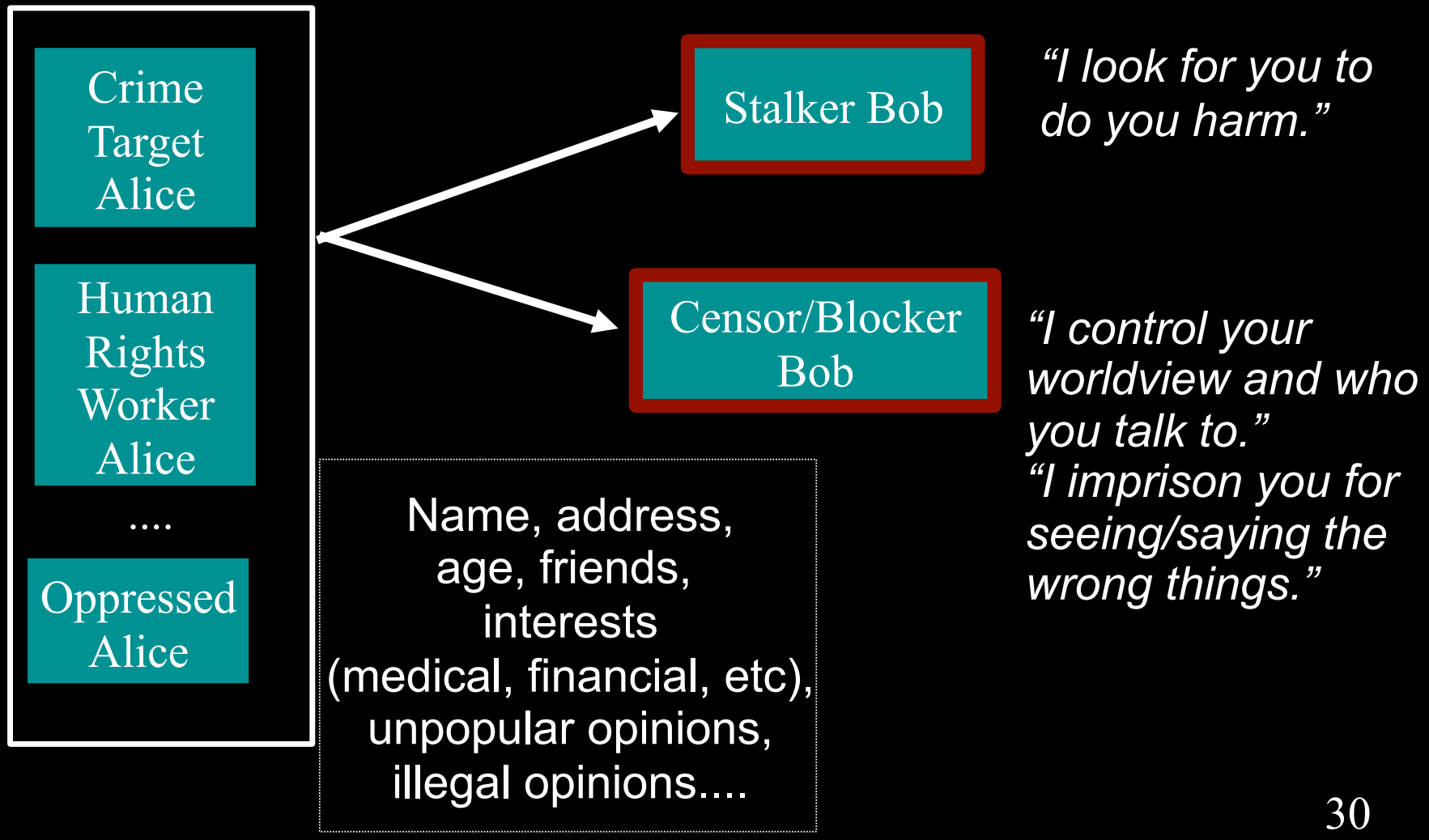



404 Not Found
http://www.google.com/transparencyreport/governmentrequests/
U.S. Diploma...pport : NPR
Saved Tabs
Apple
Google Maps
YouTube
Wikipedia
News
Google
Error
Not Found
The requested URL /transparencyreport/governmentrequests/ was not found on this server.

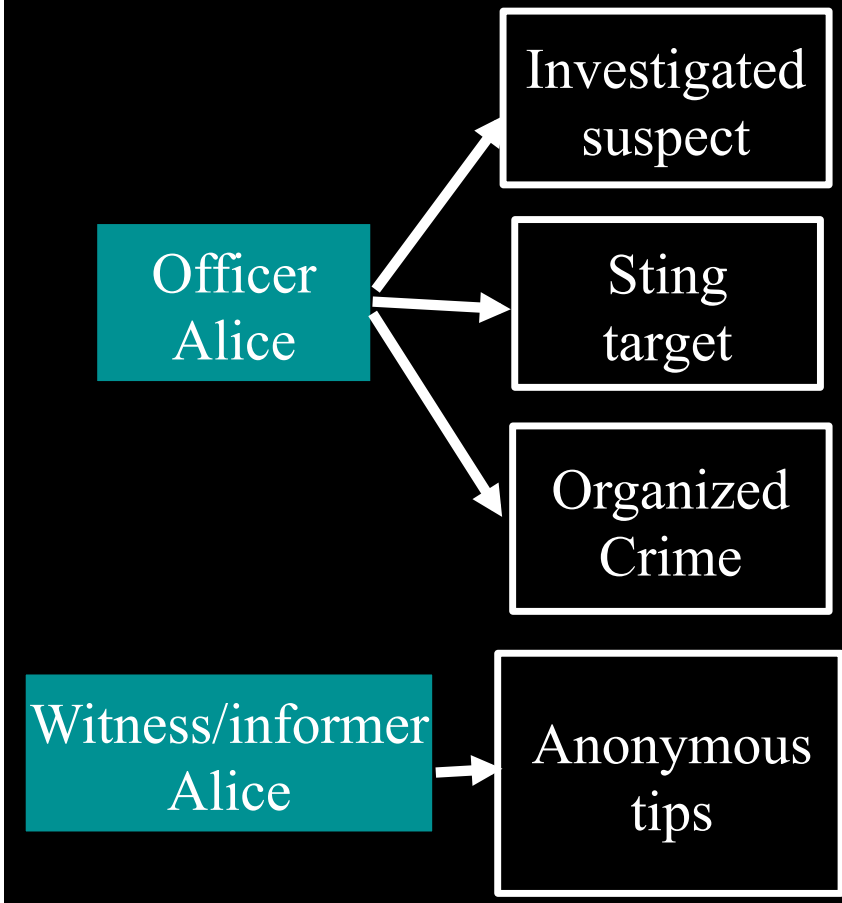
and they can't speak on the internet either...

It's not only about dissidents in faraway lands

Regular citizens don't want to be watched and tracked.



Law enforcement needs anonymity to get the job done.



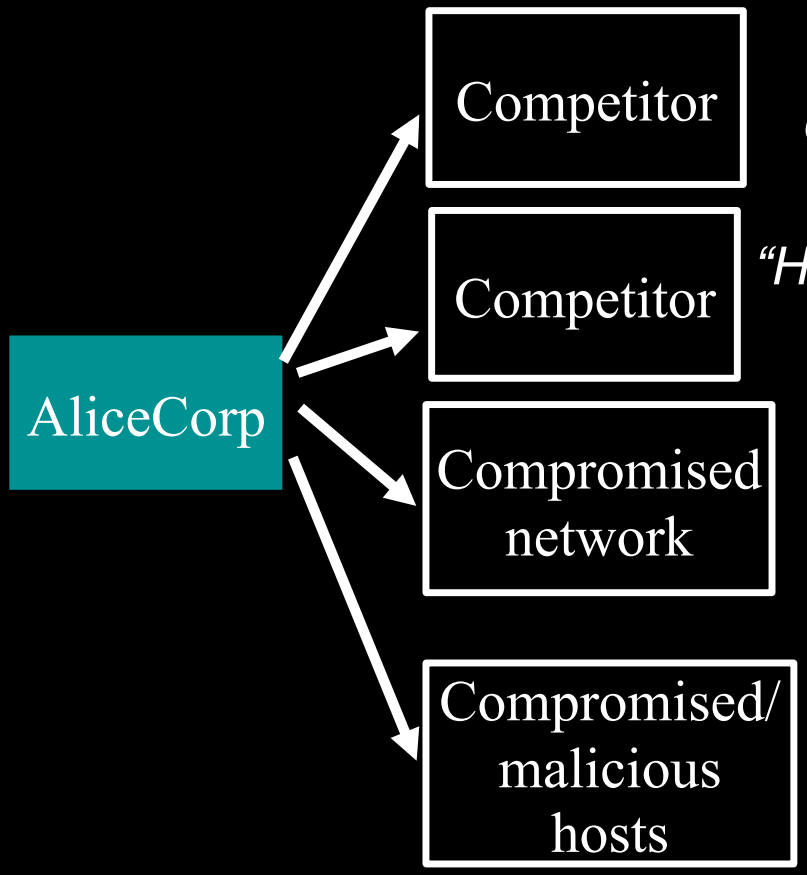
"Why is alice.fbi.gov reading my website?"

"Why no, alice.localpolice.gov! I would never sell counterfeits on ebay!"

"Is my family safe if I go after these guys?"

"Are they really going to ensure my anonymity?"

Businesses need to protect trade secrets... and their customers.



"Oh, your employees are reading our patents/jobs page/product sheets?"

"Hey, it's Alice! Give her the 'Alice' version!"

"Wanna buy a list of Alice's suppliers? What about her customers? What about her engineering department's favorite search terms?"

"We attack Alice's customers with malware, and watch for when she notices us."

Governments need anonymity for their security

"What will you bid for a list of Baghdad IP addresses that get email from .gov?"

"What bid for the hotel room from which someone just logged in to foo.navy.mil?"

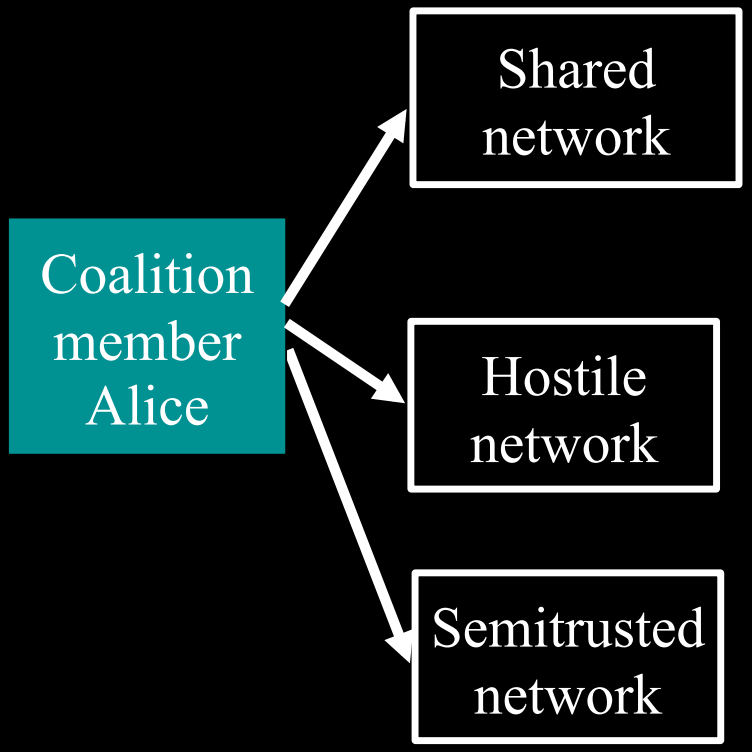
"What *does* the CIA Google for?"

Aside: other benefits of an anonymity system

Besides protecting affiliation, etc. can provide "poor man's VPN". Access to the internet despite

- Network port policy disconnects
- DNS failure

Governments need anonymity for their security



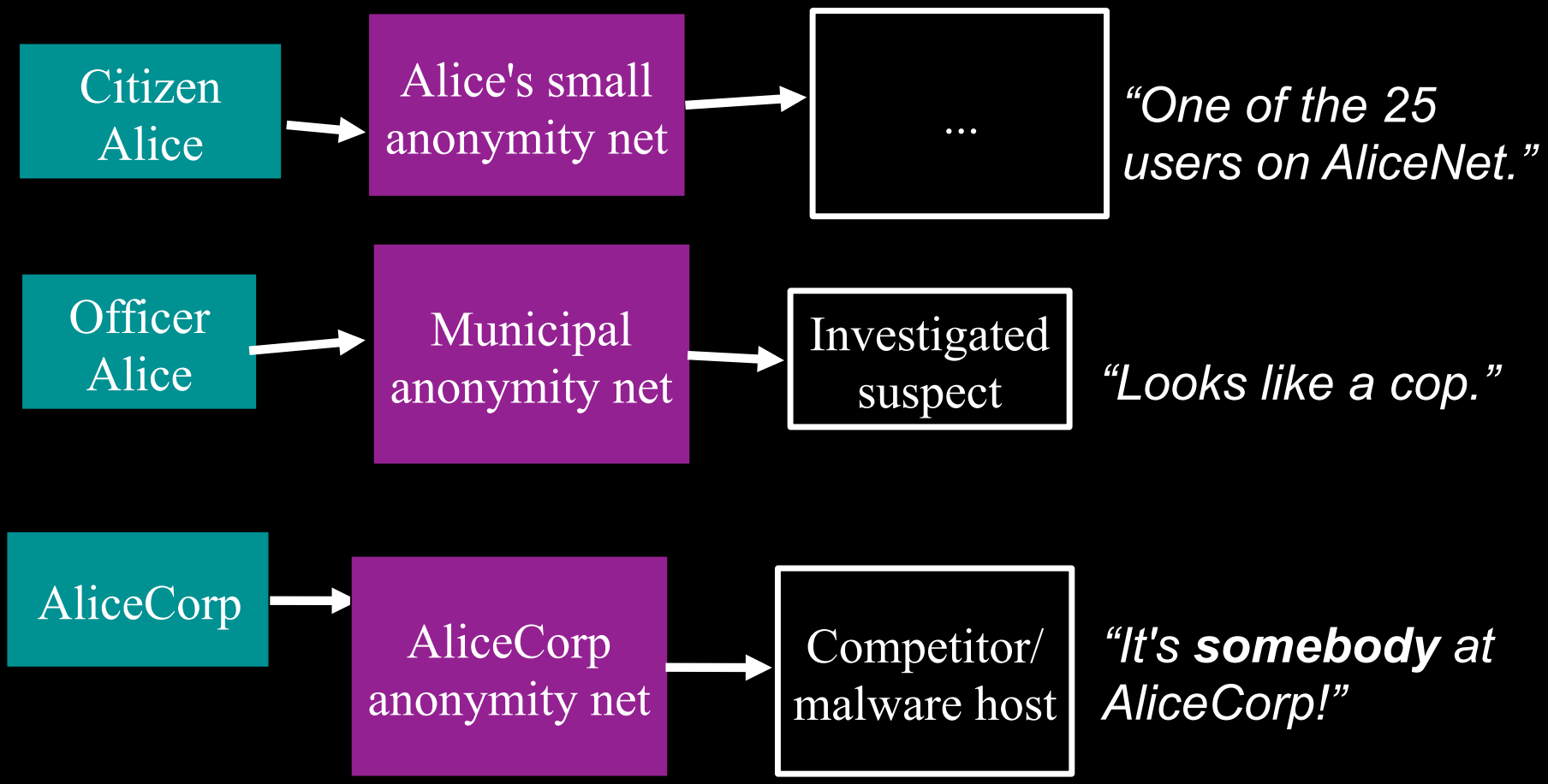
"Do I really want to reveal my internal network topology?"

"Do I want all my partners to know extent/pattern of my comms with other partners?"

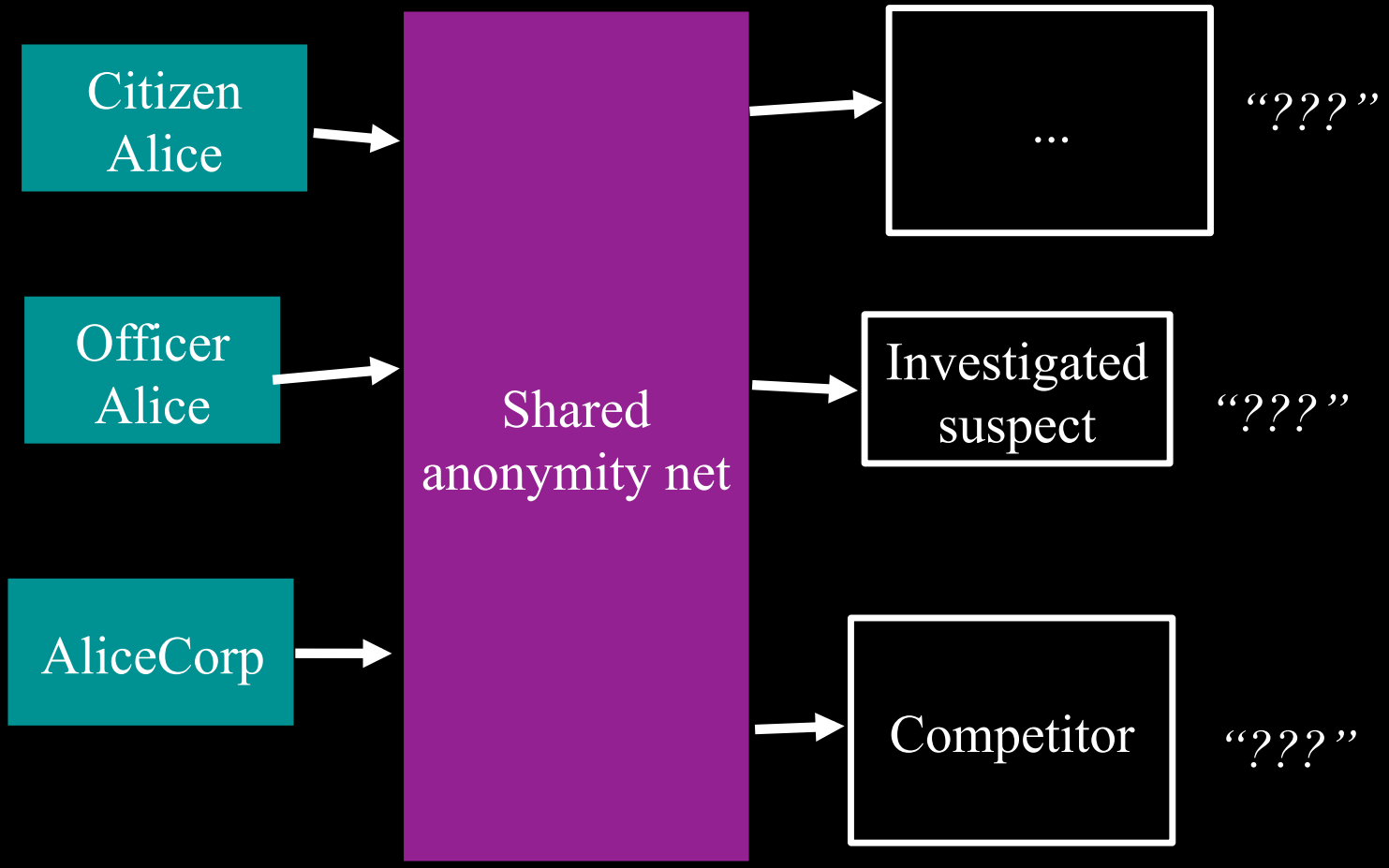
"How can I establish communication with locals without a trusted network?"

"How can I avoid selective blocking of my communications?"

You can't be anonymous by yourself: private solutions are ineffective...

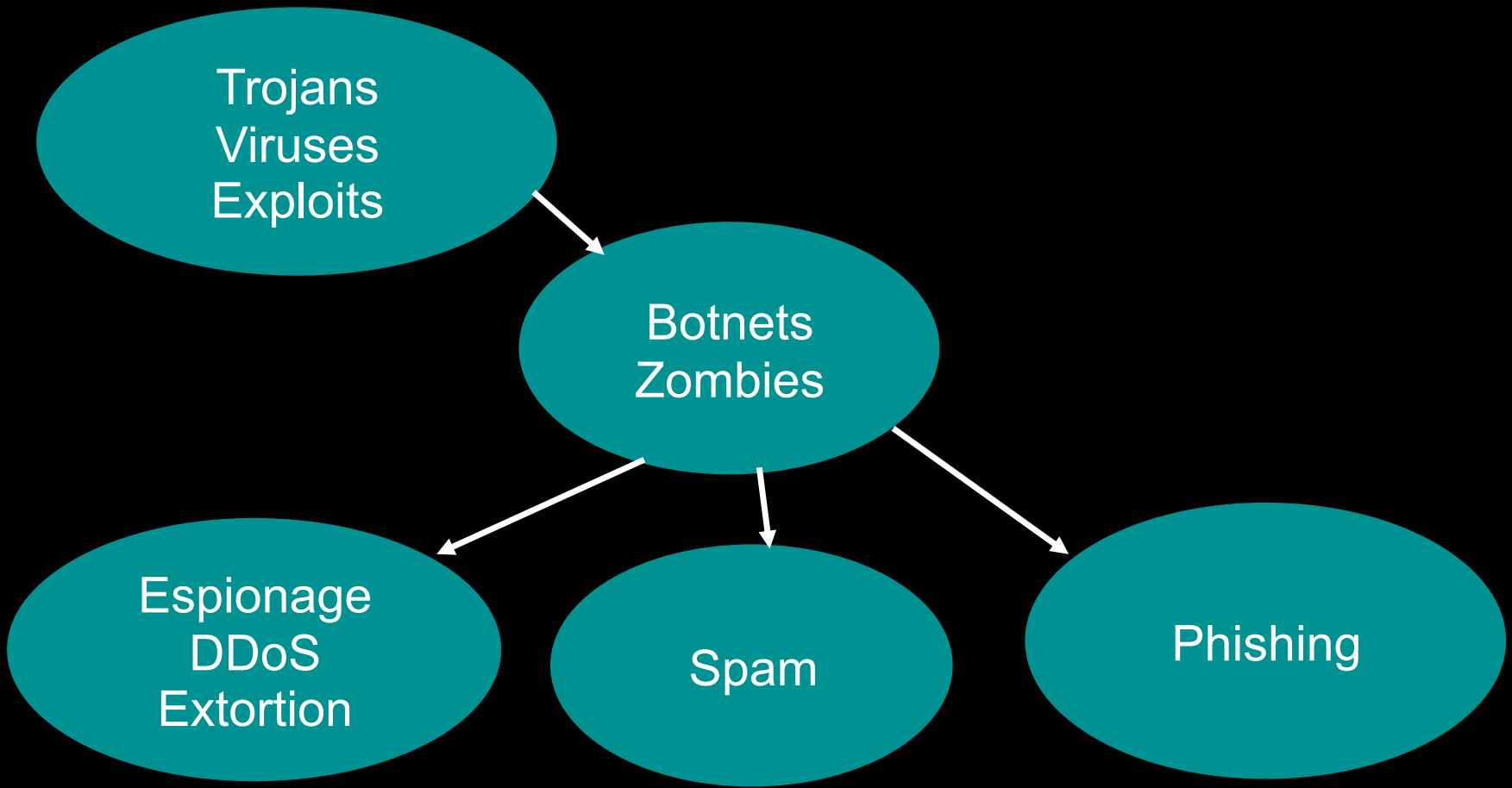


... so, anonymity loves company!

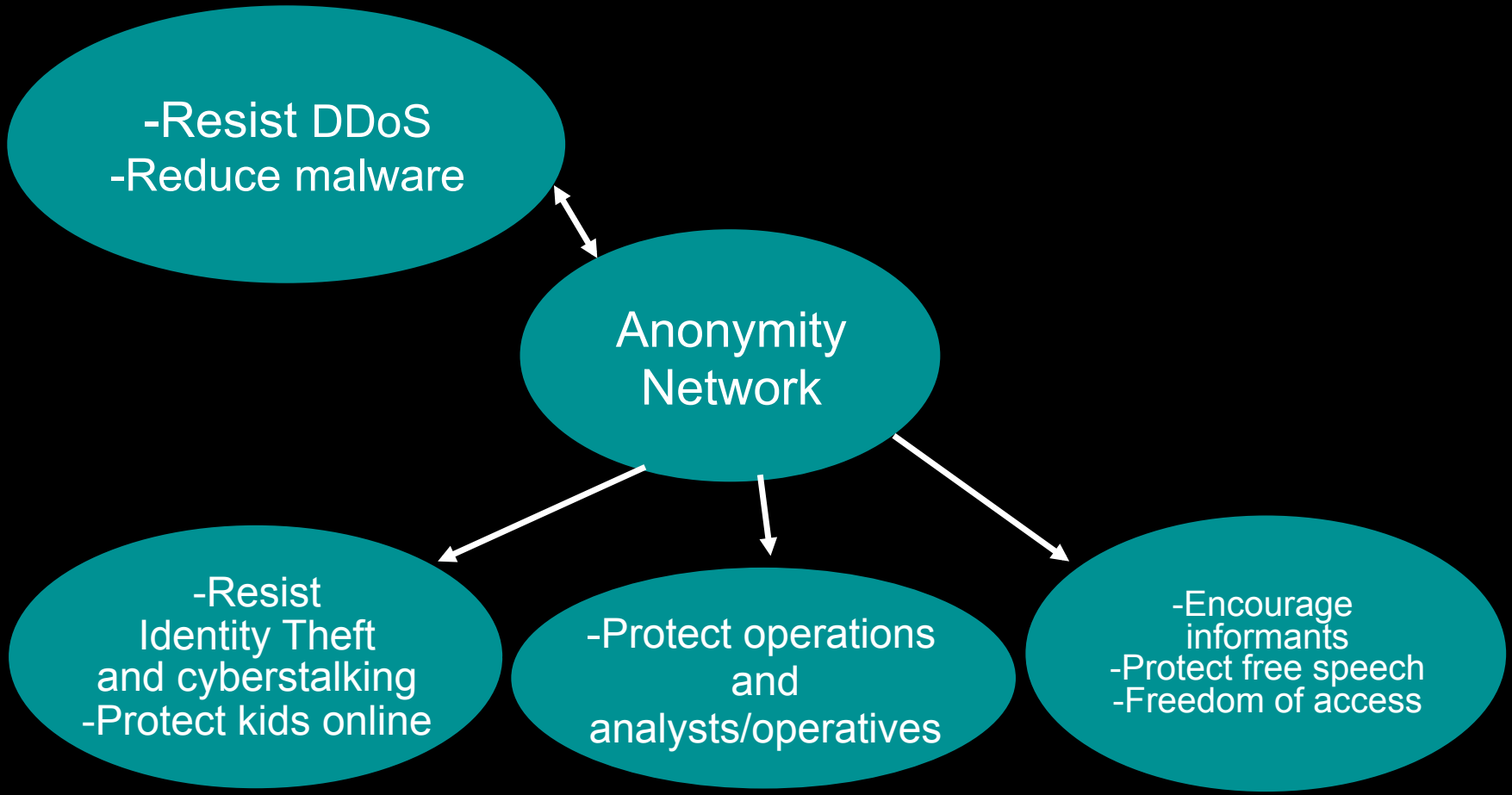


Don't bad people use anonymity?

Current situation:
Bad people on internet are doing fine

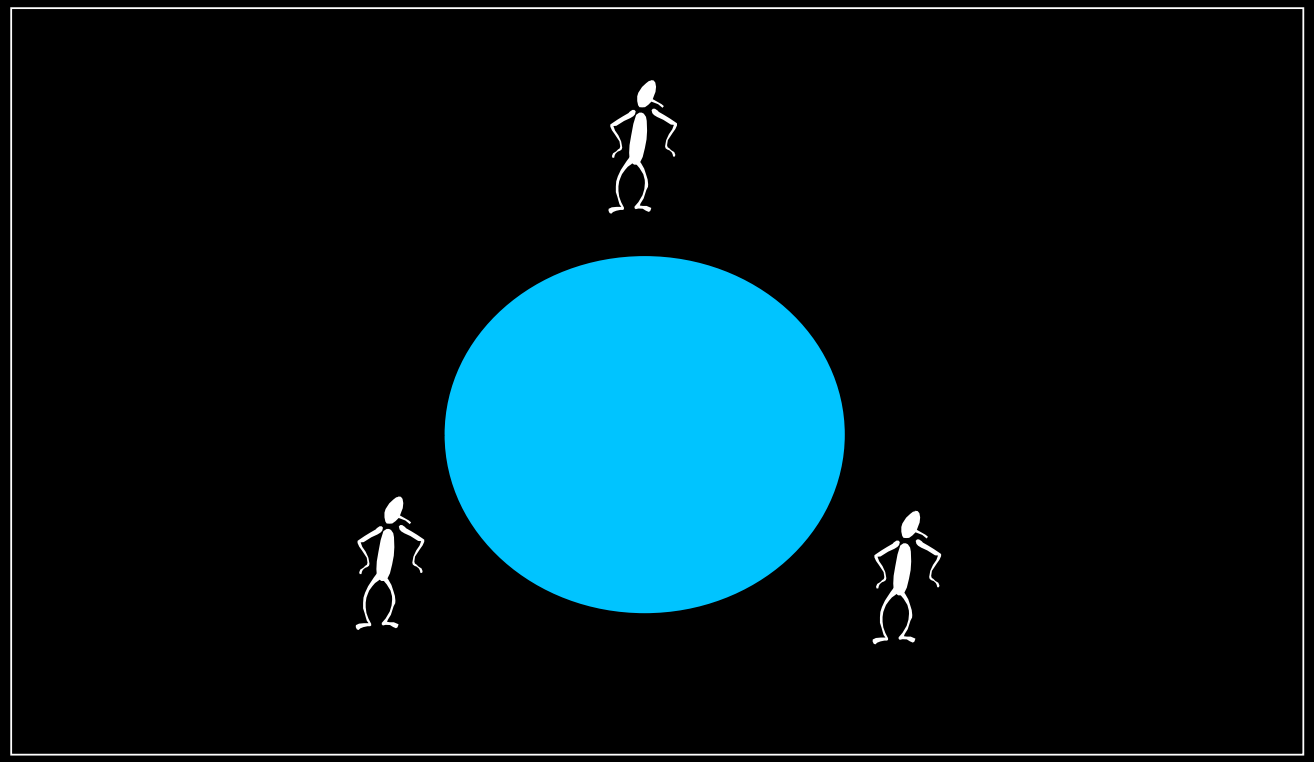


Giving good people a fighting chance

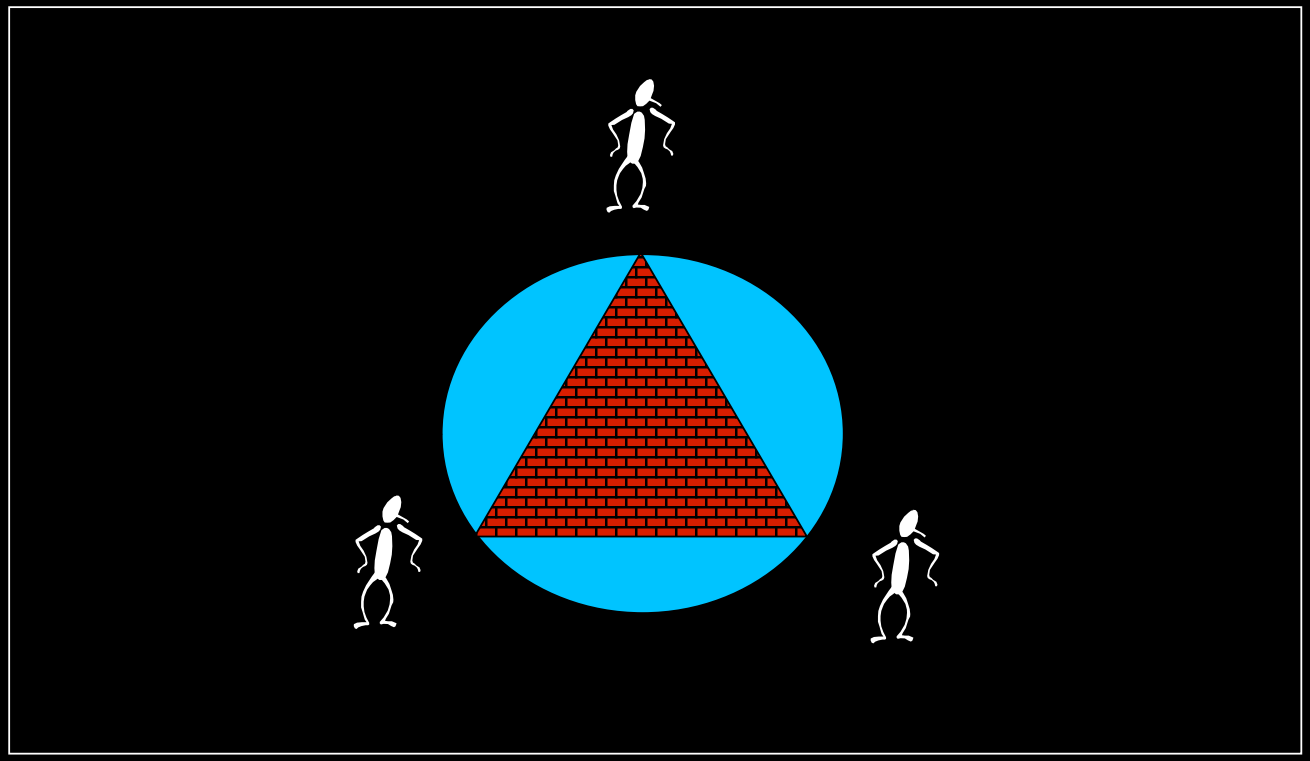


3. How does anonymity work?

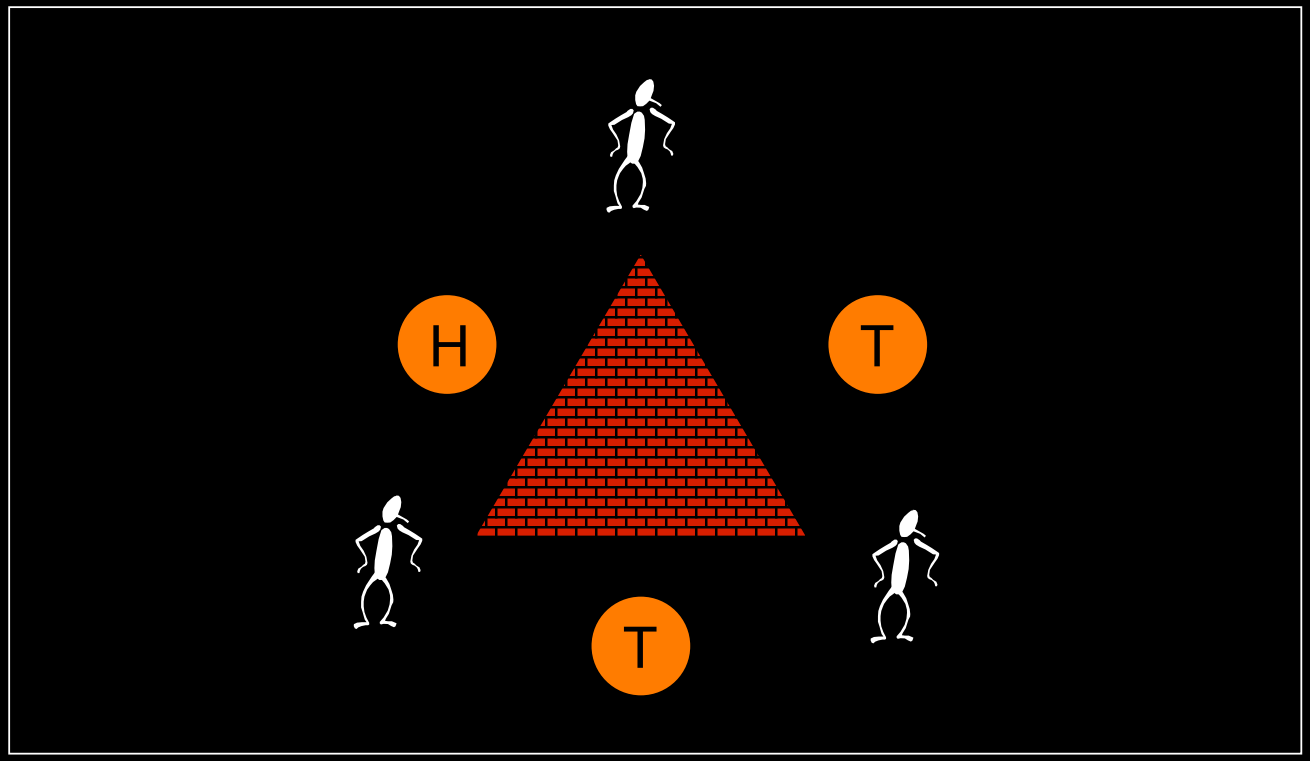
Dining Cryptographers



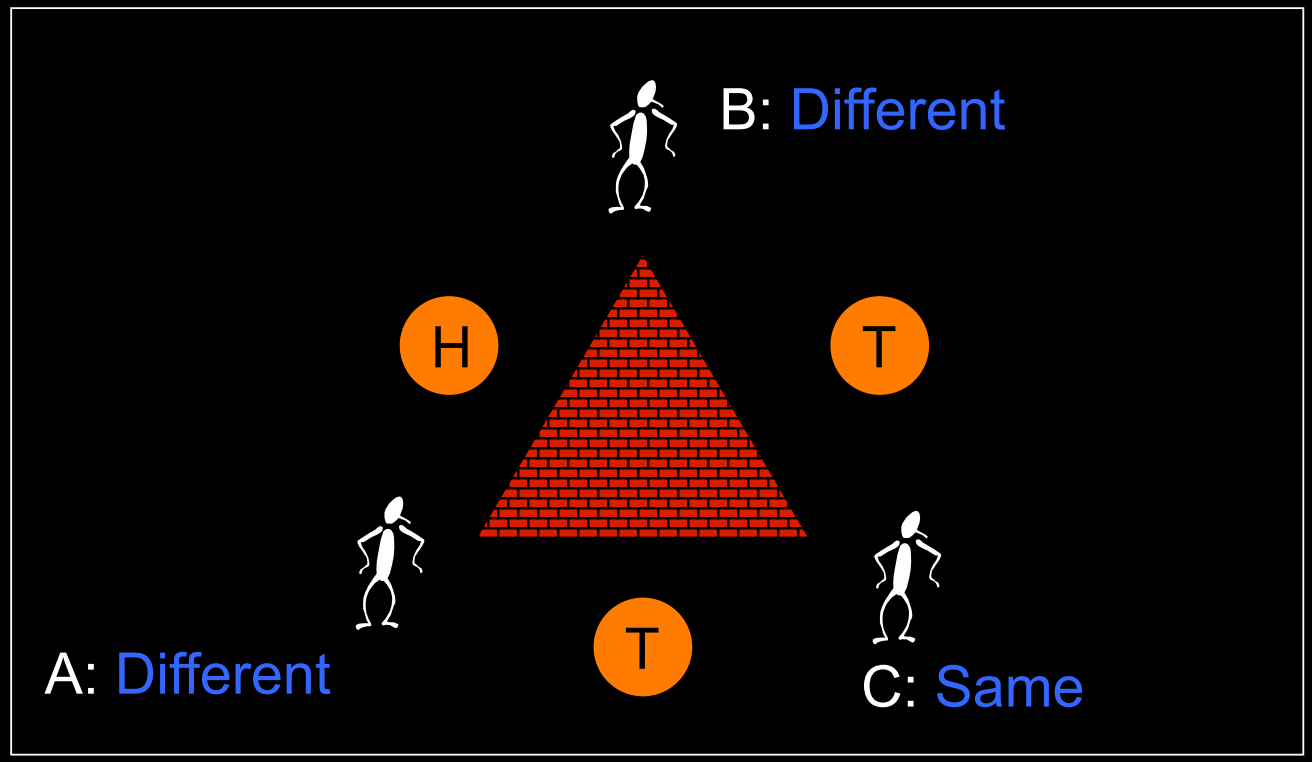
Dining Cryptographers



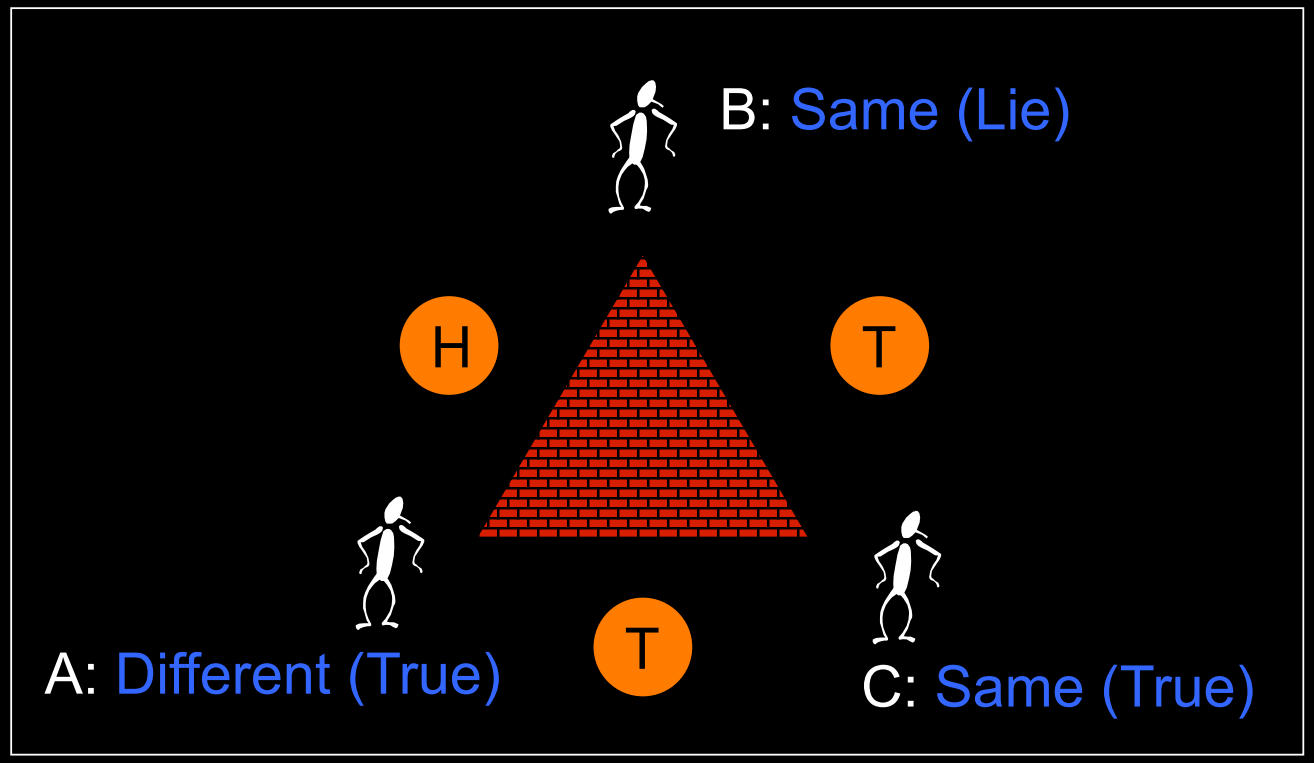
Dining Cryptographers



Dining Cryptographers



Dining Cryptographers

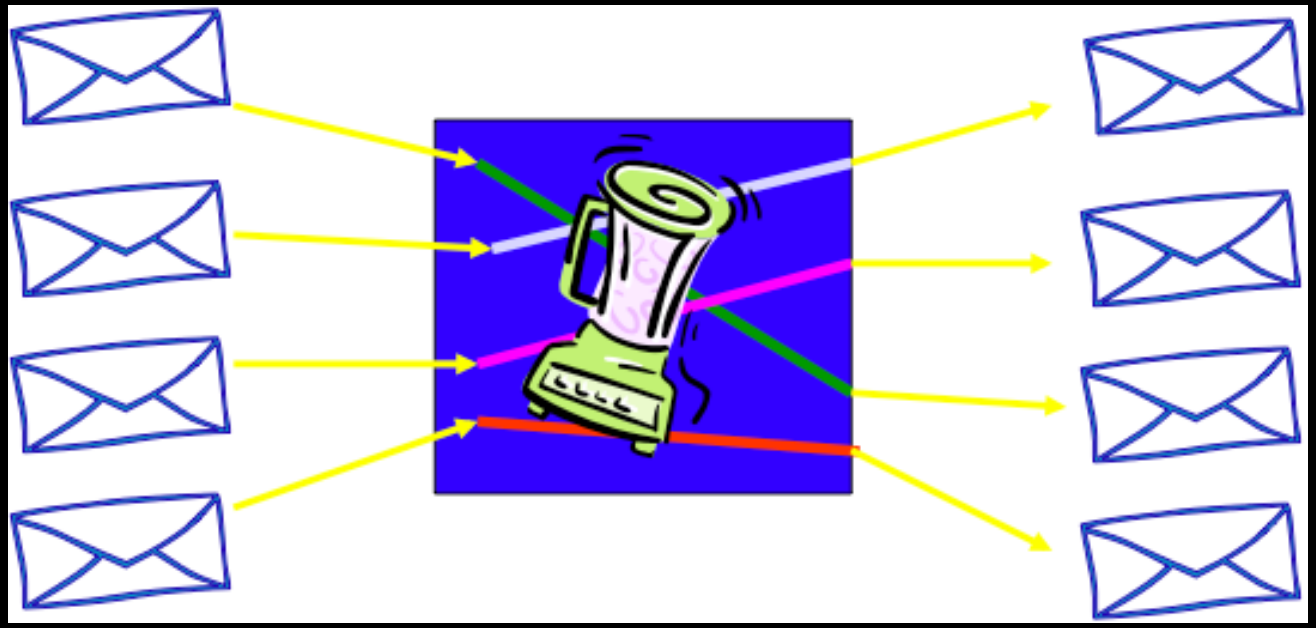


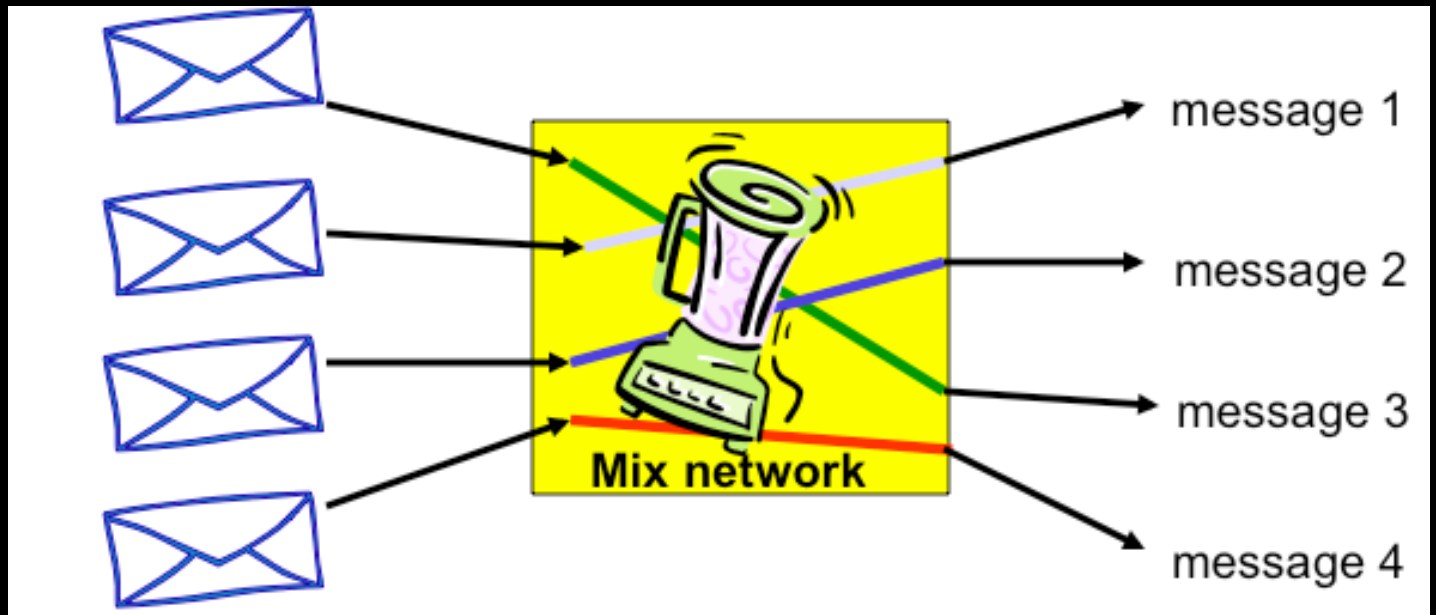
Number of "Different"s odd: Signal 1
Number of "Different"s even: No Signal 0

Dining Cryptographers (DC Nets)

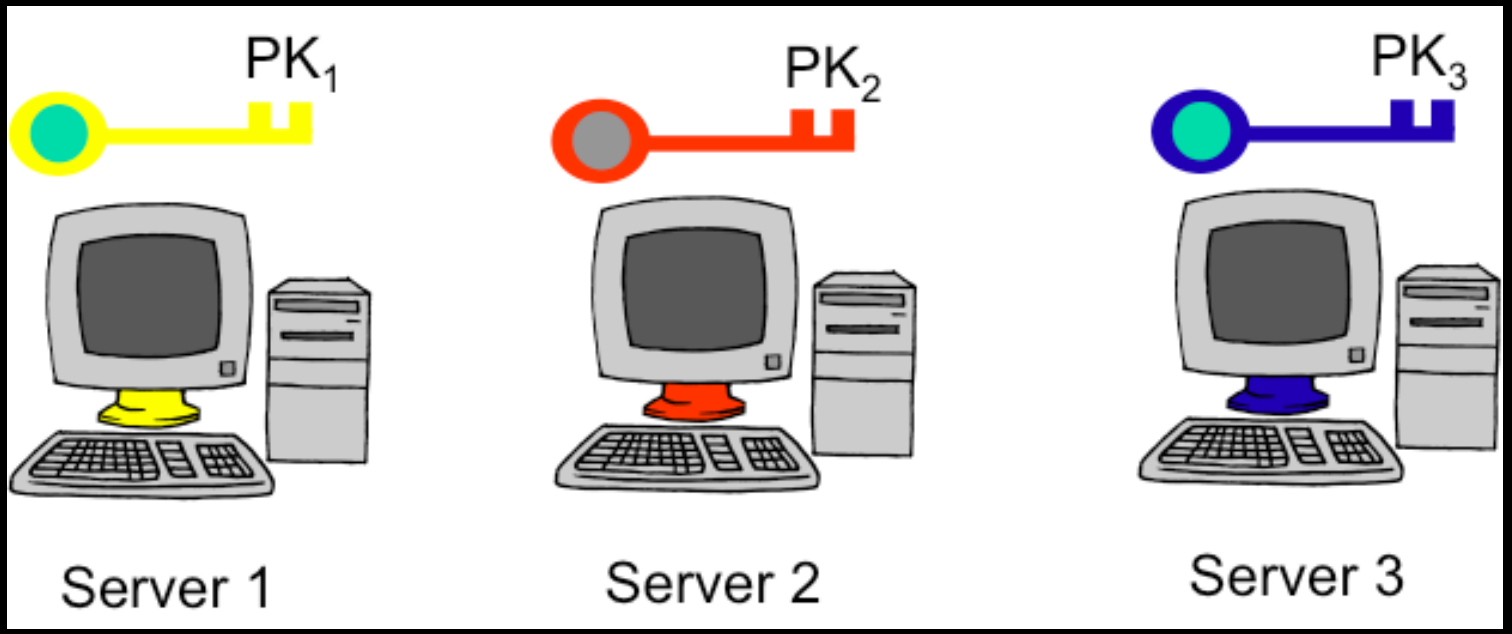
- Invented by Chaum, 1988
- Strong provable properties
- Versions without collision or abuse problems have high communication and computation overhead
- Don't scale very well

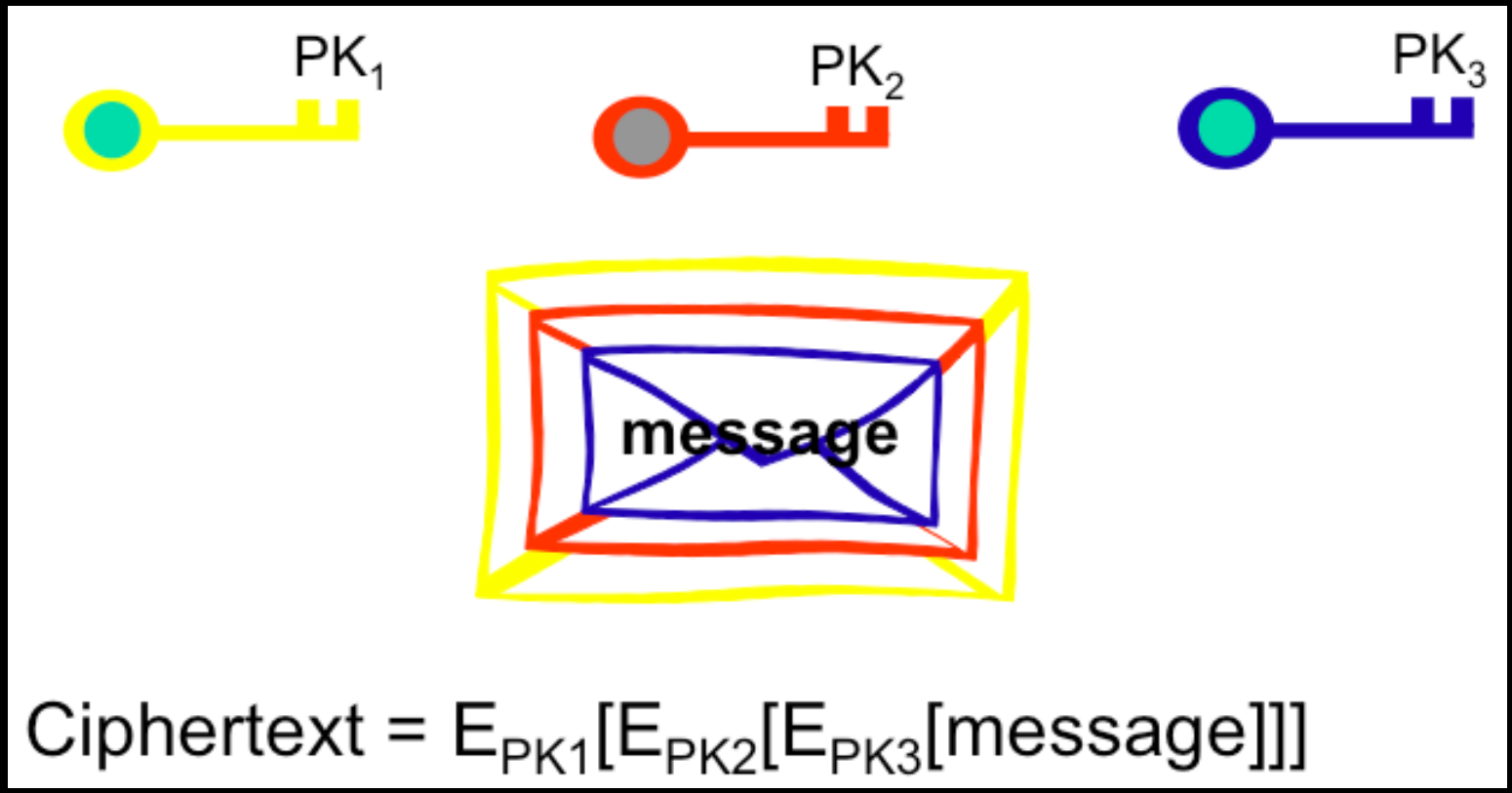
Mixes

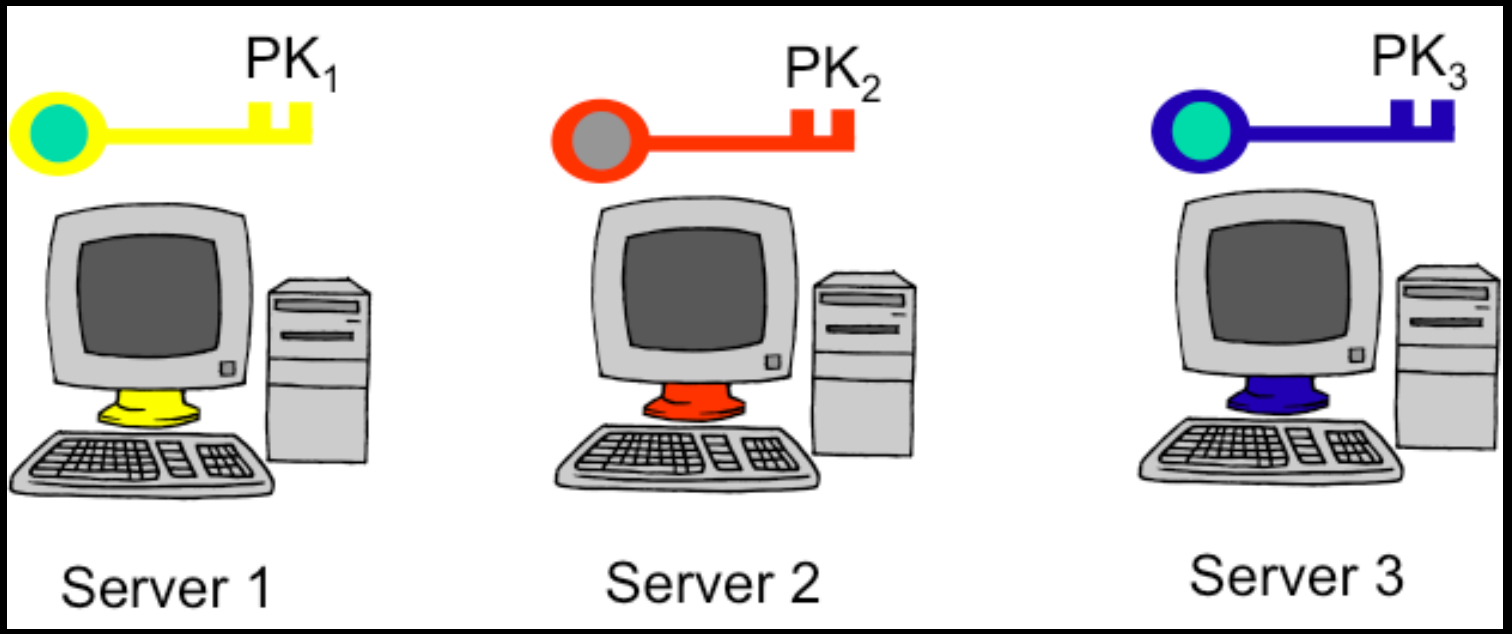


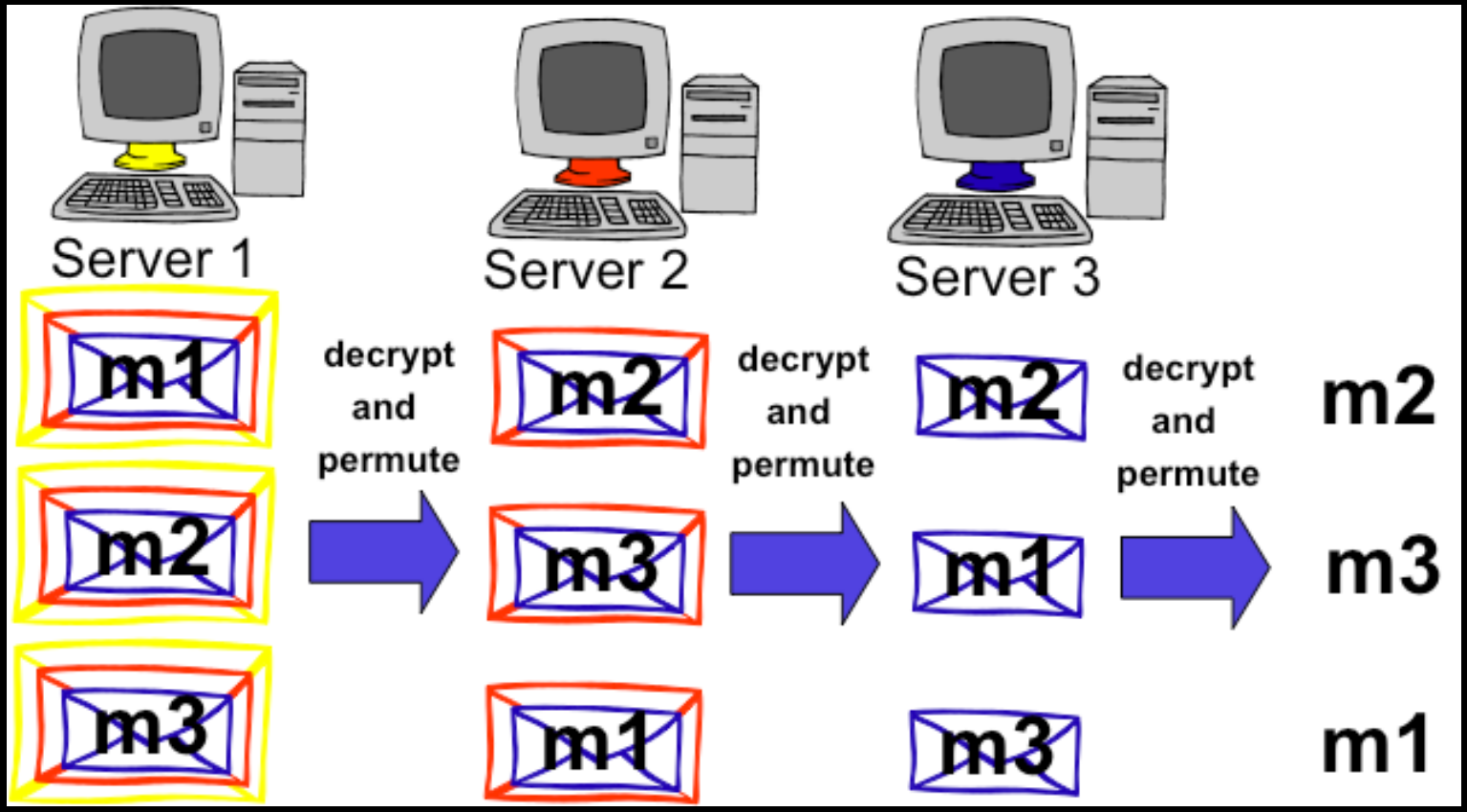


Randomly permutes and decrypts inputs


PK1
PK2
PK3
Server 1
Server 2
Server 3


PK1
PK2
PK3
message
Ciphertext = EPK1[EPK2[EPK3[message]]]

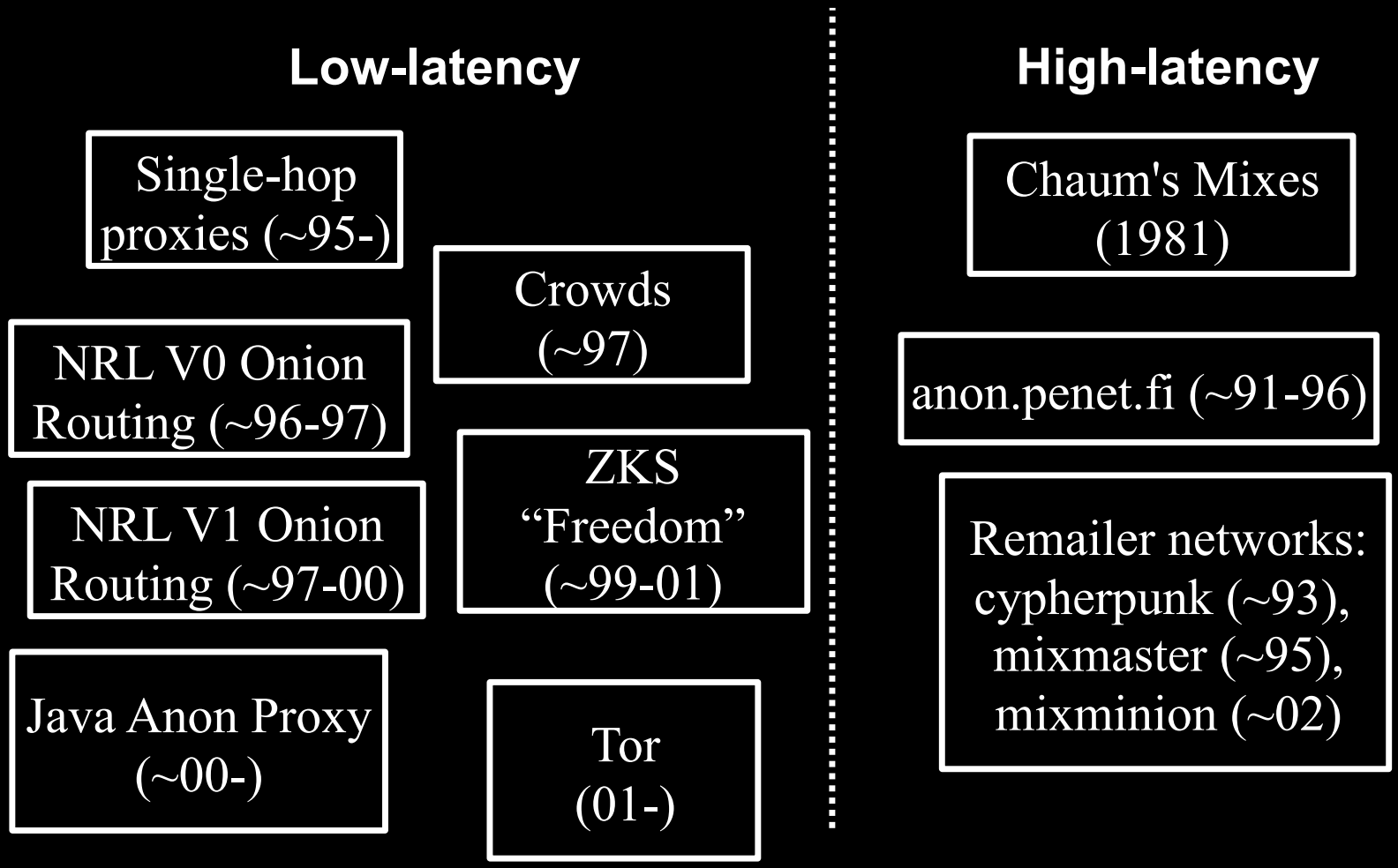

PK1
PK2
PK3
Server 1
Server 2
Server 3


Server 1
Server 2
Server 3
m1
m2
m3
decrypt and permute
decrypt and permute
decrypt and permute
m2
m3
m1
m2
m1
m3
m2
m3
m1

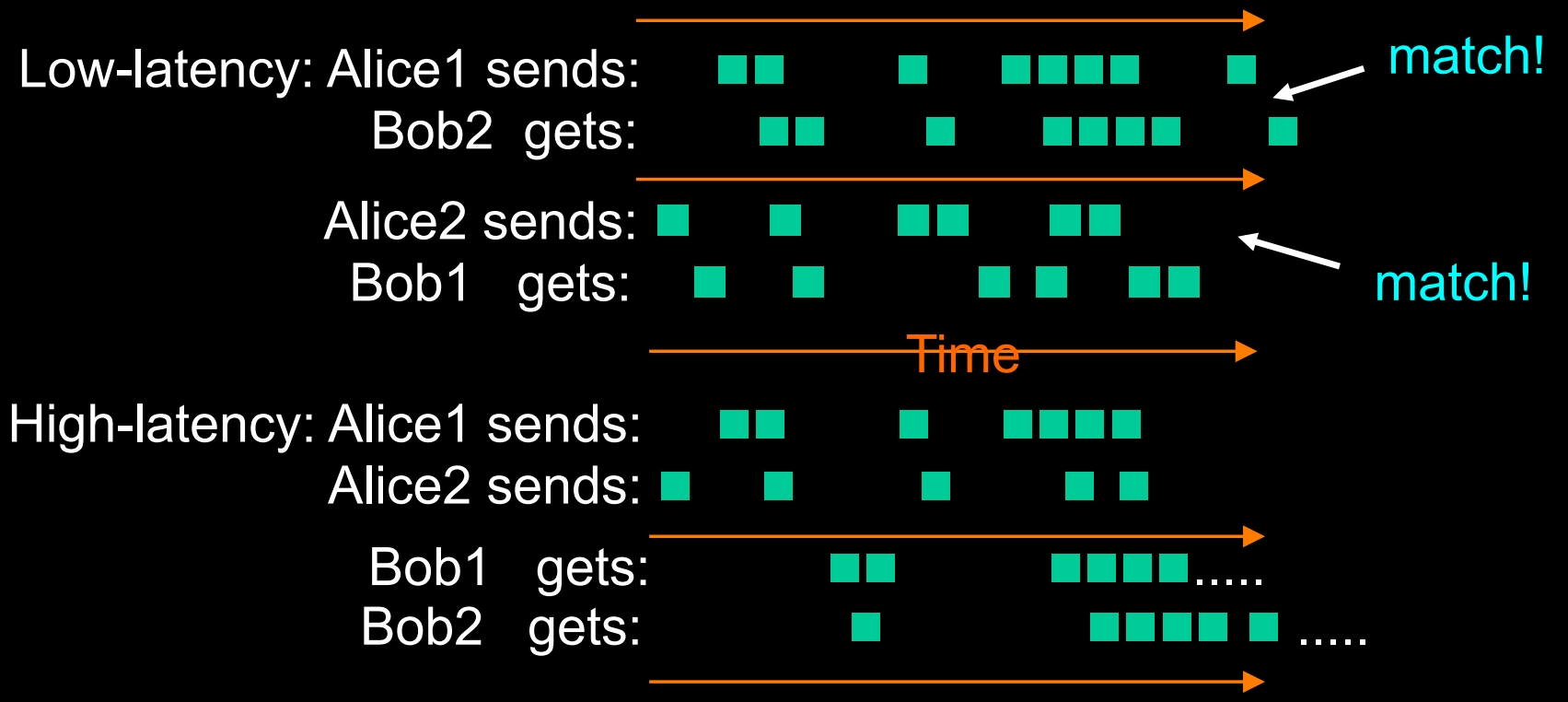
Mixes

- Invented by Chaum 1981 (not counting ancient Athens)
- As long as one mix is honest, network hides anonymity up to capacity of the mix
- Sort of
  - Flooding
  - Trickling
- Many variants
  - Timed
  - Pool
  - ...

Anonymity Systems for the Internet



Low-latency systems are vulnerable to end-to-end correlation attacks.



These attacks work in practice. The obvious defenses are expensive (like high-latency), useless, or both.

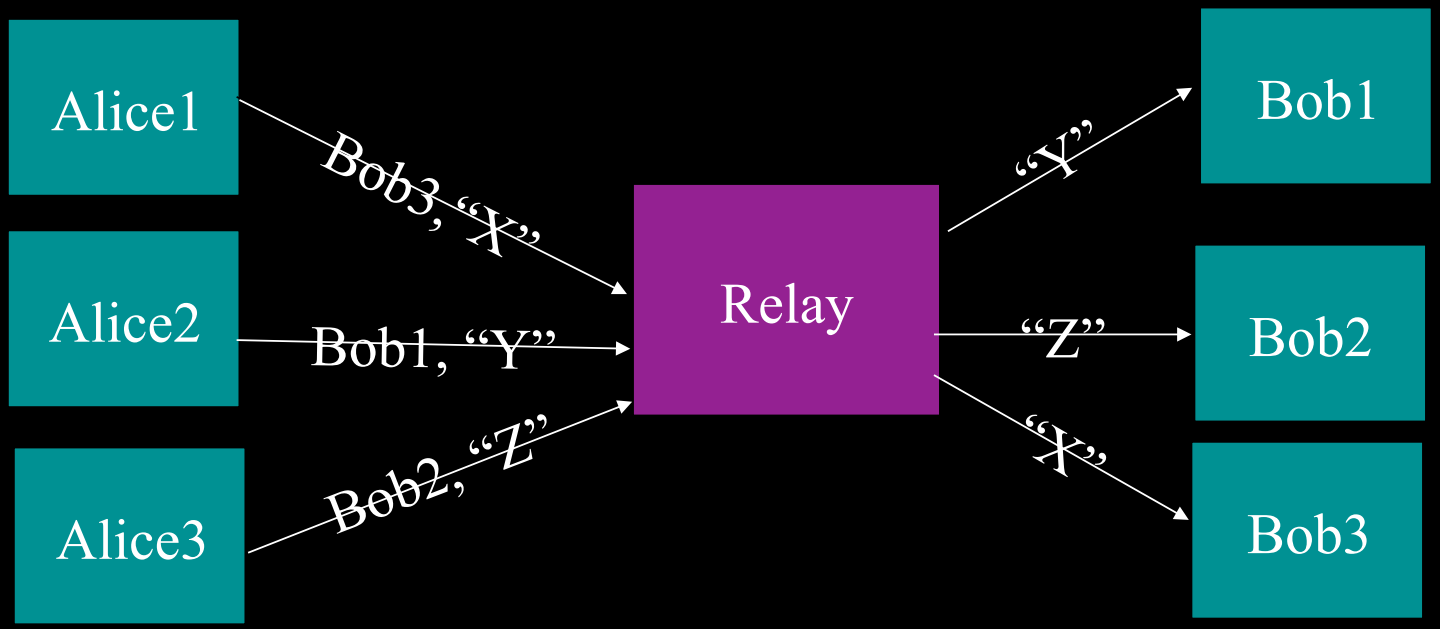
Still, we focus on low-latency, because it's more useful.

Interactive apps: web, IM, VOIP, ssh, X11, ...
 # users: millions?

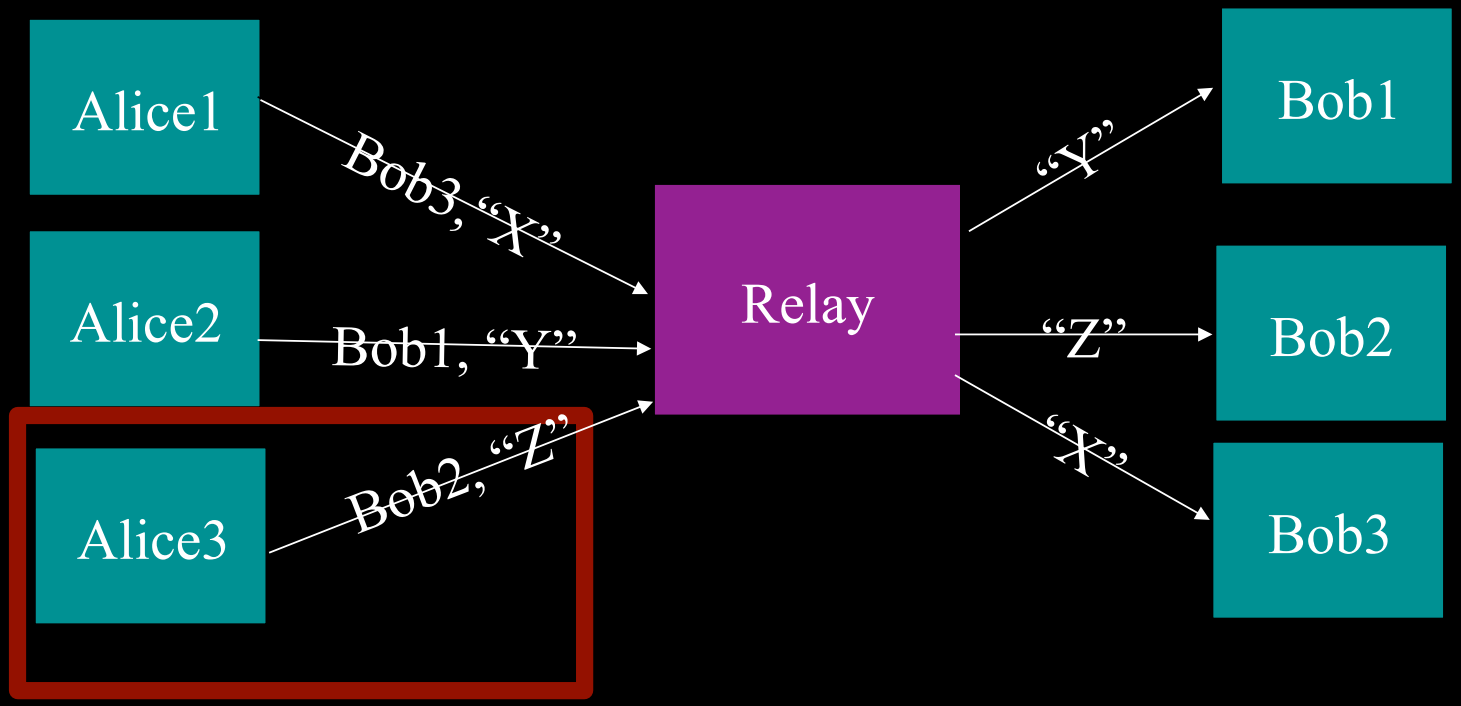
Apps that accept multi-hour delays and high bandwidth overhead: email, sometimes.
 # users: hundreds at most?

And if anonymity loves company....?

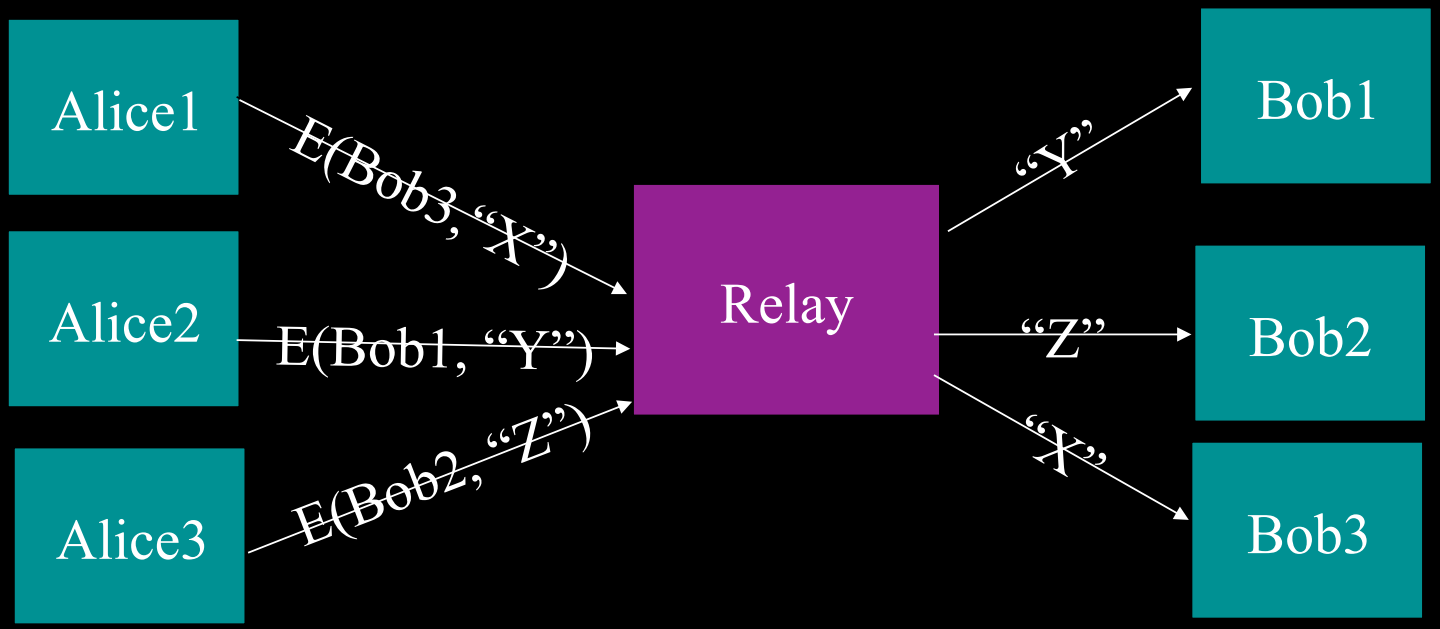
The simplest designs use a single relay to hide connections.



But an attacker who sees Alice can see who she's talking to.

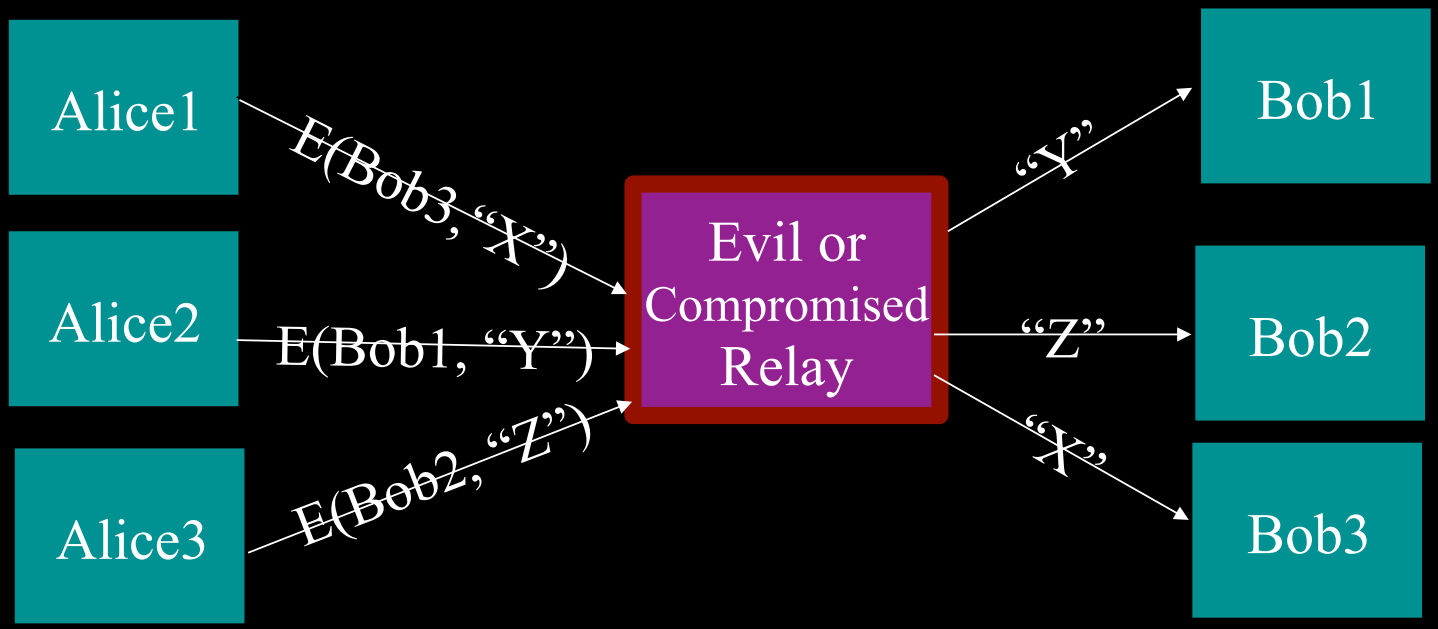


Add encryption to stop attackers who eavesdrop on Alice.

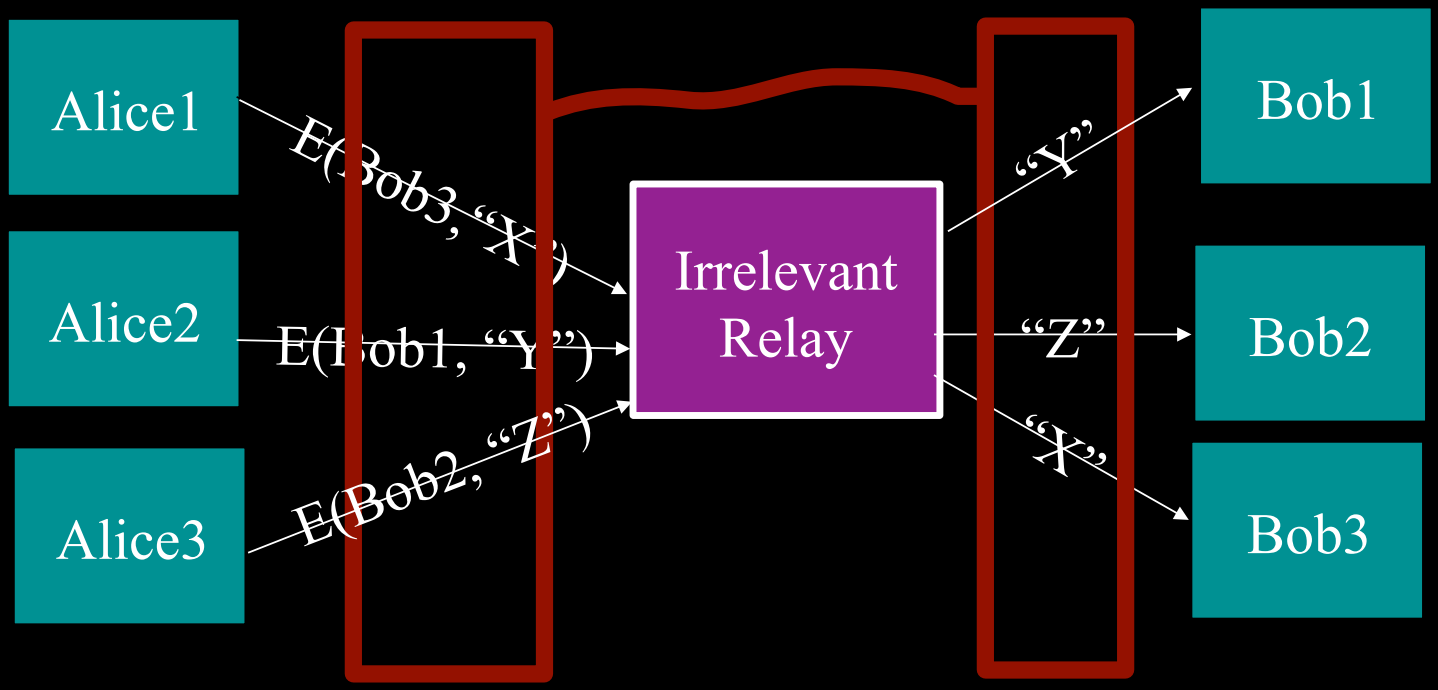


(e.g.: some commercial proxy providers, Anonymizer)

But a single relay is a single point of failure.

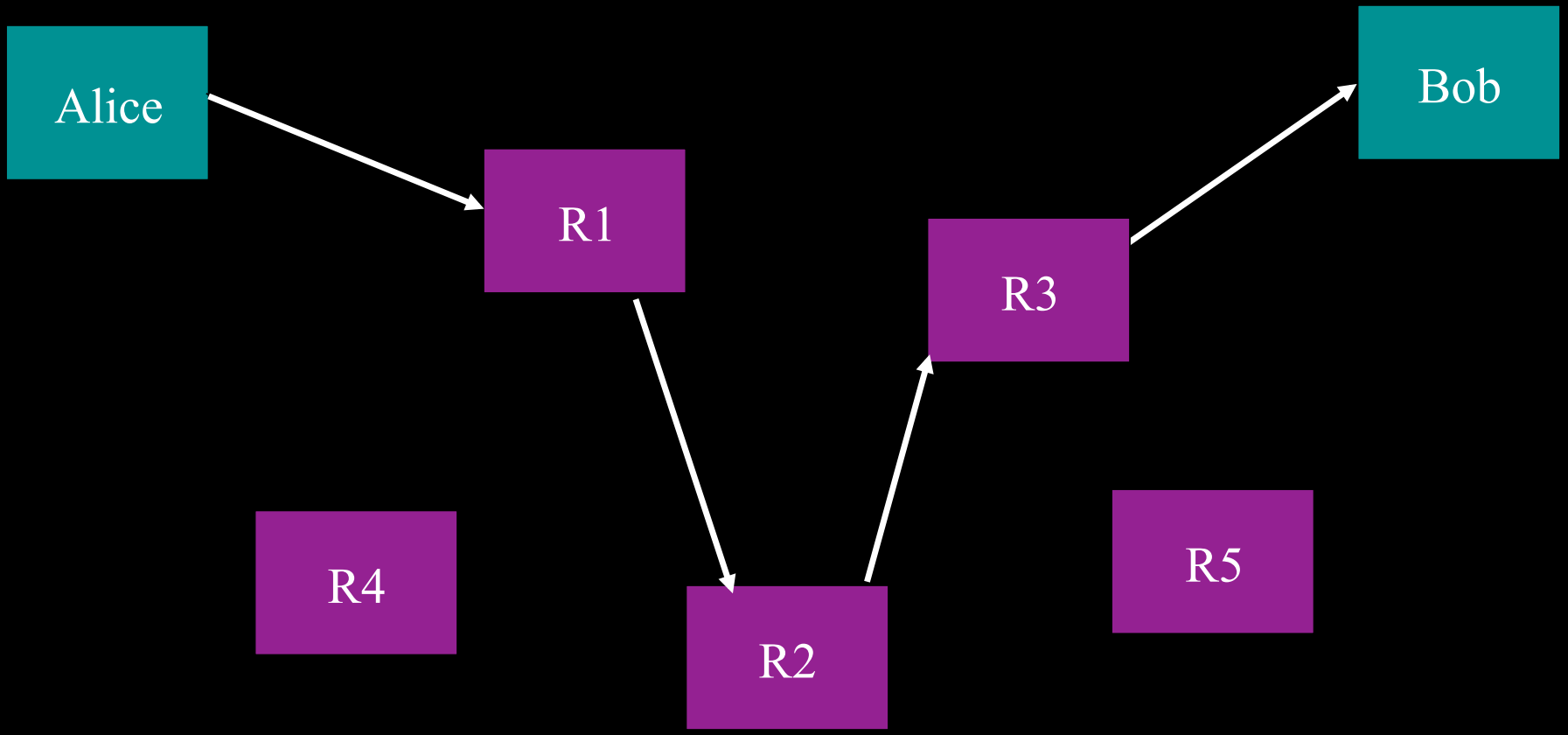


But a single relay is a single point of bypass.

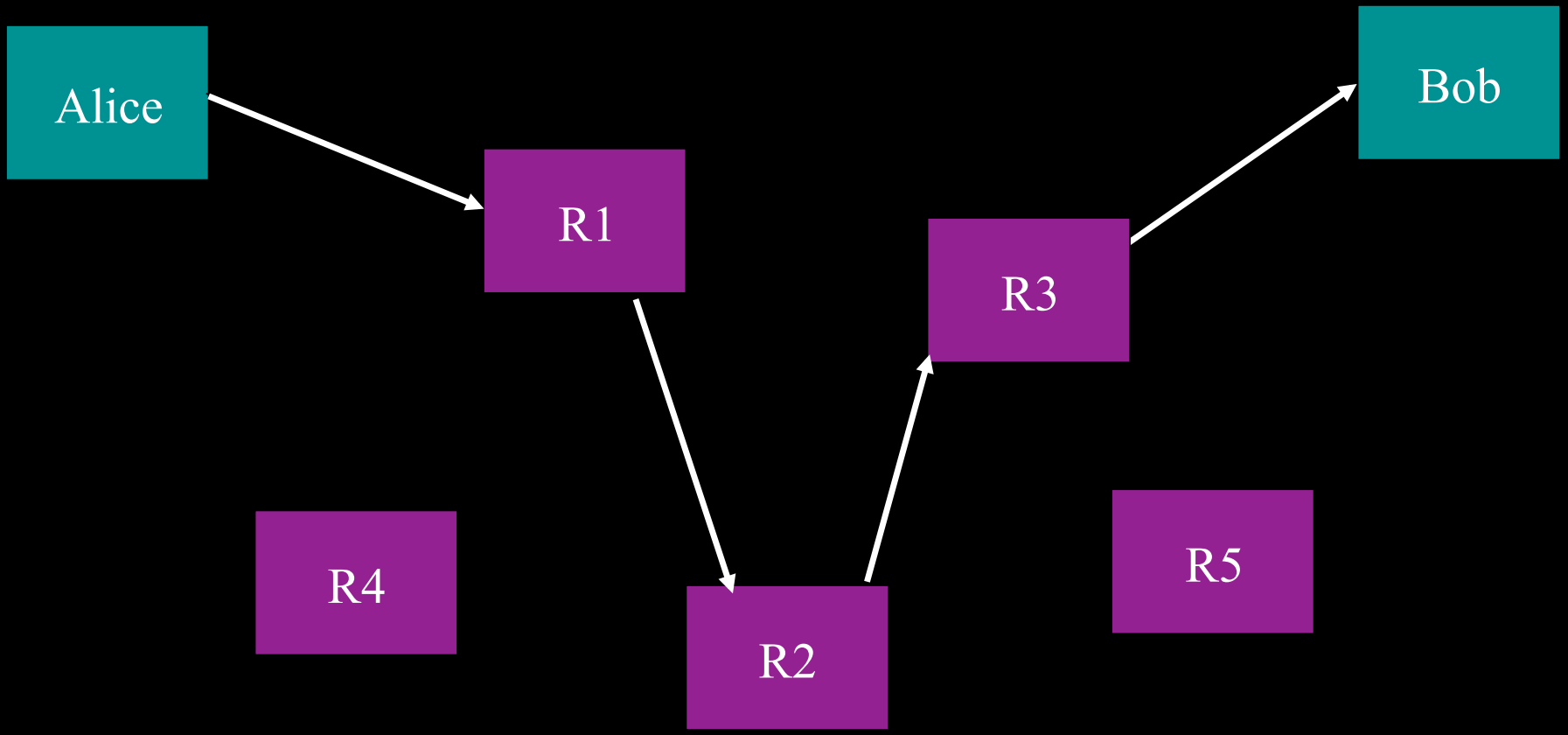


Timing analysis bridges all connections through relay ⇒ An attractive fat target

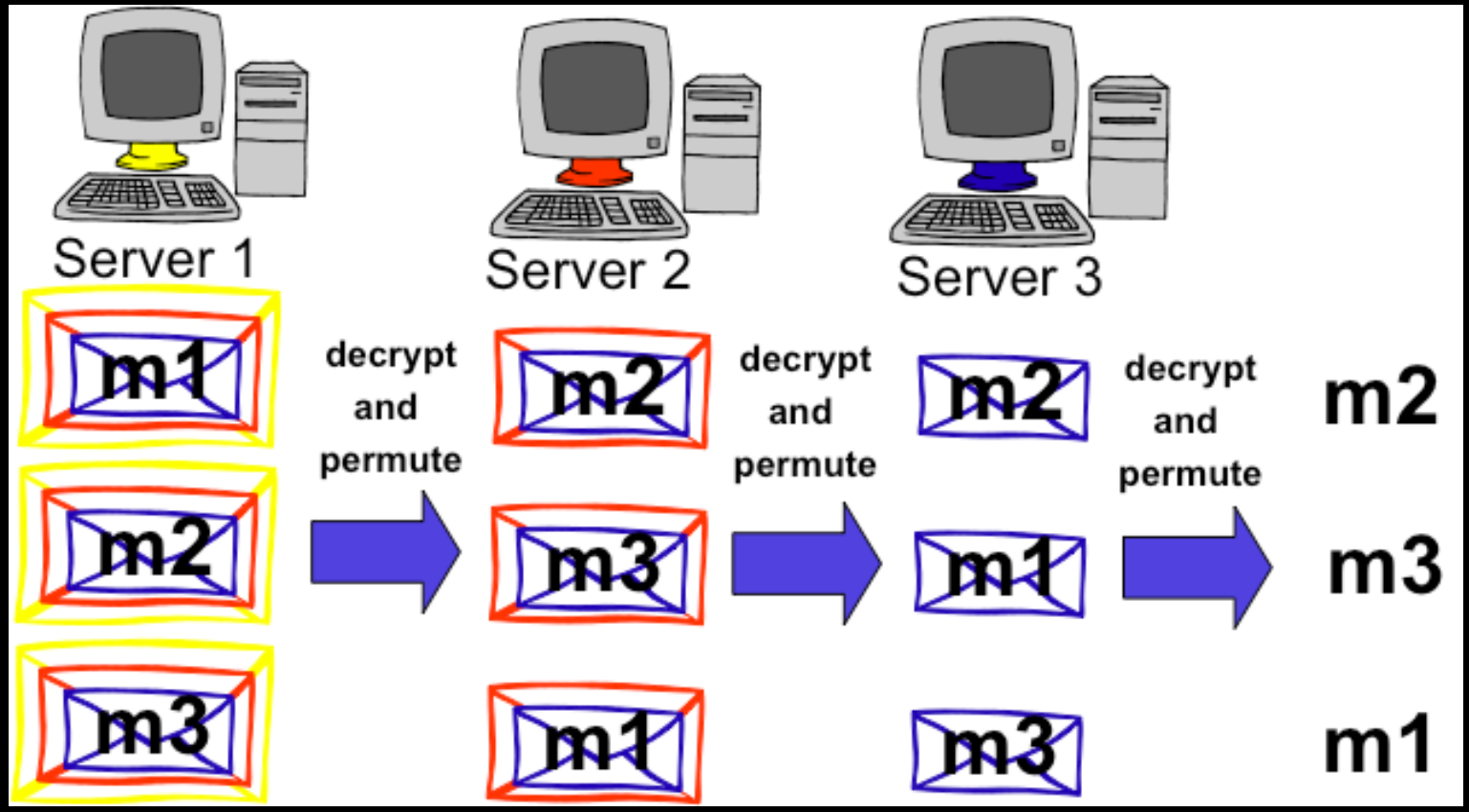
So, add multiple relays so that no single one can betray Alice.



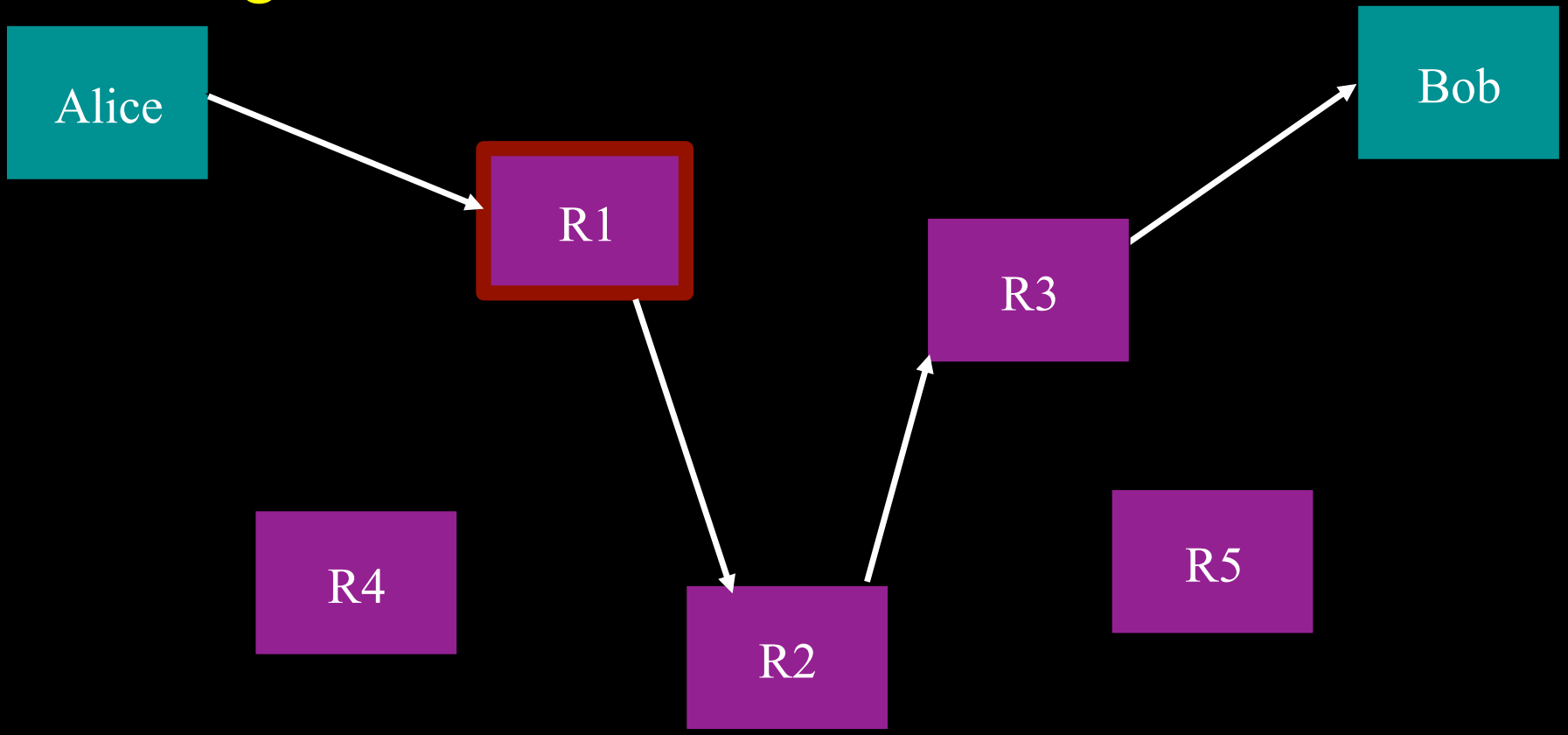
Multiple relay idea used in different ways by mix networks, Crowds, onion routing



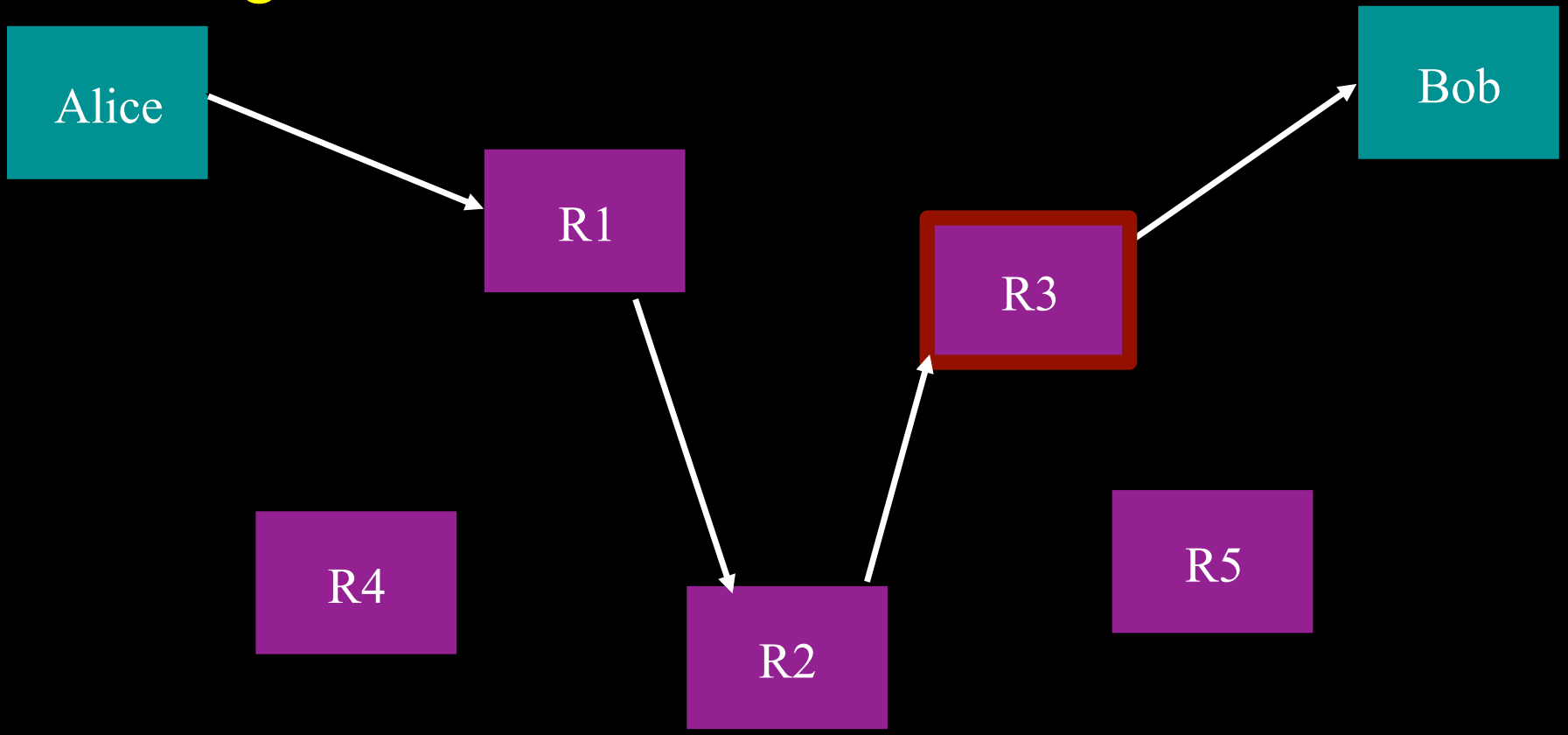
Already saw multiple relays in mix cascade



For Onion Routing and Mix Nets:
A corrupt first hop can tell that Alice is talking, but not to whom.



For Onion Routing and Mix Nets:
A corrupt last hop can tell someone is talking to Bob, but not who.



Crowds

Introduced by Reiter and Rubin in 1997

- Not the first distributed low-latency anonymity system.
  - Introduced about a year after the first onion routing deployment, and two years after Anonymizer.
- Not general purpose.
  - Exclusively for HTTP (not even HTTPS) traffic.
- Never widely deployed.
  - Largest Crowd in the wild had less than twenty users.

More Crowds limitations

- Requires all users to install and run Perl program
- Requires users to have longrunning high-speed internet connections
- Entirely new network graph needed to add new or reconnecting Crowd member
- Connection anonymity dependent on data anonymity
- Anonymity protection limited to Crowd size
- Not suitable for enclave protection
- All path members carrying your traffic have a complete pseudonymous profile of you

Why study the Crowds paper/design

Simple both in conception and implementation.

First peer-to-peer design (for any purpose? Years ahead of Napster, Gnutella, Bittorent, Chord,...).

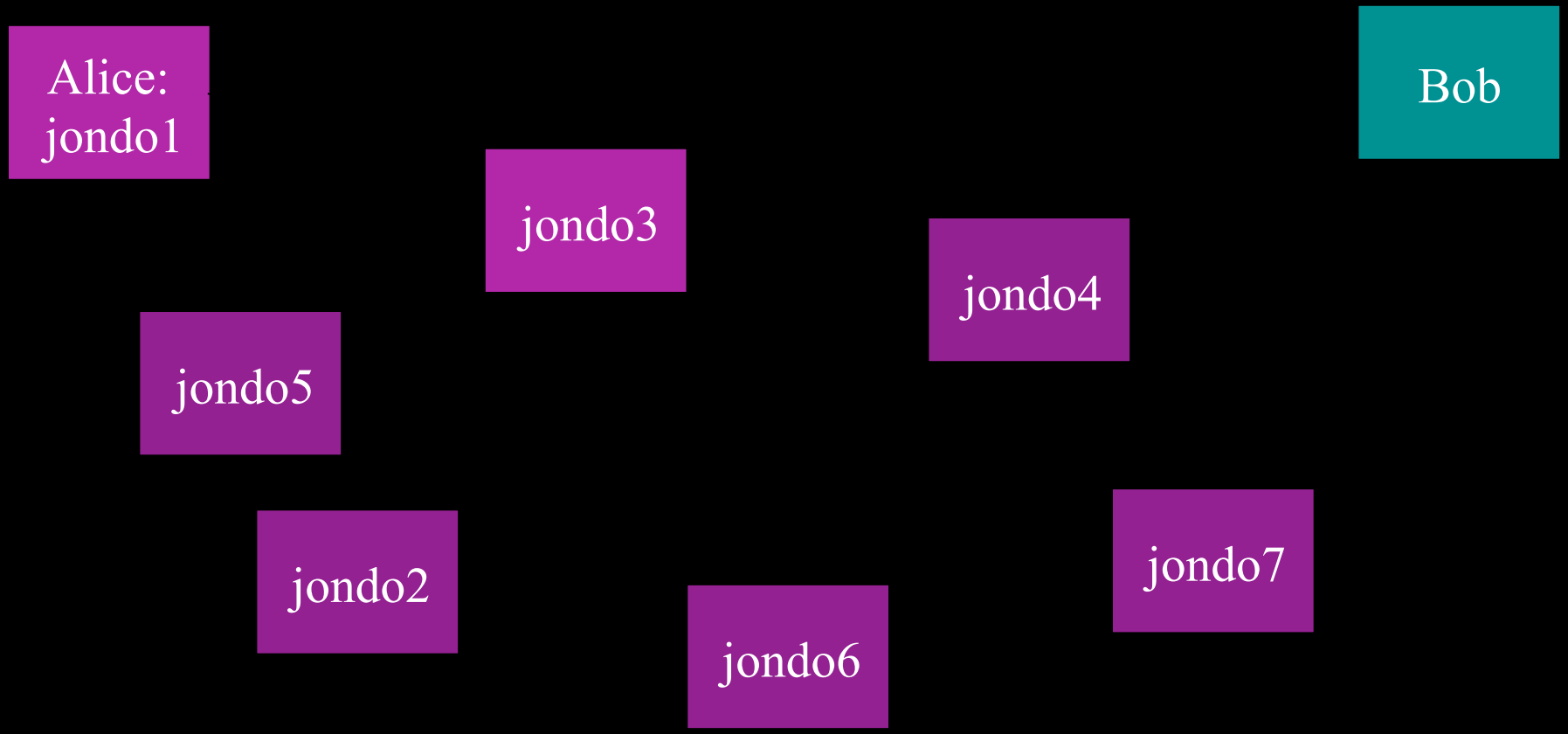
(Early onion routing was P2P in that all elements were the same, but were mostly not intended for end-user computers.)

First probabilistic analysis of anonymous communication.

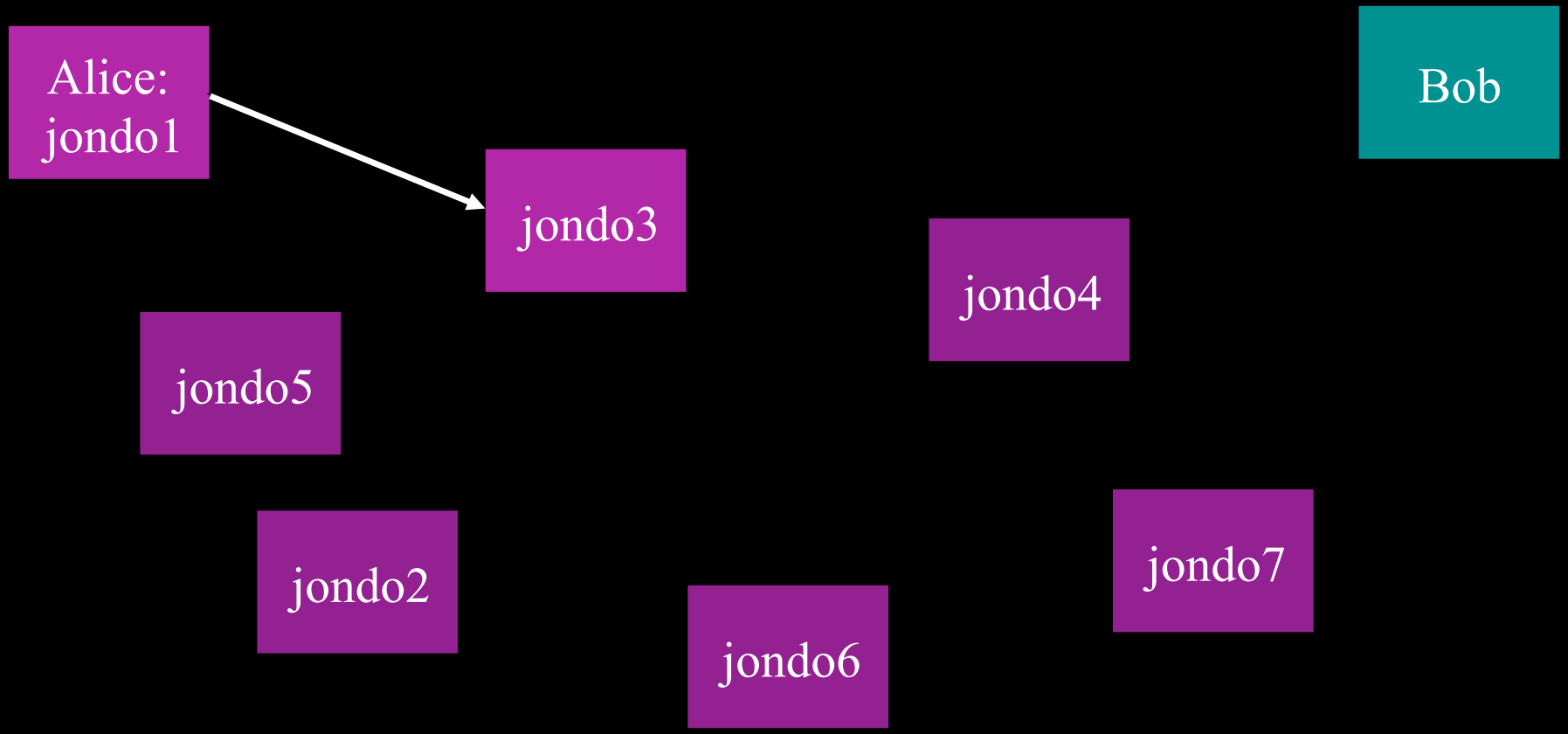
Introduced predecessor attack to the literature.

Introduced cautionary lessons about design.

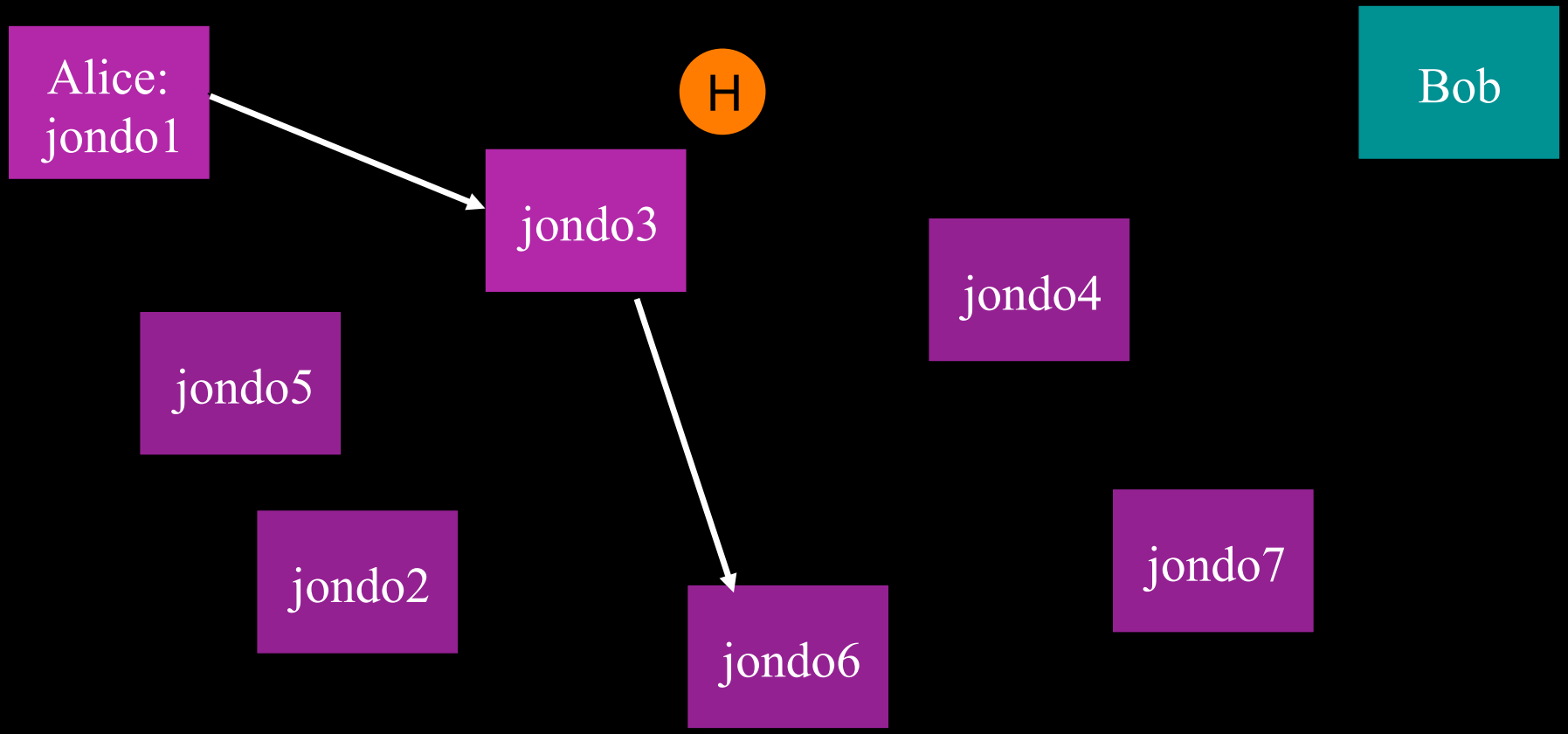
Alice is just one of the Crowd: jondo1



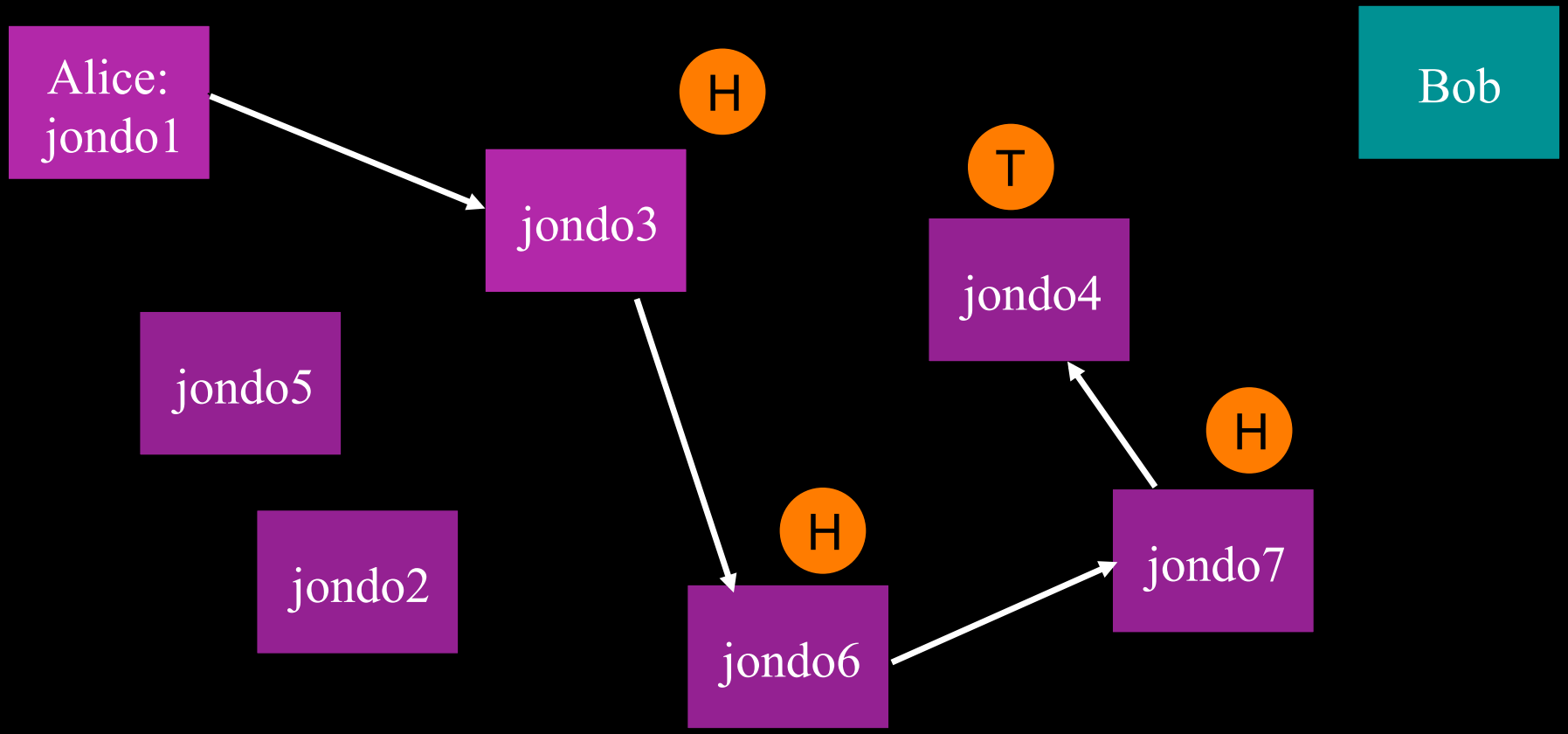
Alice connects to another Crowd member, e.g., jondo 3



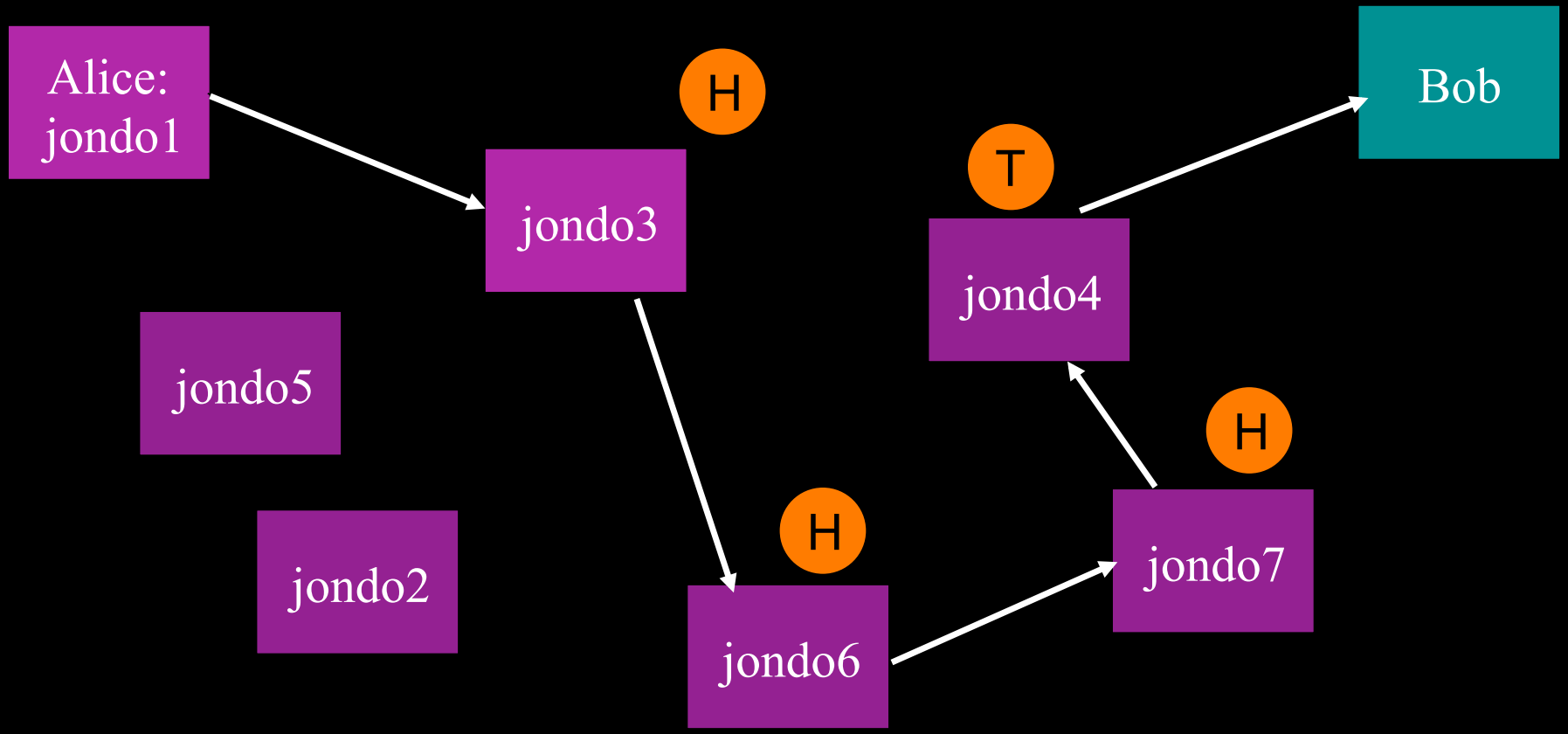
jondo3 flips weighted coin, forwards to another random crowd member if Heads



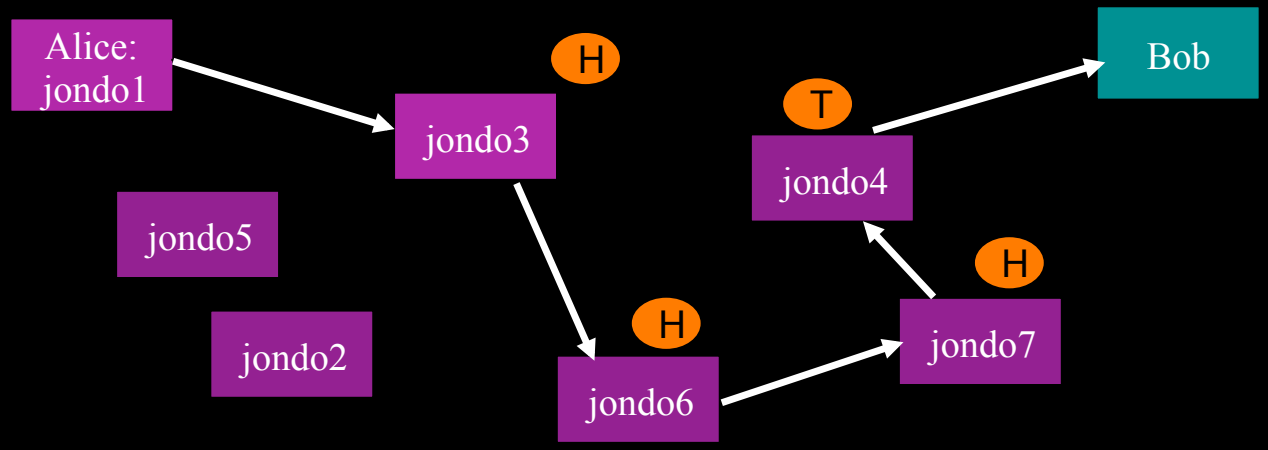
... continues until a coin comes up Tails.



... continues until a coin comes up Tails. That jondo decrypts connection request and forwards to server



- Crowd formed by a centralized “blender” that assigns membership and link keys to each pair of crowds members (limit to scaling)
- Pathkey distributed over link keys
- All path members have pathkey
- Return traffic travels back along same path
- All path members can decrypt and know destination and content
- Sender anonymity against path-members: a jondo cannot tell if predecessor is originator or not



Crowds notions of anonymity

Initiator (sender) anonymity: initiator's identity is hidden

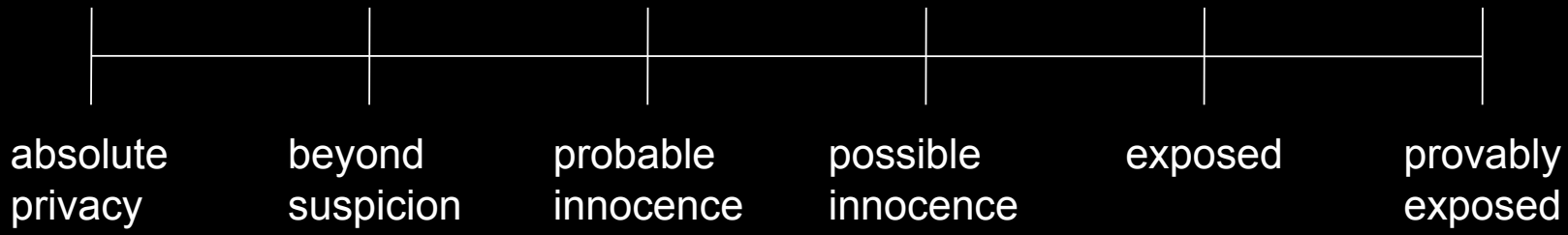
Responder (receiver) anonymity: responder's identity is hidden

Initiator-responder unlinkability: initiator and responder cannot be identified as communicating with each other

Crowds adversaries

- Local eavesdropper: can see all communication in and out of a user's computer.
- End Server: Web server interacting with user.
- Collaborating crowd member: can alter traffic patterns and content, can observe and share observations with other collaborators

Crowds degrees of anonymity



Absolute privacy: adversary sees no difference whether communication happens or not

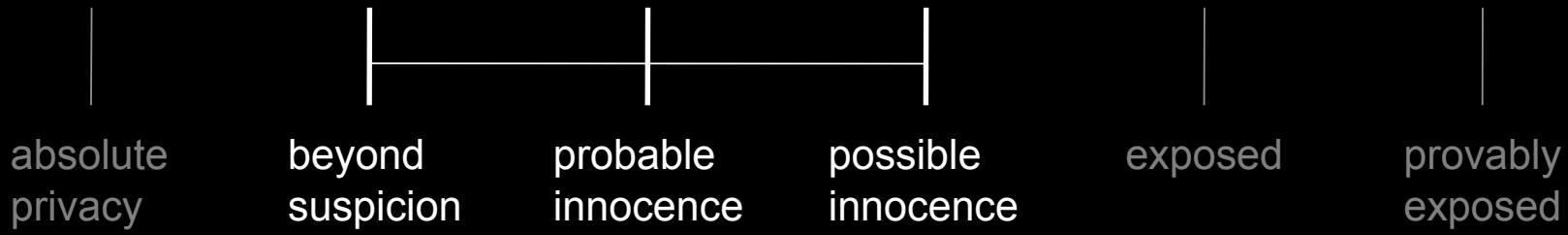
Provably exposed: initiator (responder/linking) is certain to adversary, and adversary can prove this to others

Beyond suspicion: initiator (...) is no more likely the source (...) of communication than any other potential source.

Probable innocence: initiator (...) is no more likely than not to be initiator (...)

Possible innocence: adversary places nontrivial probability on another initiator (...)

Crowds degrees of anonymity



Absolute privacy: adversary sees no difference whether communication happens or not

Provably exposed: initiator (responder/linking) is certain to adversary, and adversary can prove this to others

Beyond suspicion: initiator (...) is no more likely the source (...) of communication than any other potential source.

Probable innocence: initiator (...) is no more likely than not to be initiator (...)

Possible innocence: adversary places nontrivial probability on another initiator (...)

Crowds anonymity properties proven

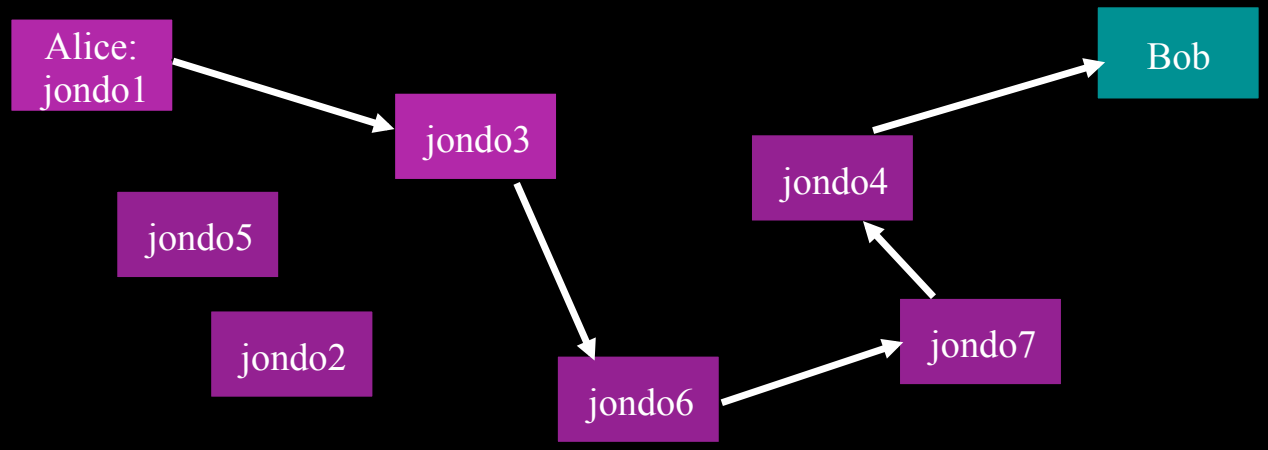
Table 1. Anonymity properties provided by Crowds

Attacker	Sender anonymity	Receiver anonymity
local eavesdropper	exposed	$P(\text{beyond suspicion}) \underset{n\to\infty}{\longrightarrow} 1$
c collaborating members, $n \geq \frac{p_f}{p_f - 1/2}(c+1)$	**probable innocence** $P(\text{absolute privacy}) \underset{n\to\infty}{\longrightarrow} 1$	$P(\text{absolute privacy}) \underset{n\to\infty}{\longrightarrow} 1$
end server	**beyond suspicion**	N/A

Table from ACM TISSEC ’98 Crowds paper

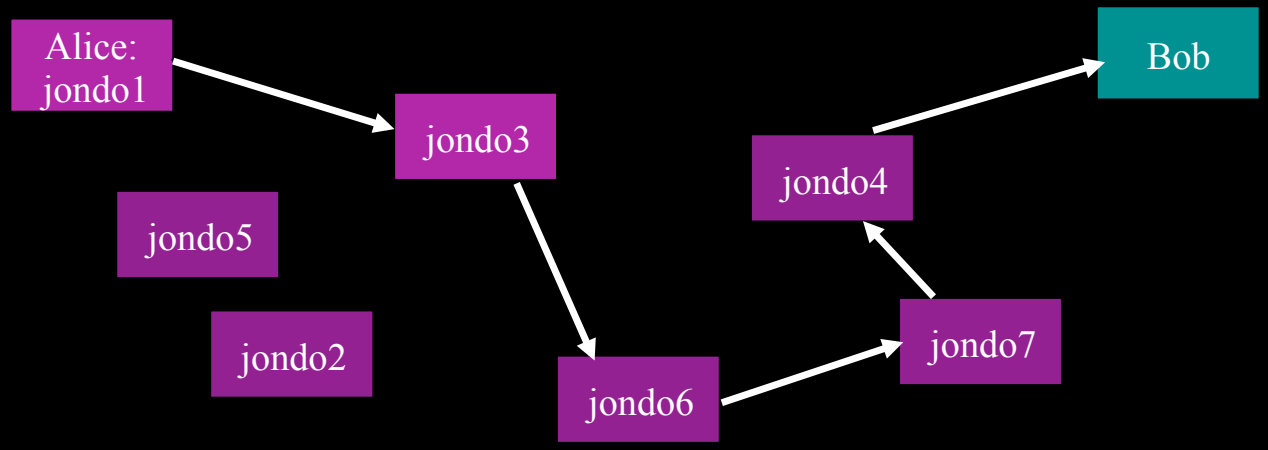
Timing attacks on Crowds

- For autoloaded content, e.g, embedded image requests: jondos can use response-request timing to determine position in path
- Crowds's solution: Last jondo automatically makes such response-requests and propagates the server response down the path
- The first jondo automatically blocks such requests and feeds responses to browser when the arrive
- Is this still a statistical threat for manual requests?
- Note side effect: Exit jondo does not simply forwarded content in each direction. This may have legal implications.



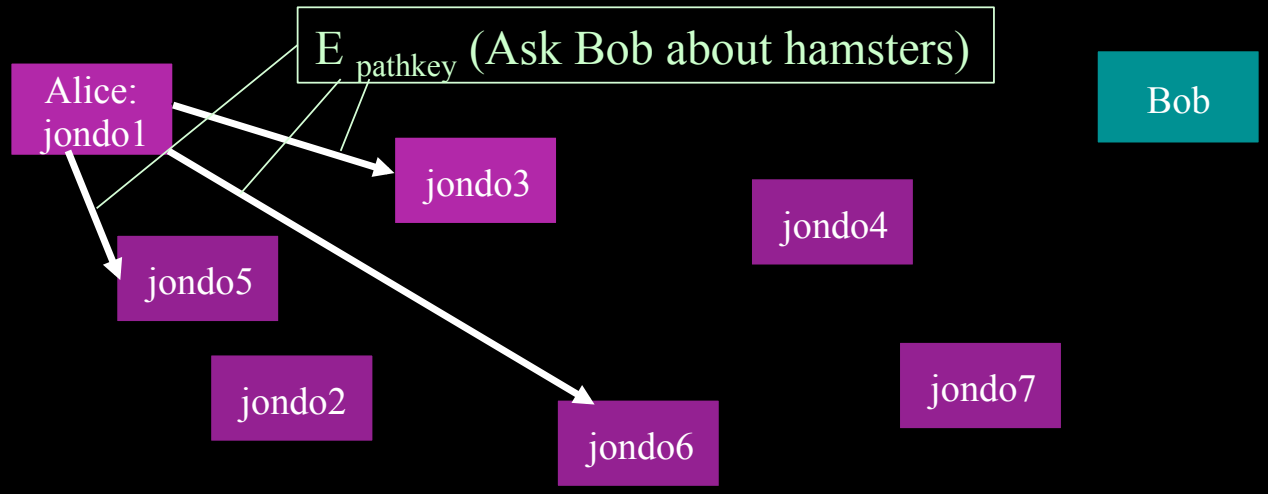
Connection capture, static paths, & forward anonymity

- Any corrupt path member can read or insert anything into path
  - Can try to insert malicious code or identifying scripts (path anonymity dependent on filter quality)
  - Chances of malicious path members increase with path length
- Static paths: path essentially remains for lifetime of crowd.
  - Route capture is more cost effective (one attack works longer)
  - Richer profile attack (all HTTP connections during crowd in a single profile)
  - Bad forward anonymity (identification of any transaction links to whole profile)



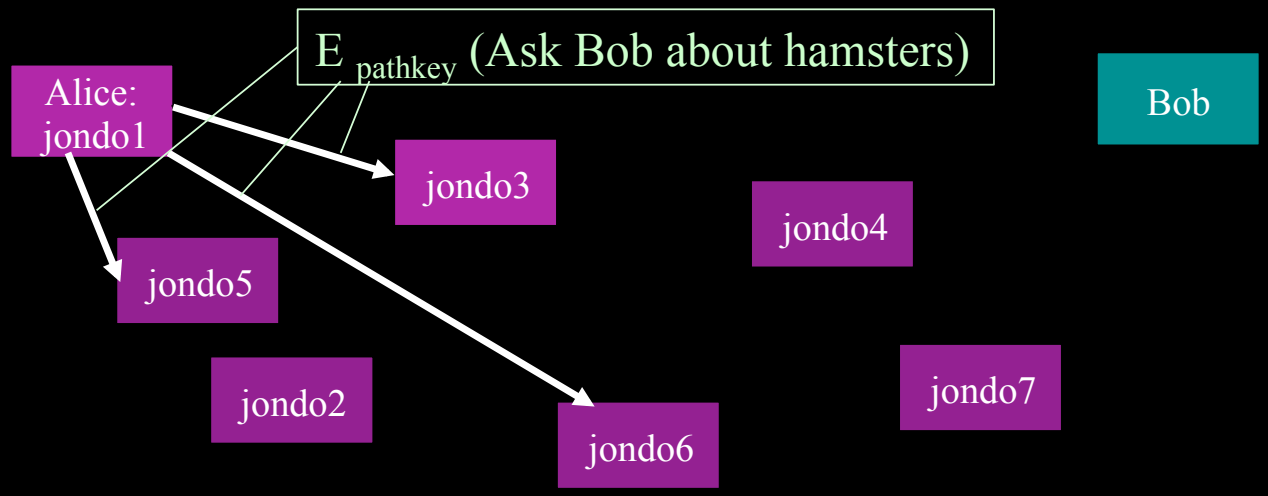
Dynamic paths & predecessor attacks

- Dynamic paths would reduce the pseudonymous profiling
- Because content is known to path members, dynamic paths could lead to intersection attacks
- Paths are rebuilt in only two circumstances
  - If a connection breaks, path is just rebuilt from that point on
  - When a new member (re)joins the network, the whole crowd reforms to protect it



Predecessor attacks on reformation

- Wright et al., Adonieh et al., Shmatikov all c. 2002 looked at predecessor attacks on Crowds and other systems
- Shmatikov showed precision of predecessor attack increases with crowd size (Prob (no false pos | positive))
  - using PRISM (probabilistic model checker) that crowd size, not just number of path reformations matters
  - Anonymity degrades fairly fast



Predecessor results from PRISM

Crowd		Path reformulations 3	4	5	6
5 honest, 1 corrupt	P_{pos}	13.8%	23.5%	33.3%	42.7%
	P_{conf}	100.0%	97.4%	93.1%	**86.9%**
	P_{fpos}	5.1%	9.1%	12.9%	15.8%
10 honest, 2 corrupt	P_{pos}	10.4%	18.1%	26.3%	34.6%
	P_{conf}	100.0%	98.9%	96.2%	**92.5%**
	P_{fpos}	2.9%	5.5%	8.2%	10.8%
15 honest, 3 corrupt	P_{pos}	9.4%	16.5%	24.1%	31.8%
	P_{conf}	100.0%	98.9%	97.5%	**95.0%**
	P_{fpos}	2.0%	3.9%	5.9%	7.9%
20 honest, 4 corrupt	P_{pos}	**8.9%**	**15.6%**	**23.0%**	**30.5%**
	P_{conf}	100.0%	99.4%	97.8%	**96.1%**
	P_{fpos}	1.6%	3.0%	4.6%	6.3%
10 honest, 1 corrupt	P_{pos}	3.7%	6.8%	10.5%	14.5%
	P_{conf}	100.0%	99.6%	98.1%	96.6%
	P_{fpos}	1.6%	3.0%	4.8%	16.8%
20 honest, 2 corrupt	P_{pos}	3.0%	5.5%	8.6%	12.0%
	P_{conf}	100.0%	99.6%	98.8%	98.3%
	P_{fpos}	0.8%	1.6%	2.6%	3.8%

Table 3. Probabilities of observations by the adversary.

Table from Journal of Computer Sec. ’04 paper

Wisdom from Crowds

Anonymity is tricky: Even when you know there is a threat, you might underestimate how bad it is

Anonymity is tricky: Doing something to make you more secure can make you less secure

- Static paths to avoid predecessor attacks ➔ worse against profiling (likewise for higher prob. of forwarding)
- Larger anonymity set ➔ less risk of single-path identifying initiator but great risk of confident exposure
- HTTPS reduces risk from data exposure but implies an evil successor exposes initiator with high probability

- Anonymity is tricky: Danezis et al., ESORICS 2009 showed that attempts to vary probability of forwarding reduced anonymity and that Crowds had made optimal choice

What's up next (and what questions do you have now?)

Lecture 1:

- Usage examples, basic notions of anonymity, types of anonymous comms systems
- Crowds: Probabilistic anonymity, predecessor attacks

Lecture 2:

- Onion routing basics: simple demo of using Tor, network discovery, circuit construction, crypto, node types and exit policies
- Economics, incentives, usability, network effects