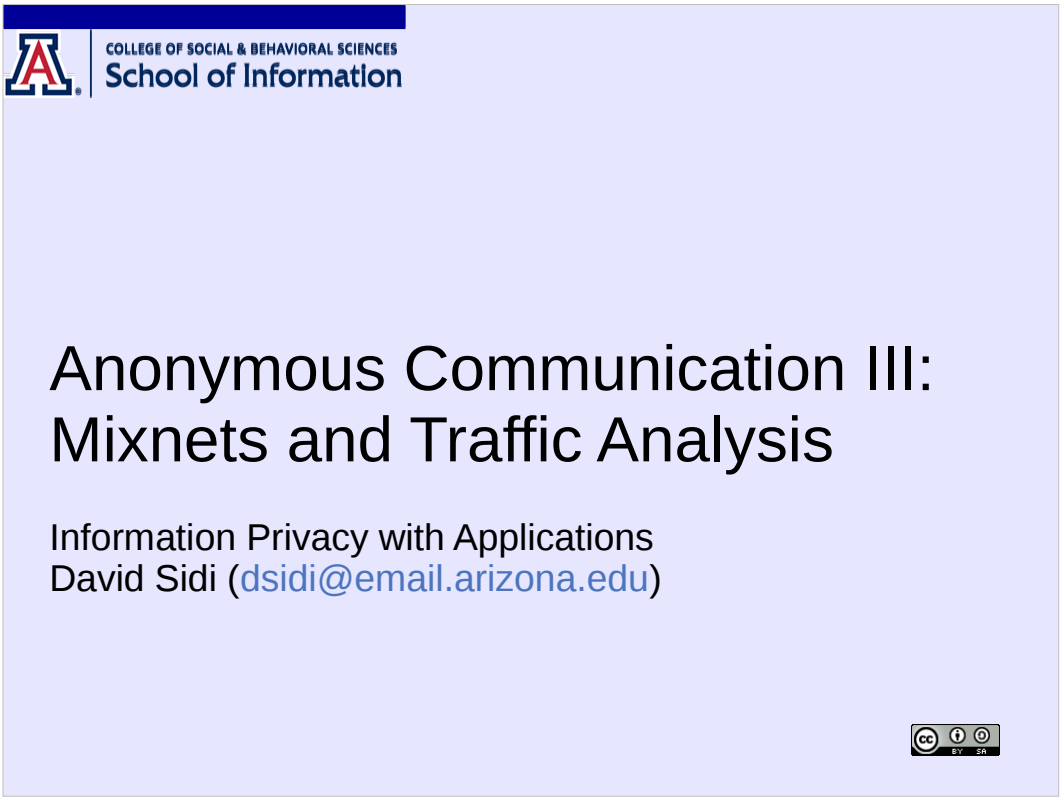
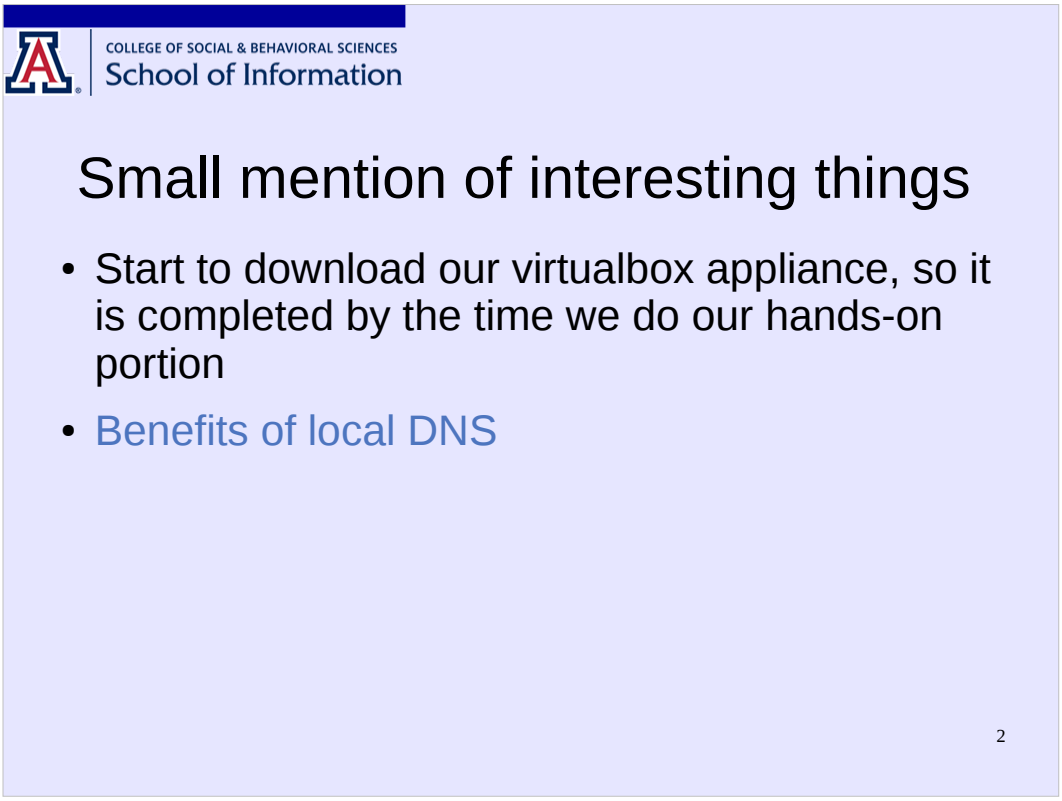


We’re going to discuss mixnets and onion routing today, and do a little traffic analysis ourselves



recall from last time that DNS leaks information about your browsing.

COLLEGE OF SOCIAL & BEHAVIORAL SCIENCES
School of Information

Anonymity networks

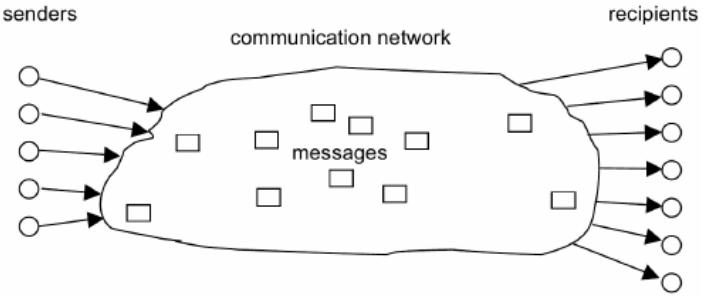


Image credit (before modification):
Christina Pöpper
Ruhr-University Bochum

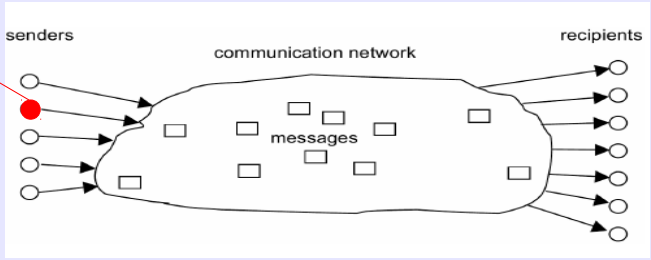
senders communicating messages over a channel with recipient.

Anonymity is a property of a channel. Loosely, such channels obscure who communicates with whom, and when (relationship anonymity for Chaum). Contrast with message confidentiality

How do we understand parties to the communication?

COLLEGE OF SOCIAL & BEHAVIORAL SCIENCES
School of Information

	MAC	Browser_fingerprint	IP	Sites_visited
SNDER_1	00:a0:ef:eb:5v:ff	af7f098c39728f8cb676e3df82ced01a149ee3aa92af2b88c20c4948a5fad5fd		torproject.org, ischool.arizona.edu, maps.google.com...
SNDER_2	00:c0:ff:dd:ff:ef	a5fad5fdd01a149eeaf7f098c39728f8cb676e3df82ce3aa92af2b88c20c4948		nytimes.com, purple.com



- you can think of senders/recipients as records in a database: things that can be identified (“subjects,” “persons,” etc...) are descriptions via a set of attribute value
- an attacker seeking to reduce anonymity seeks a combination of message attributes that she can measure with confidence, and that together distinguish participants from one another

COLLEGE OF SOCIAL & BEHAVIORAL SCIENCES
School of Information

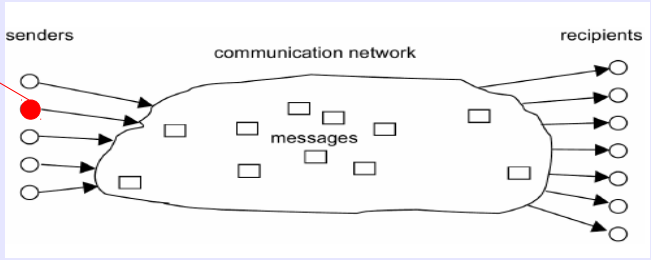
CC-SA License by David Sidi

- Suppose I include in the records of a database of UA students' medical history a person's weight and height as 150 lbs, 5'3". Is the person anonymous?

- Now suppose further that I do so for a database of male UA basketball players. Obviously, the player is not (as) anonymous. some attributes contribute more to identifiability than others in different sets of possible identities.
- Where might you find combinations of control information for messages that could be used to identify people on computer networks?
  - IP, Ethernet, DNS, third-party tracking, browser fingerprinting

COLLEGE OF SOCIAL & BEHAVIORAL SCIENCES
School of Information

	MAC	Browser_fingerprint	IP	Sites_visited
SNDER_1	00:a0:ef:eb:5v:ff	af7f098c39728f8cb676e3df82ced01a149ee3aa92af2b88c20c4948a5fad5fd		torproject.org, ischool.arizona.edu, maps.google.com...
SNDER_2	00:c0:ff:dd:ff:ef	a5fad5fdd01a149eeaf7f098c39728f8cb676e3df82ce3aa92af2b88c20c4948		nytimes.com, purple.com



- there are different steps along the way to identifying a communicant via attributes: whether they are present on the network (note that if this is known, it can in some cases be used to bound the size of the anonymity set), whether a set of communications can be bundled as sharing one or both endpoints. Ultimately, what an attacker wants is to link /possible/ communications

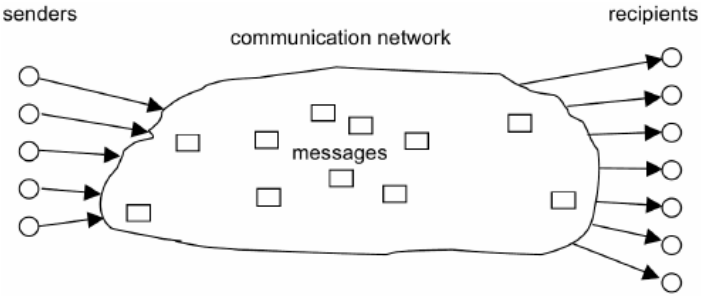
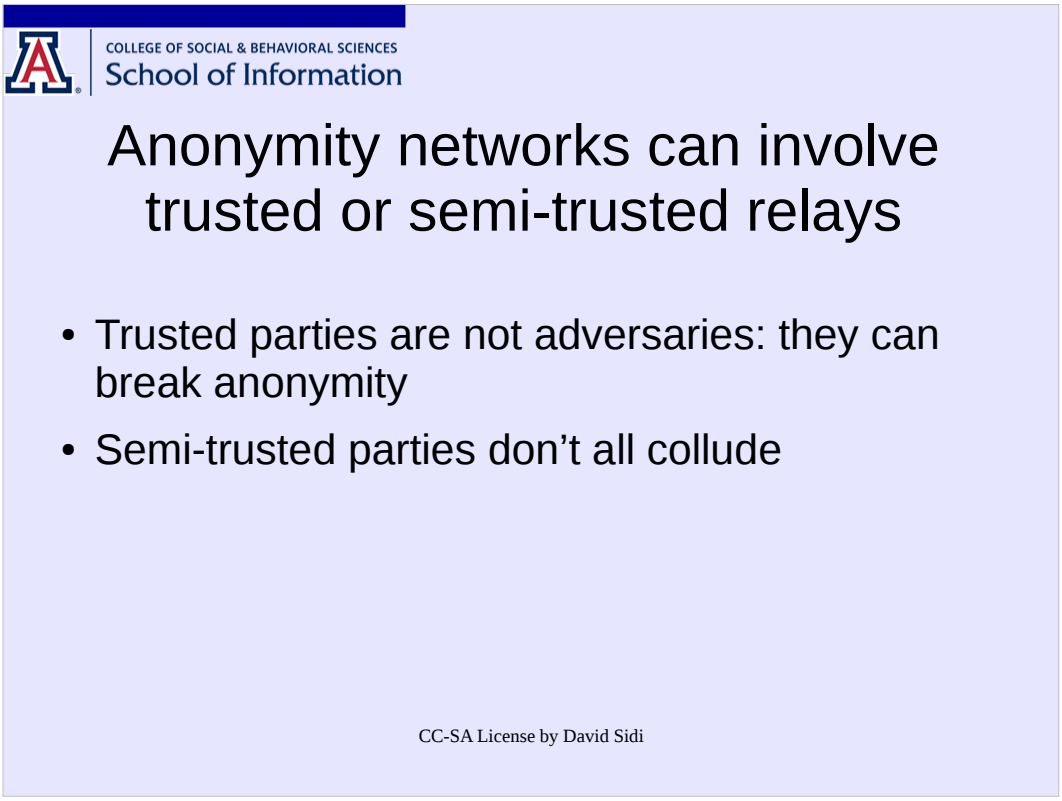


Image credit (before modification):
Christina Pöpper
Ruhr-University Bochum

in addition to the steps toward identification, there are different targets: sender, recipient, and relationship. Most anonymity metrics focus on sender anonymity---which active sender did a message exiting the network come from? Case for recipient anonymity is similar.

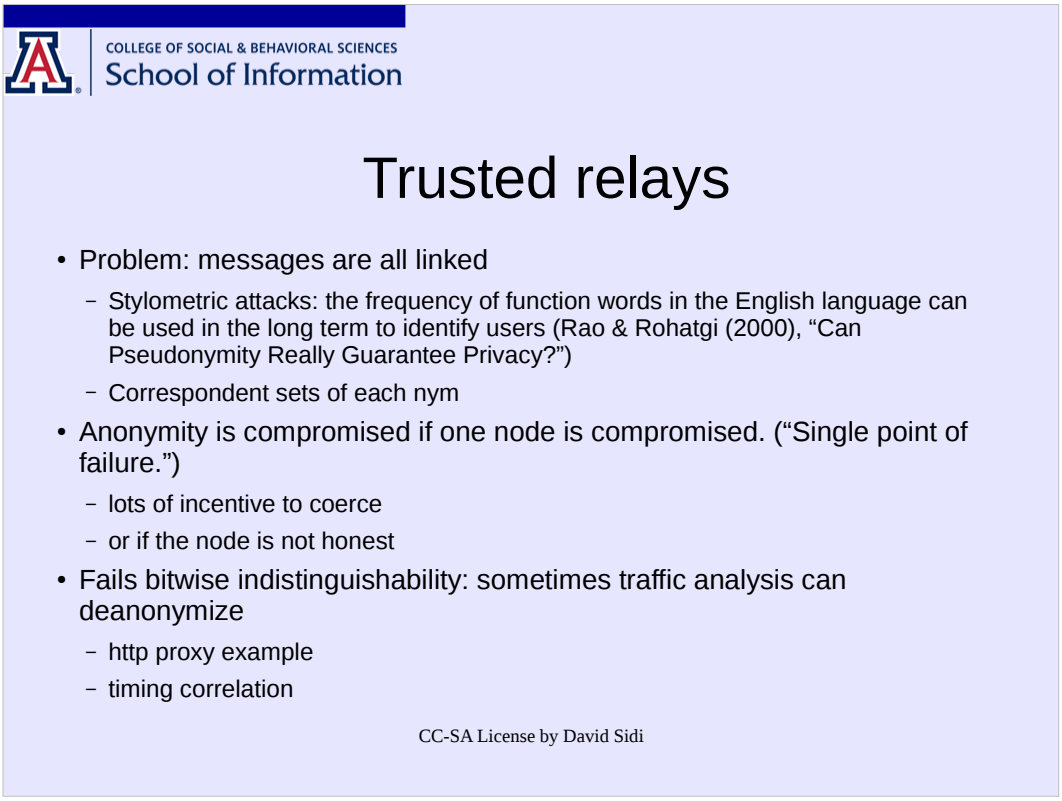
Shmatikov and Wang point out that even with sender anonymity, relationship anonymity may fail. Consider a property over many messages: at least one came from person S. Same argument applies to recipient anonymity.

Anonymity networks aim to prevent traffic analysis from succeeding in deanonymizing a channel (they often also employ technologies of confidentiality).



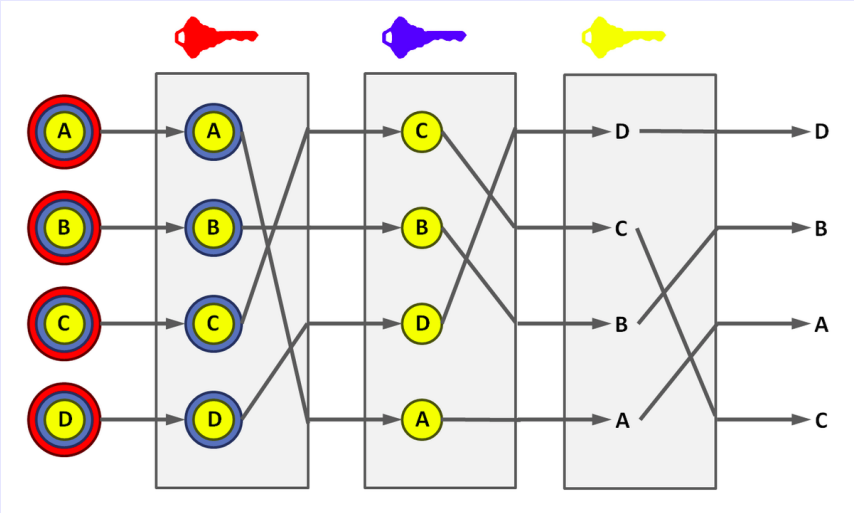
- Example of trusted relays: Nym servers
  - a server keeps a dictionary between real and pseudonymous emails
  - request comes to the remailer, which forwards it, gets the response, and returns it to the user
  - Example: anon.penet.fi
- Other Examples: Anonymous proxies (startpage.com), VPNs

- semi-trusted means not all relays collude



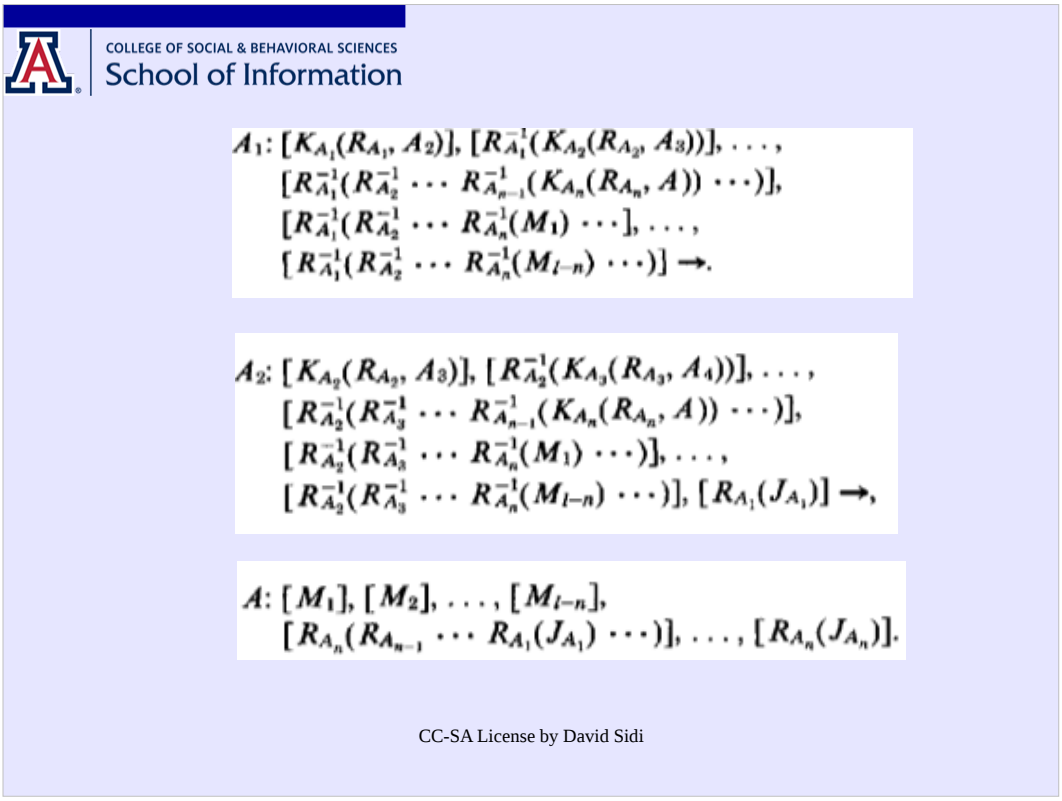
- messages are all linked as coming from the same sender, and having the same destination. How to associate that bundle of messages with a person involves linking the (potentially rich) information it contains. examples are stylometry, sets of all correspondents for a nym, single point of failure.
- bitwise indistinguishability is roughly the inability to associate patterns of bits going into a relay with a pattern of bits going out: if a trusted relay is used in a simple way this may fail

COLLEGE OF SOCIAL & BEHAVIORAL SCIENCES
School of Information

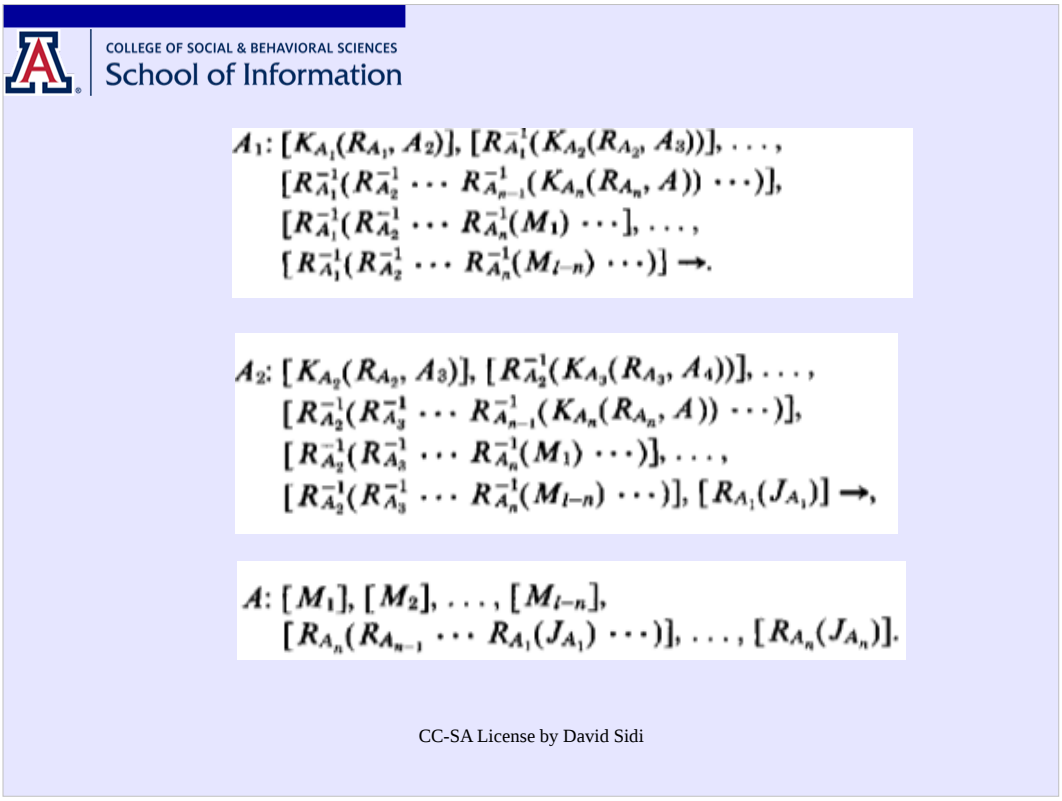


wikipedia.org

- Mixnets are anonymity systems with semi-trusted relays
- Routing protocol with a cascade of cryptographic relays called ‘mixes’
- Mixes only know their neighbors
- User-specifiable routing is an elaboration (Chaum’s “new kind of mix”)



- hybrid scheme with symmetric and asymmetric cryptosystems
- Suppose we are at a mix A1, which receives an encrypted message m.
- before being packaged, m has been split to fit into a final message with a fixed size ℓ. That is, there are ℓ-n pieces of the message in the encrypted blob sent to the mixes, where n is the number of mix nodes in the network.
- The first block is a bit like headers we've seen in TCP/IP: it contains the symmetric key R_{A1} and address A2 for the next hop, both encrypted using the public key for the mix (this is the hybrid part). The header part is stripped off, and a padding ("junk") block is encrypted using the symmetric key and added to the end
- The rest of the blocks are, first, the header blocks for all remaining routers in the cascade, then the final destination exiting the mixnet, and finally the message pieces.



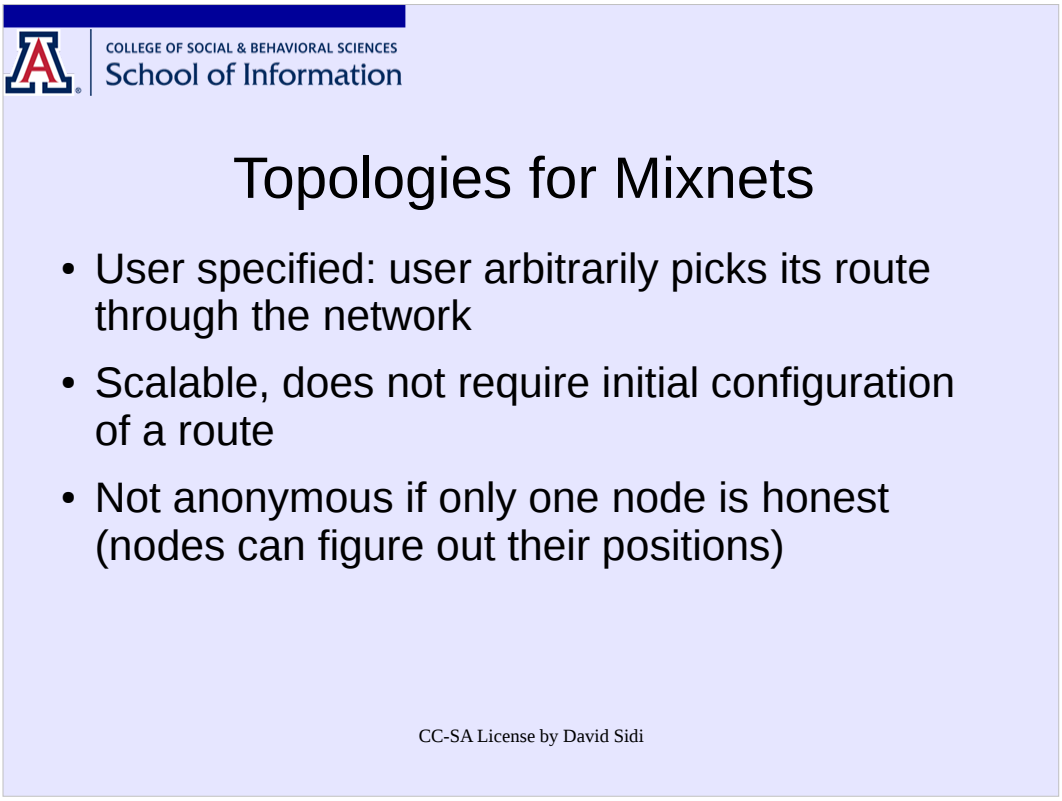
- A_1 uses the R_{A1} it now has to decrypt the the header for the next hop, one layer of the multiply-encrypted final recipient of the message once it exits the mixnet, and one layer of the multiply encrypted message.
- The result is then passed to the next mix node A_2

- Mixes only know their neighbors. (Question: Why?)
- Notice confidentiality is not handled at the exit node
- Weaknesses
  - active attacks: tagging attacks (blind signing attack), replay attacks
  - slow (public-key cryptography, latency in anonymous remailers).

COLLEGE OF SOCIAL & BEHAVIORAL SCIENCES
School of Information

Topologies for Mixnets

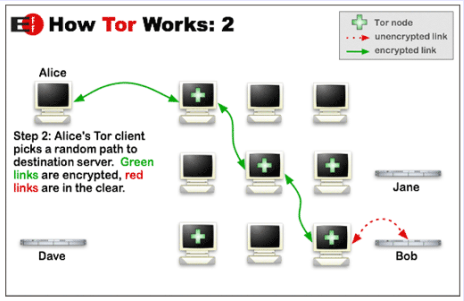
- Cascading: All nodes are always used, in the same order
- Scalability is a problem, requires setting up a fixed route with all nodes
- Only requires one honest node to preserve anonymity



relays can determine their position in the chain, which can be used to deanonymize with enough collusion

Onion Routing

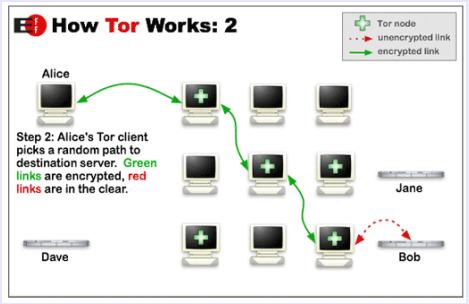
- First was from the US Naval Laboratory, 1996
  - pure peering at this stage, loafers!
- Freedom Network was an independent onion routing network from Zero Knowledge Systems
- Tor is a third-gen. onion routing network



COLLEGE OF SOCIAL & BEHAVIORAL SCIENCES
School of Information

Tor is an onion routing system

- Example: simplified, slightly out-of-date Tor (*link*)
- Distributed TCP overlay network
- Sets up a “virtual circuit” as a cascade of three onion relays (OR) from the initial client onion proxy (OP)
- guard (from “helper nodes”), relay, and exit nodes
  - each node only knows its immediate predecessor and successor



each time a user creates a circuit, there is a small chance that the circuit will be compromised. However, most users create a large number of Tor circuits, so with the original path selection algorithm, these small chances would build up into a potentially large chance that at least one of their circuits will be compromised.

For users who have good guard nodes, the situation is much better, and for users with bad guard nodes the situation is not much worse than before.

COLLEGE OF SOCIAL & BEHAVIORAL SCIENCES
School of Information

Tor is an onion routing system

- originally, onion routing systems sent an initial onion message that was “just layers” to set up the circuit; Tor does it in stages (“telescoping”)
- Next hop in the circuit is determined by unwrapping an “extend” relay cell with a symmetric key, which causes the OR to send its own “create” control cell

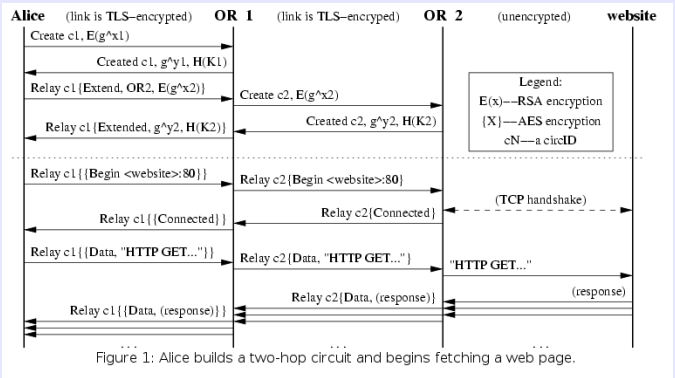


Figure 1: Alice builds a two-hop circuit and begins fetching a web page.

compare onion messages to SURBs

walk through building a 2-hop circuit

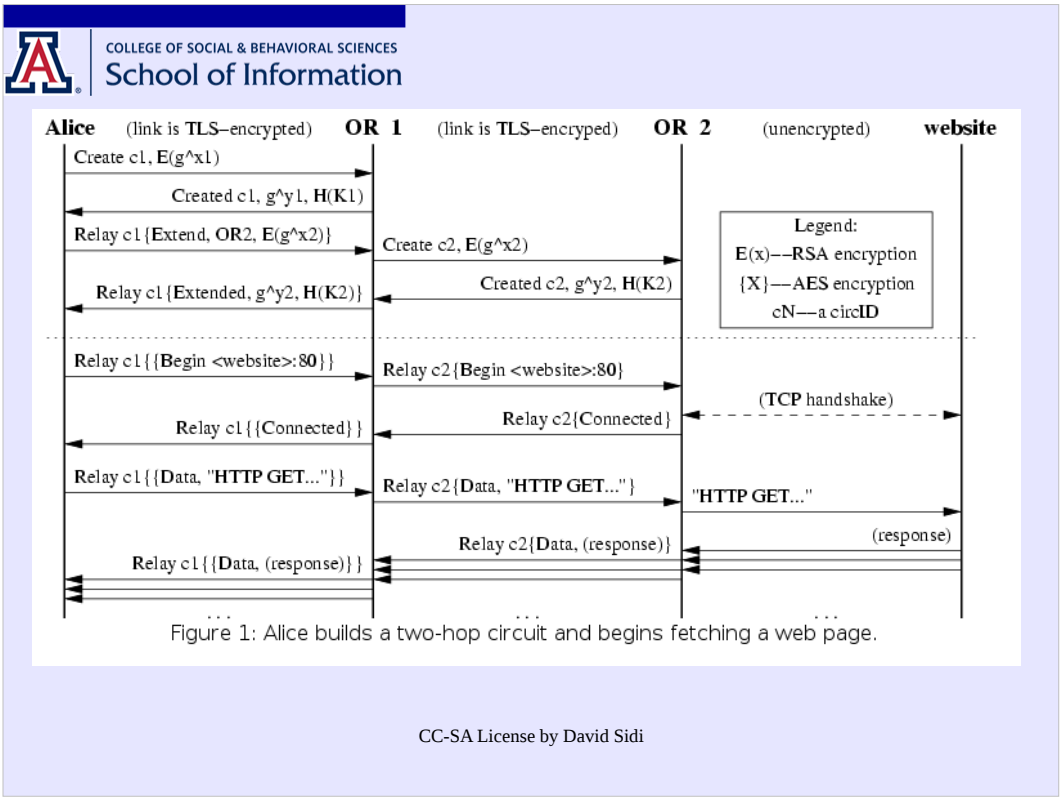


Figure 1: Alice builds a two-hop circuit and begins fetching a web page.

the rest will be next time