COLLEGE OF SOCIAL & BEHAVIORAL SCIENCES
School of Information

End to Database Privacy; Start to Communication Anonymity

Information Privacy with Applications

David Sidi (dsidi@email.arizona.edu)

Administrative

- Integrated session assignment

Alternative 1 to the Approach in the Assignment: Subsampling

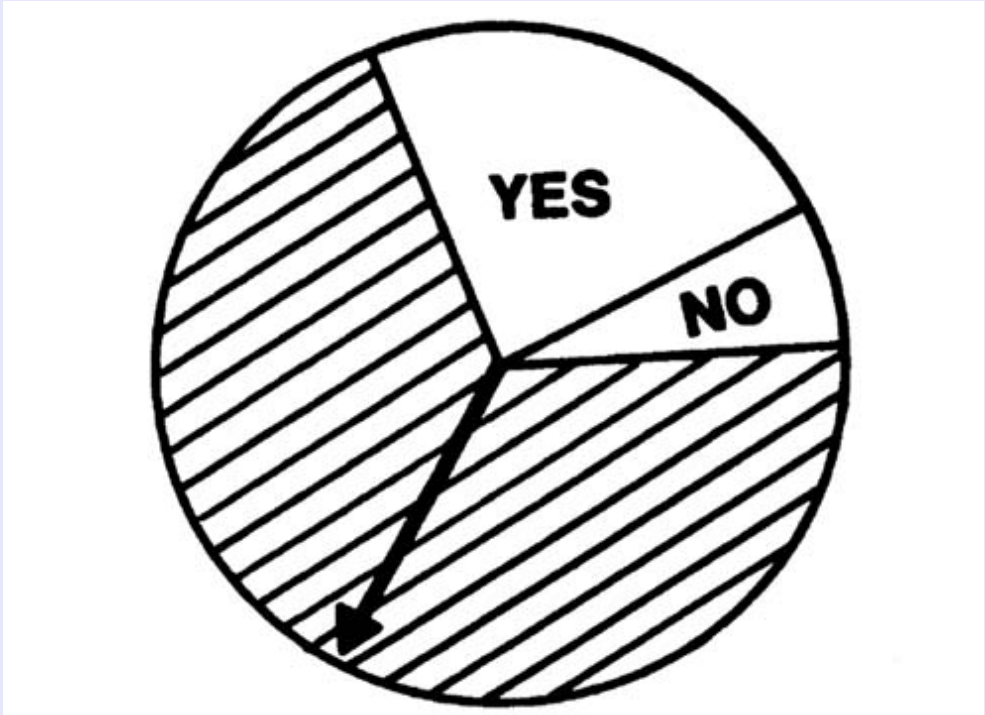
- Generate a representative subsample of the dataset, and compute statistics on that
- If it's small, then every respondent is unlikely to appear in the sample
- **Problem**: Some unlucky set of people are, against all probability, in the sample. So problem of protecting the privacy of their data recurs
- Are those in the subsample better off than those who would have been in the full dataset, though?
  - Does *plausible deniability* help? It depends (Rocher et al. 2019; Sidi and Bambauer 2020)

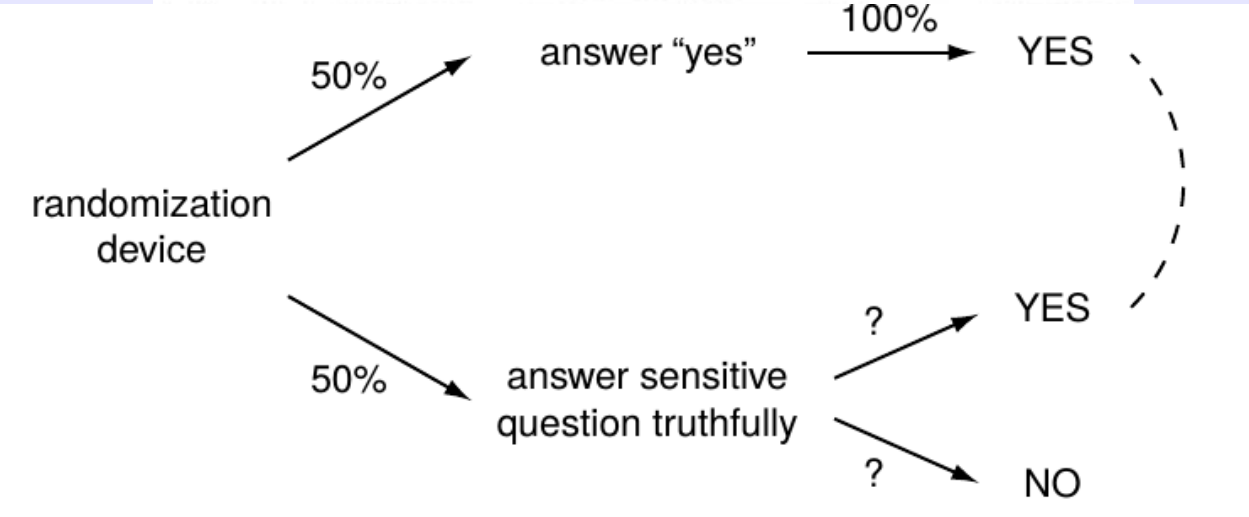
“[our] results reject [that] ... sampling or releasing partial datasets provide plausible deniability.” (Rocher et al. 2019)

- **Problem**: “The individual likelihood estimation of uniqueness is a good measure of plausible deniability only under the assumption that the intruder has auxiliary information about all of the variables used to render the data subject unique.” Modeling the case where an attacker has identified records with all variables, but wrong or missing values, can be done with a simulation-based approach. (Sidi and Bambauer 2020)
- “Despite ... exemplary work, it has taken several years to fully appreciate the importance of taking auxiliary information into account in privacy-preserving data release.” (Dwork 2011)

Alternative 2 to the Approach in the Assignment: Input Perturbation

(These are from different sources; the numbers don't line up)





- Consider how the following ways that a Randomized Response Technique might be set up affect (1) reduction in privacy risk; and (2) utility of the data:
  - The size of the “tell the truth” region in the spinner (more generally, the probability that you’ll have to give a truthful answer in the protocol)
  - The mode of administration (online in a locked down browser, online in your own browser, on a computer in a lab setting, in person over video conference, in person IRL)
  - Whether the protocol for randomizing is a physical device, or a program you run locally on your computer, or on a remote server.

Alternative 3 to the Approach in the Assignment: Output perturbation

- Interactive: curator of the data is involved in answering the query. “Did you ask too many disclosive questions relative to the current query? Then it is denied.”
- Noninteractive: Fire and forget. Provably cannot be done

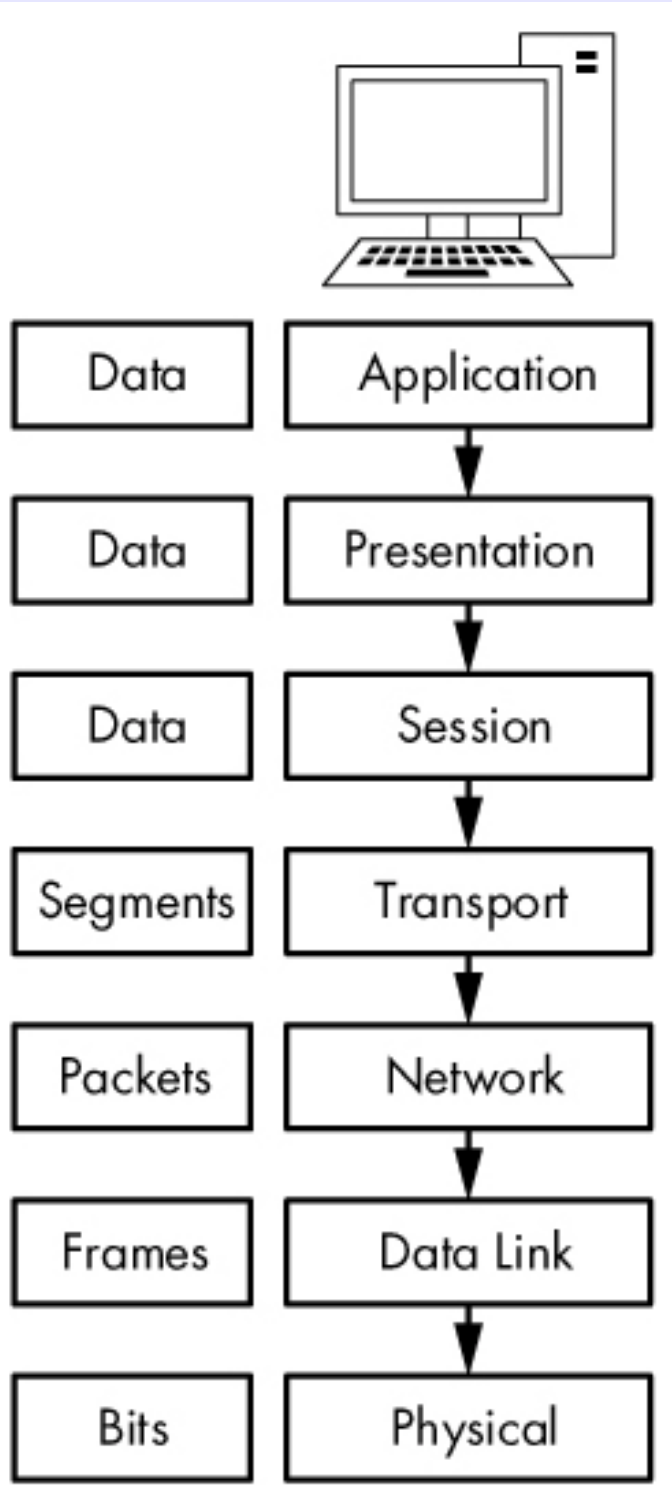
Big picture

- What is “semantic security?” Why is database privacy in Dwork (2011) understood differently
  - p.90
- Should “Dalenius’s Desideratum” be preserved?

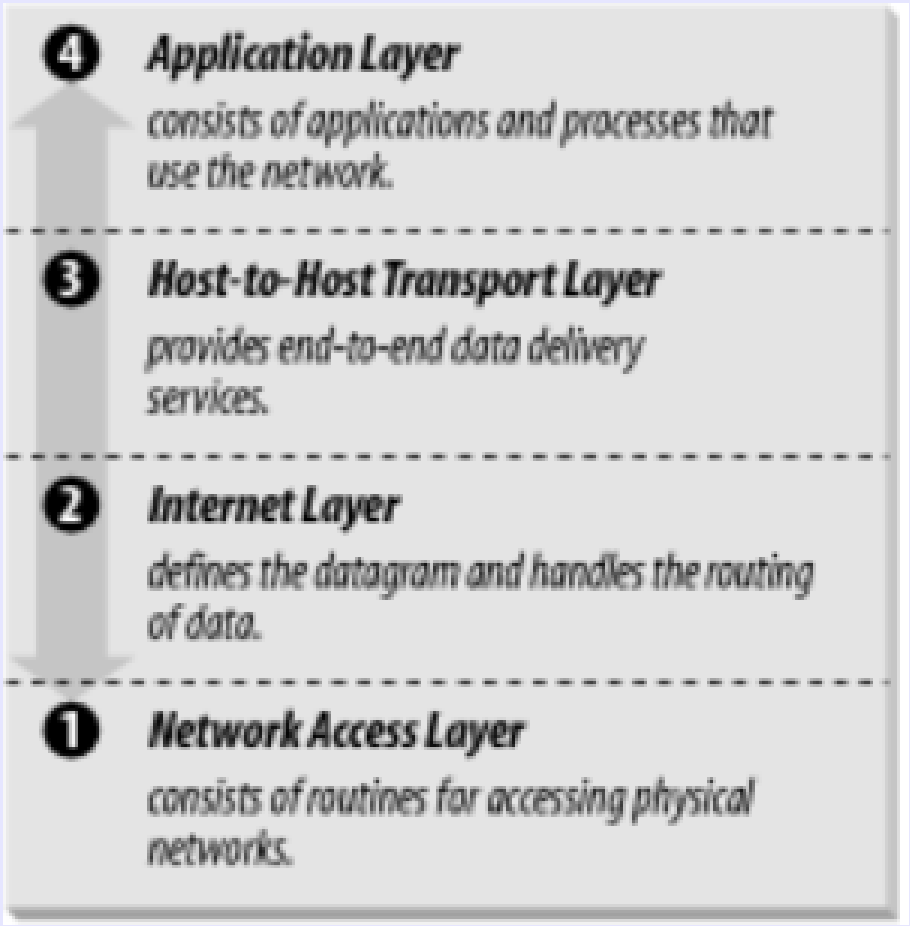
Anonymity

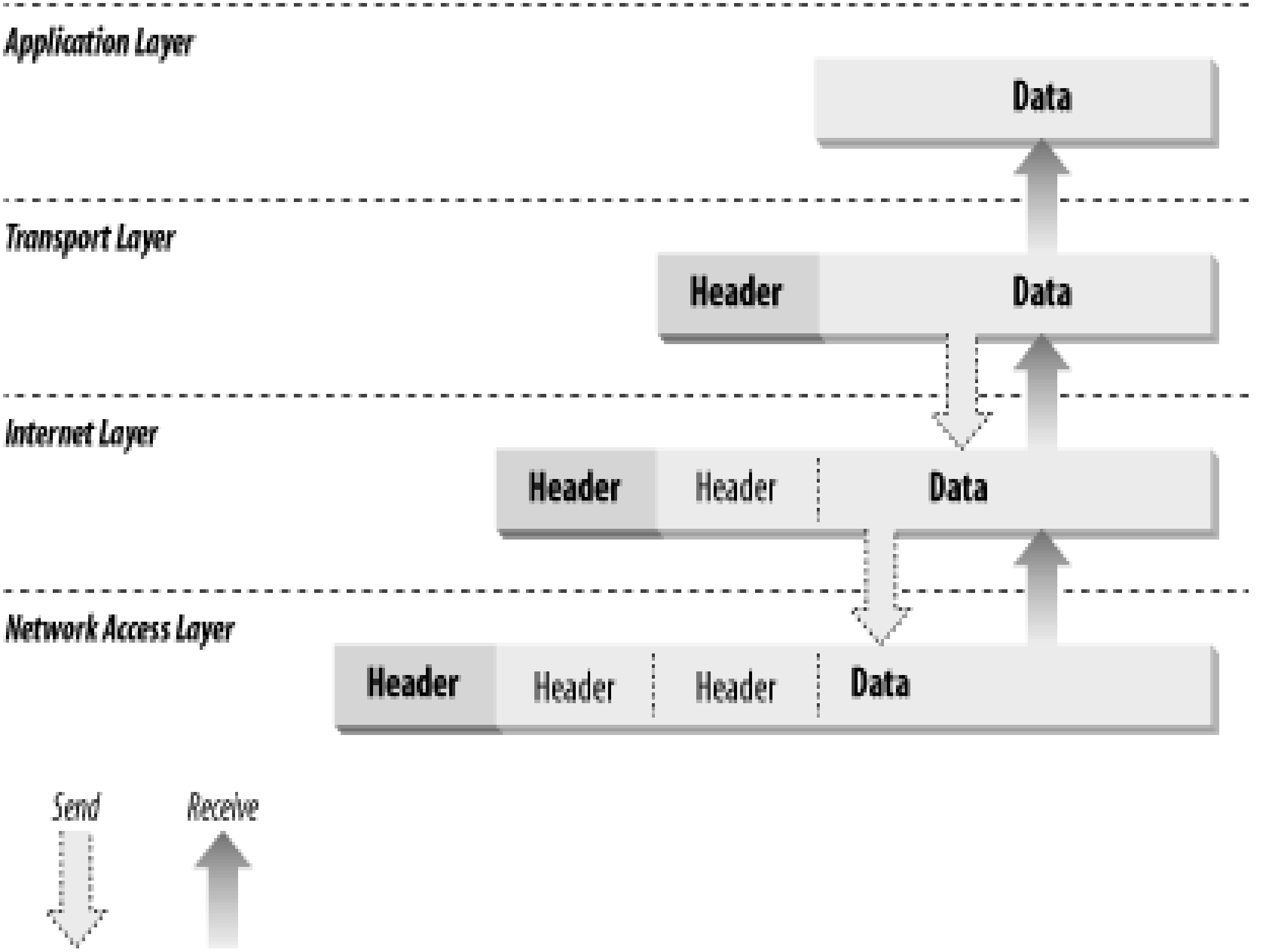
The Seven-Layer OSI Reference Model

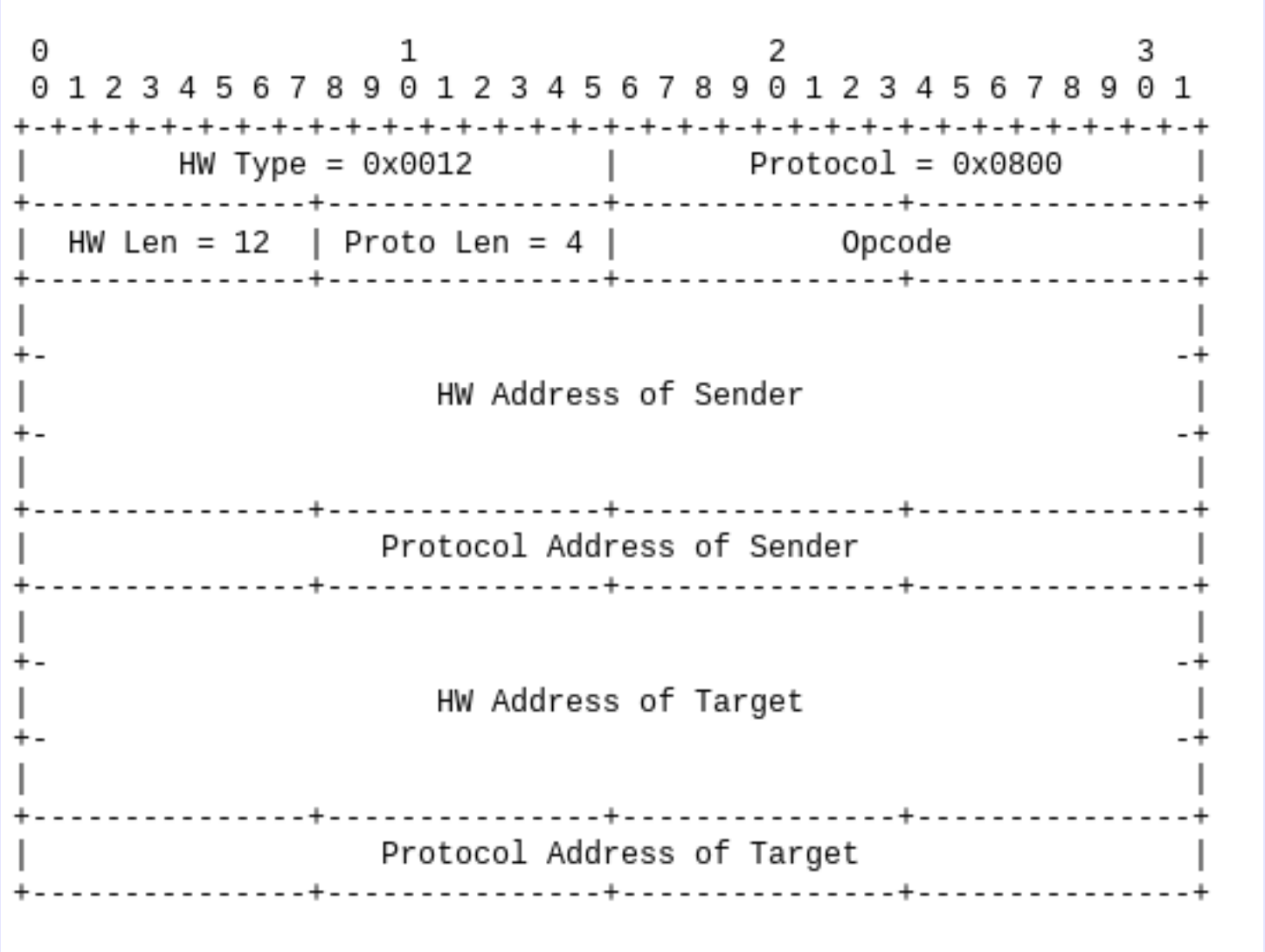
“the way up is the way down.” -Heraclitus

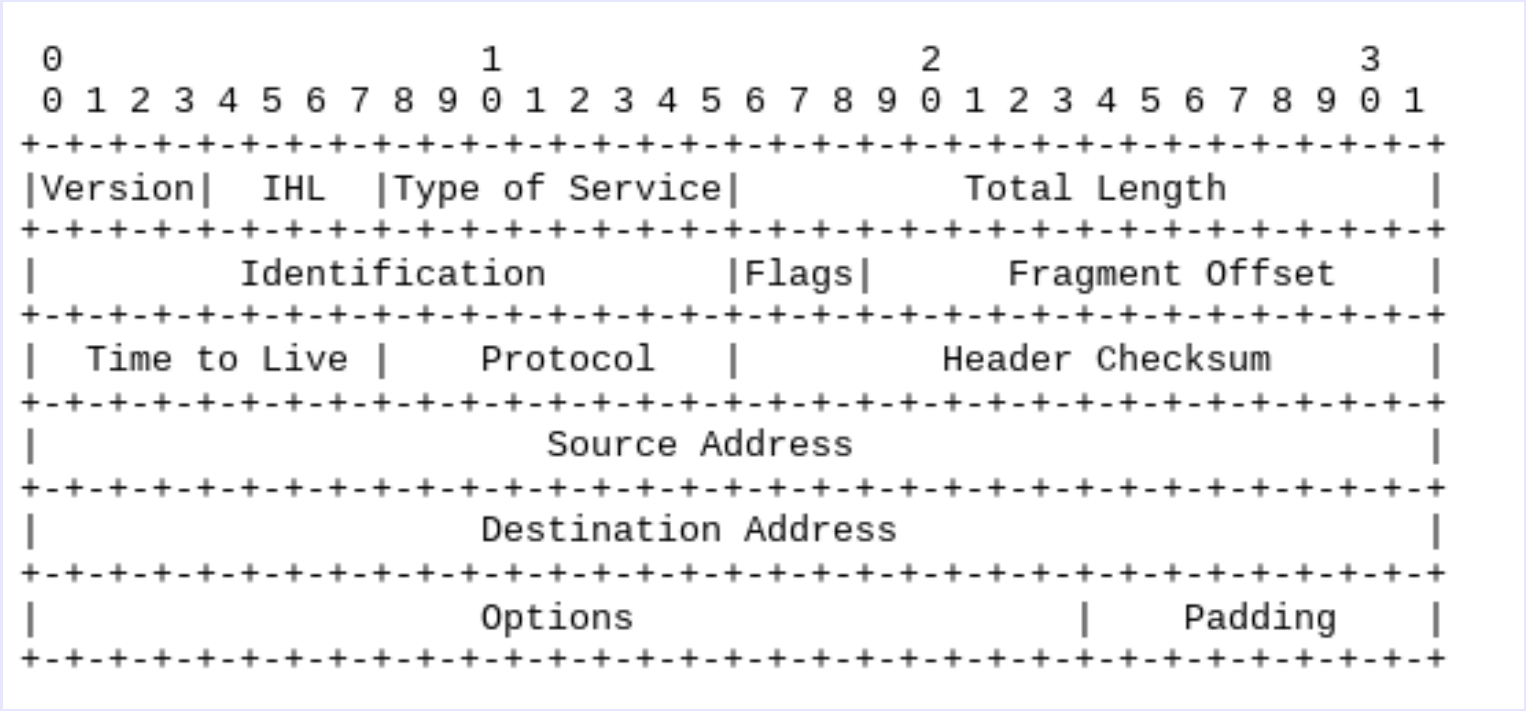


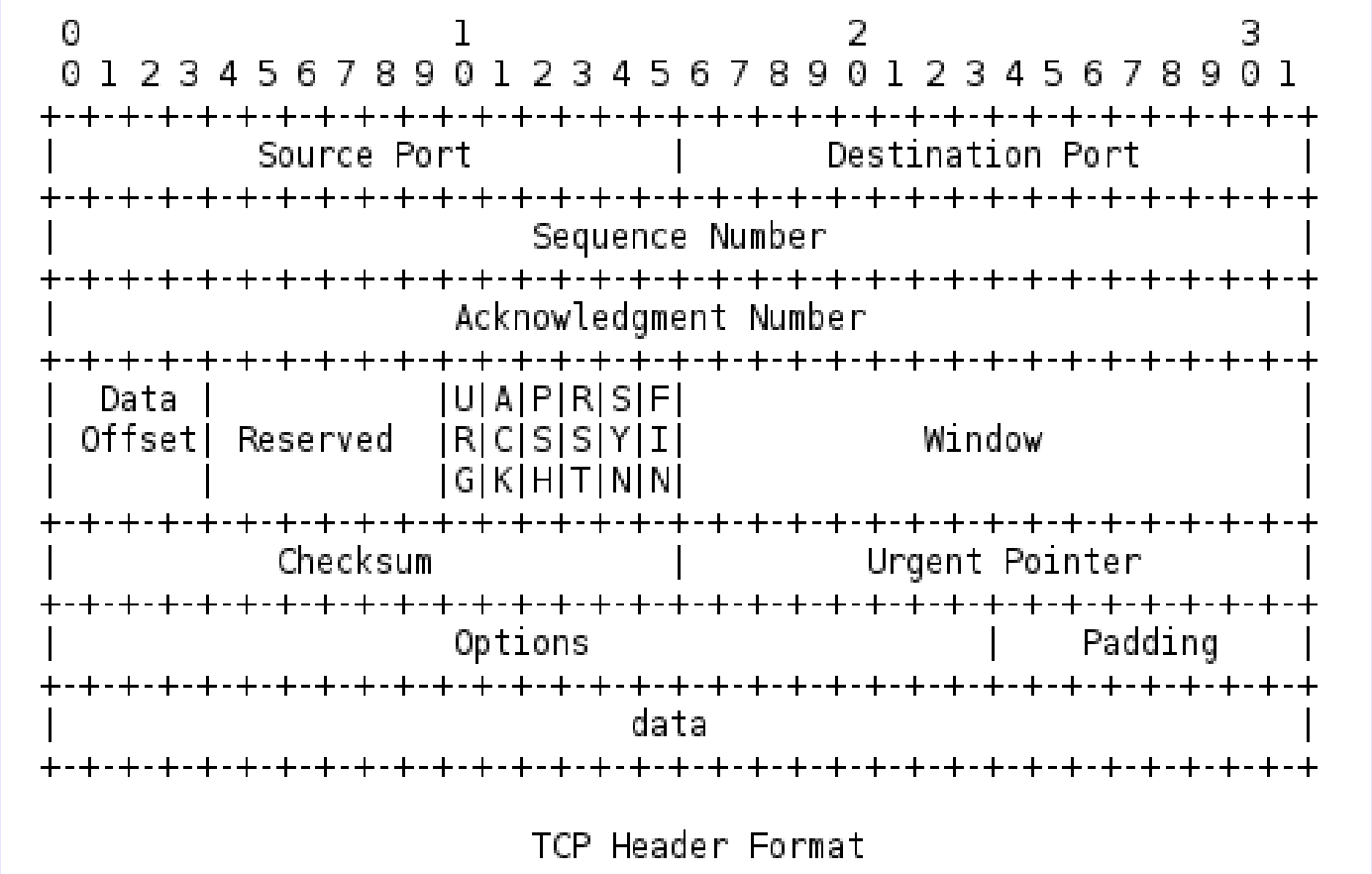
TCP/IP Protocol Stack




Application Layer
Data
Transport Layer
Header
Data
Internet Layer
Header
Header
Data
Network Access Layer
Header
Header
Header
Data
Send
Receive


0 1 2 3
0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
HW Type = 0x0012
Protocol = 0x0800
HW Len = 12
Proto Len = 4
Opcode
HW Address of Sender
Protocol Address of Sender
HW Address of Target
Protocol Address of Target


0 1 2 3
0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
Version
IHL
Type of Service
Total Length
Identification
Flags
Fragment Offset
Time to Live
Protocol
Header Checksum
Source Address
Destination Address
Options
Padding



TCP Header Format

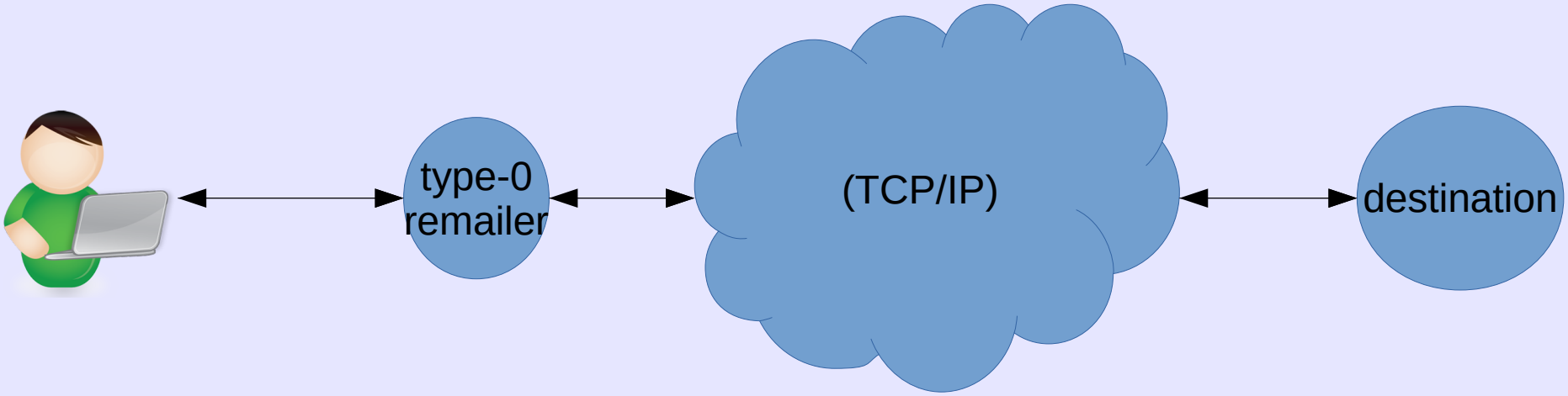
```
# Network services, Internet style
#
# Note that it is presently the policy of IANA to assign a single well-known
# port number for both TCP and UDP; hence, officially ports have two entries
# even if the protocol doesn't support UDP operations.
#
# Updated from http://www.iana.org/assignments/port-numbers and other
# sources like http://www.freebsd.org/cgi/cvsweb.cgi/src/etc/services .
# New ports will be added on request if they have been officially assigned
# by IANA and used in the real-world or are needed by a debian package.
# If you need a huge list of used numbers please install the nmap package.

tcpmux          1/tcp                           # TCP port service multiplexer
echo            7/tcp
echo            7/udp
discard         9/tcp           sink null
discard         9/udp           sink null
systat          11/tcp          users
daytime         13/tcp
daytime         13/udp
netstat         15/tcp
qotd            17/tcp          quote
msp             18/tcp                          # message send protocol
msp             18/udp
chargen         19/tcp          ttytst source
chargen         19/udp          ttytst source
ftp-data        20/tcp
ftp             21/tcp
fsp             21/udp          fspd
ssh             22/tcp                          # SSH Remote Login Protocol
ssh             22/udp
telnet          23/tcp
smtp            25/tcp          mail
time            37/tcp          timserver
time            37/udp          timserver
rlp             39/udp          resource        # resource location
nameserver      42/tcp          name            # IEN 116
whois           43/tcp          nicname
tacacs          49/tcp                          # Login Host Protocol (TACACS)
tacacs          49/udp
re-mail-ck      50/tcp                          # Remote Mail Checking Protocol
re-mail-ck      50/udp
domain          53/tcp                          # Domain Name Server
domain          53/udp
mtp             57/tcp                          # deprecated
tacacs-ds       65/tcp                          # TACACS-Database Service
tacacs-ds       65/udp
bootps          67/tcp                          # BOOTP server
bootps          67/udp
bootpc          68/tcp                          # BOOTP client
bootpc          68/udp
tftp            69/udp
gopher          70/tcp                          # Internet Gopher
```

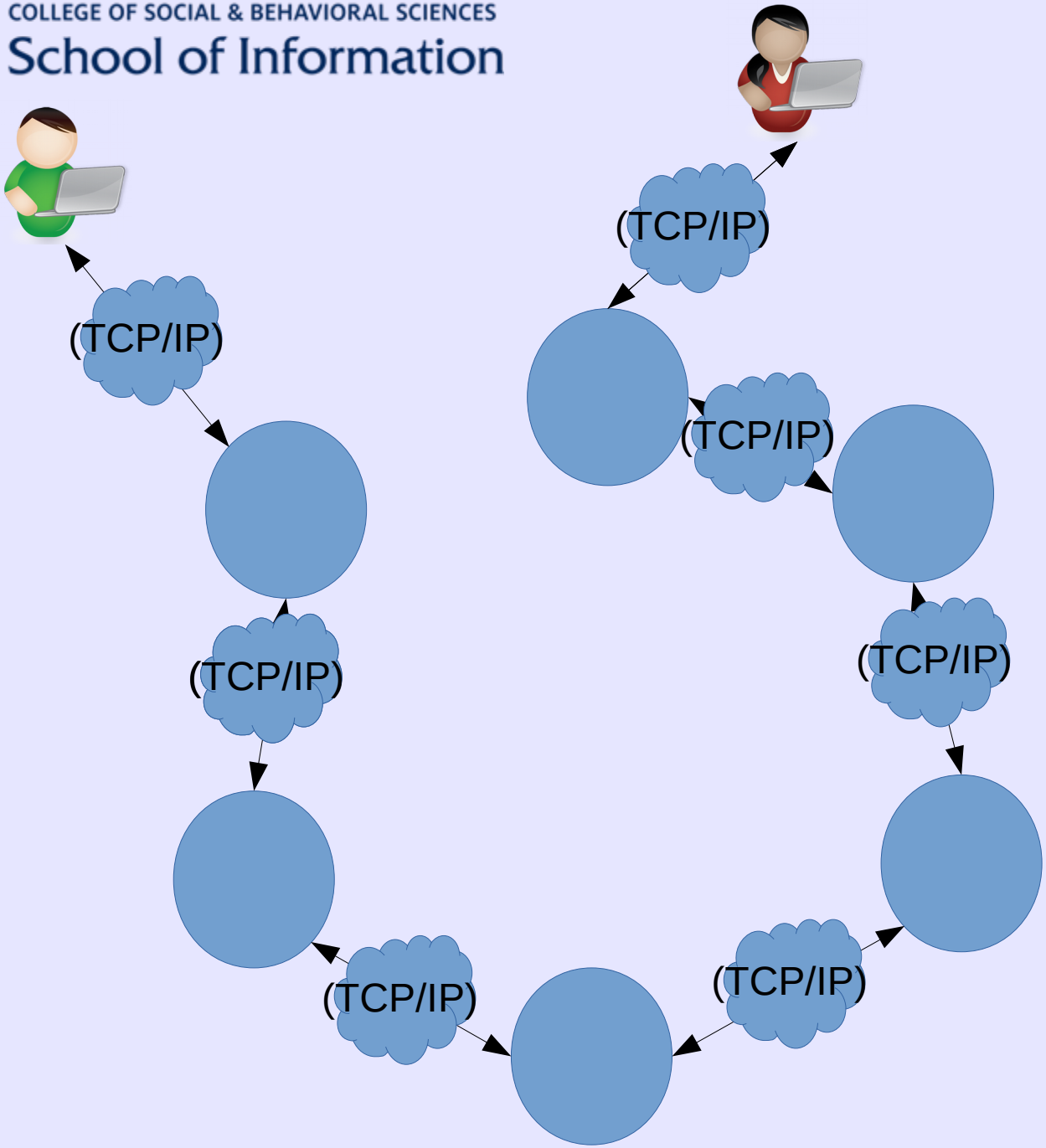
Chaum 1981:

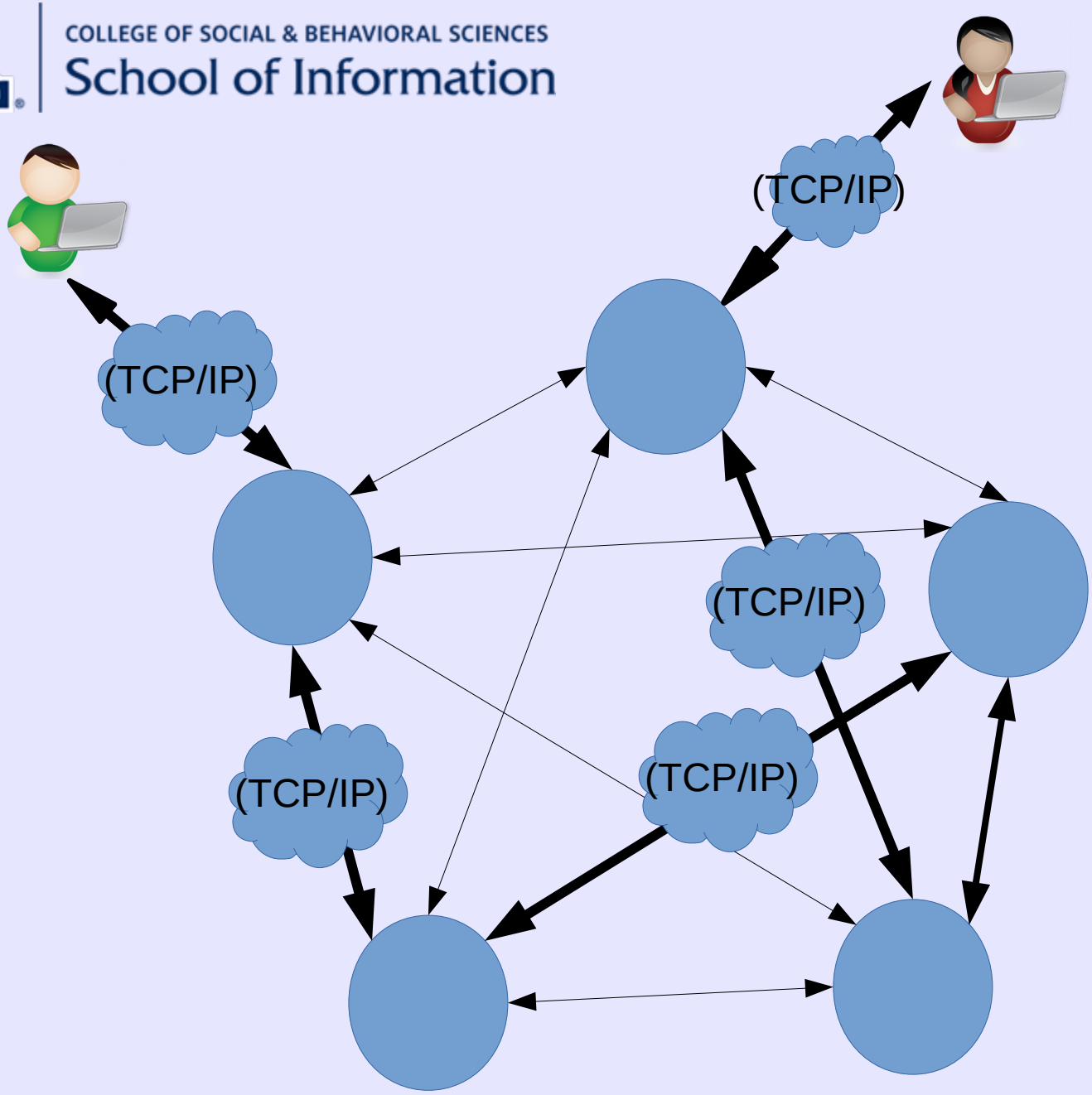
process each item of mail before it is delivered. A participant prepares a message M for delivery to a participant at address A by sealing it with the addressee's public key K_a, appending the address A, and then sealing the result with the mix's public key K_1. The left-hand side of the following expression denotes this item which is input to the mix:

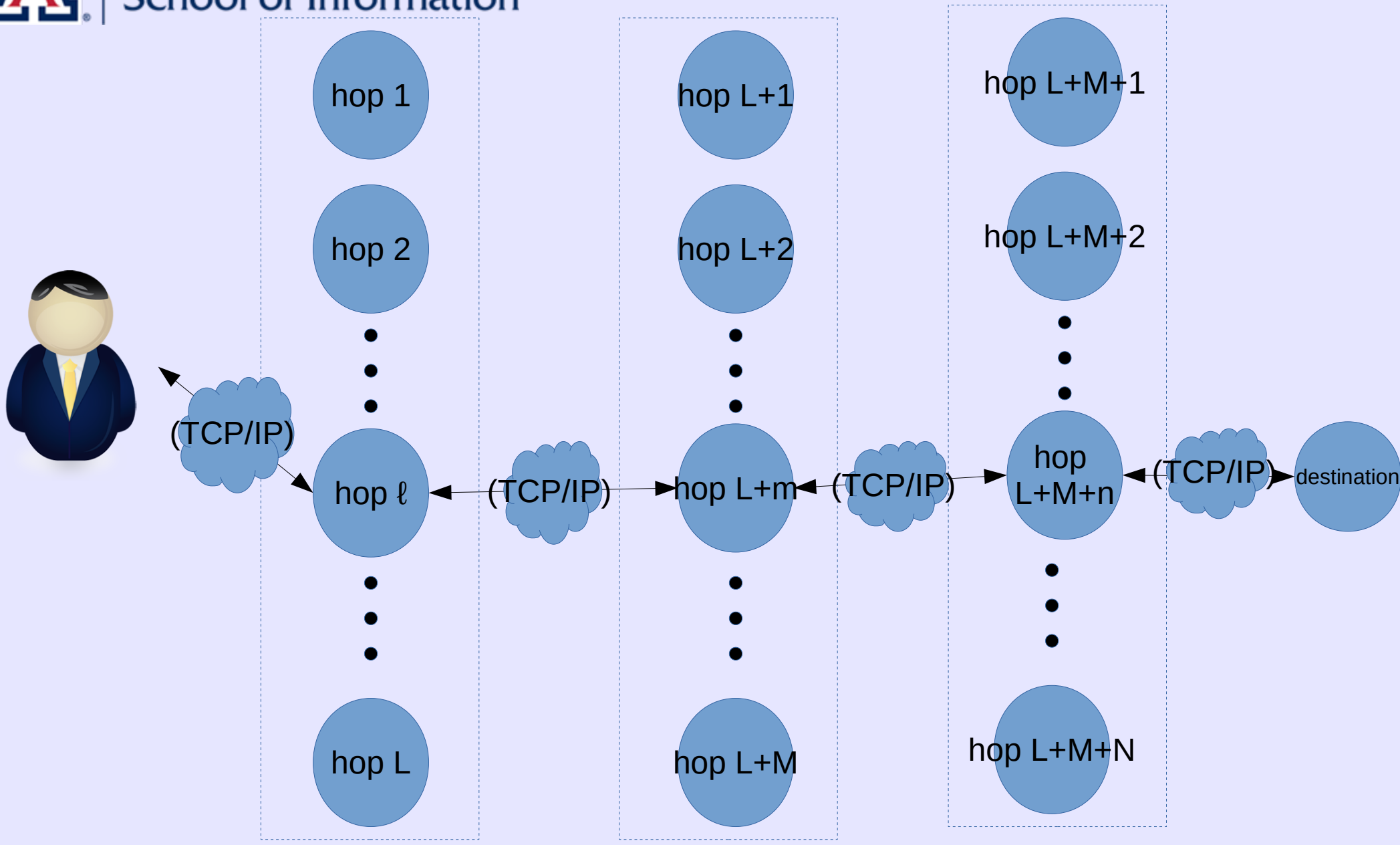
$$K_1(R_1, K_a(R_0, M), A) \rightarrow K_a(R_0, M), A.$$

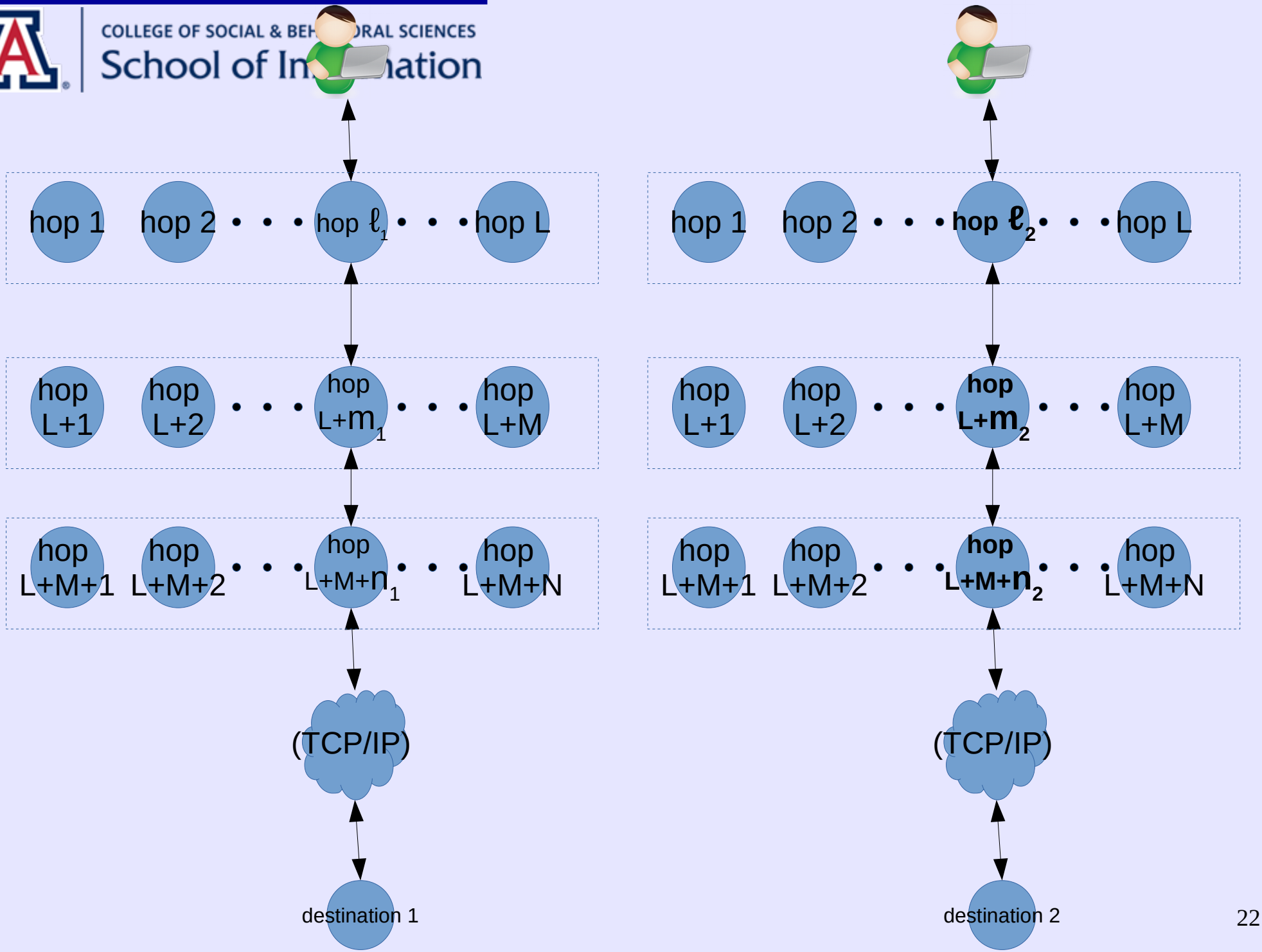
How does the message get to A?
One answer: overlay network


type-0
remailer
(TCP/IP)
destination


(TCP/IP)
(TCP/IP)
(TCP/IP)
(TCP/IP)
(TCP/IP)
(TCP/IP)
(TCP/IP)


(TCP/IP)
(TCP/IP)
(TCP/IP)
(TCP/IP)
(TCP/IP)


hop 1
hop 2
hop ℓ
hop L
hop L+1
hop L+2
hop L+m
hop L+M
hop L+M+1
hop L+M+2
hop L+M+n
hop L+M+N
(TCP/IP)
(TCP/IP)
(TCP/IP)
(TCP/IP)
destination


hop 1
hop 2
hop ℓ_1
hop L
hop L+1
hop L+2
hop L+m_1
hop L+M
hop L+M+1
hop L+M+2
hop L+M+n_1
hop L+M+N
(TCP/IP)
destination 1
hop 1
hop 2
hop ℓ_2
hop L
hop L+1
hop L+2
hop L+m_2
hop L+M
hop L+M+1
hop L+M+2
hop L+M+n_2
hop L+M+N
(TCP/IP)
destination 2

Anonymity

COLLEGE OF SOCIAL & BEHAVIORAL SCIENCES
School of Information

Terminology Review

Anonymity set

- `Anonymity' is defined with respect to a subset of the possible senders, called the anonymity set.
- Think of it as answering "who might you be?"

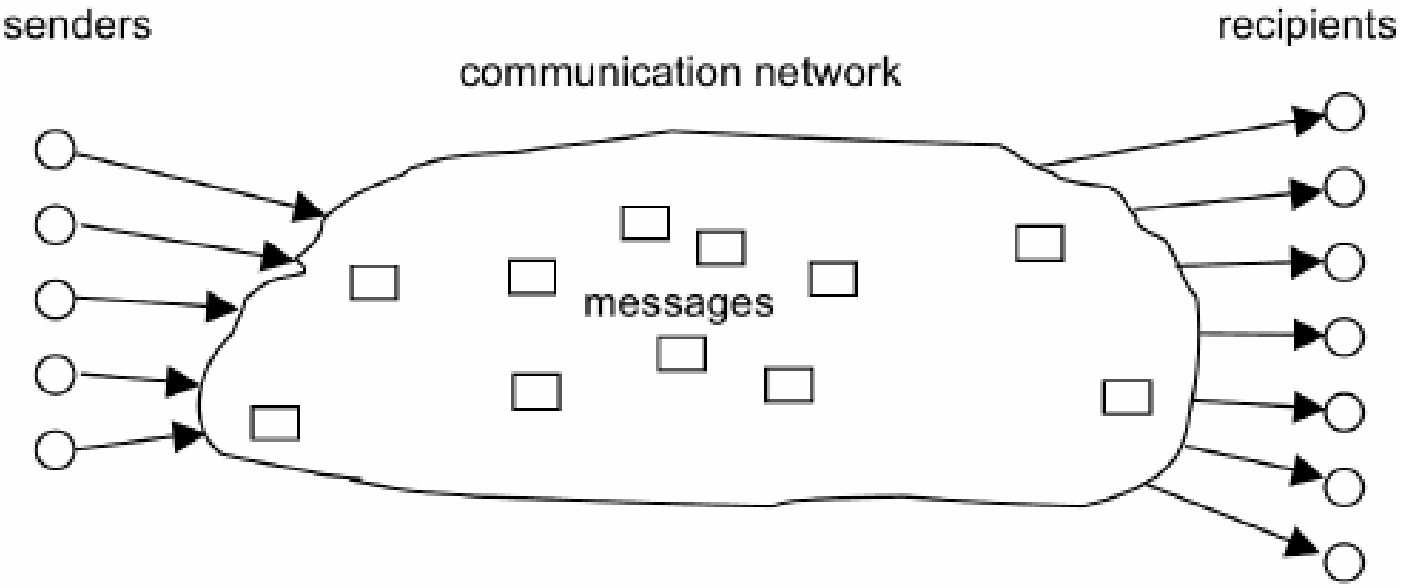
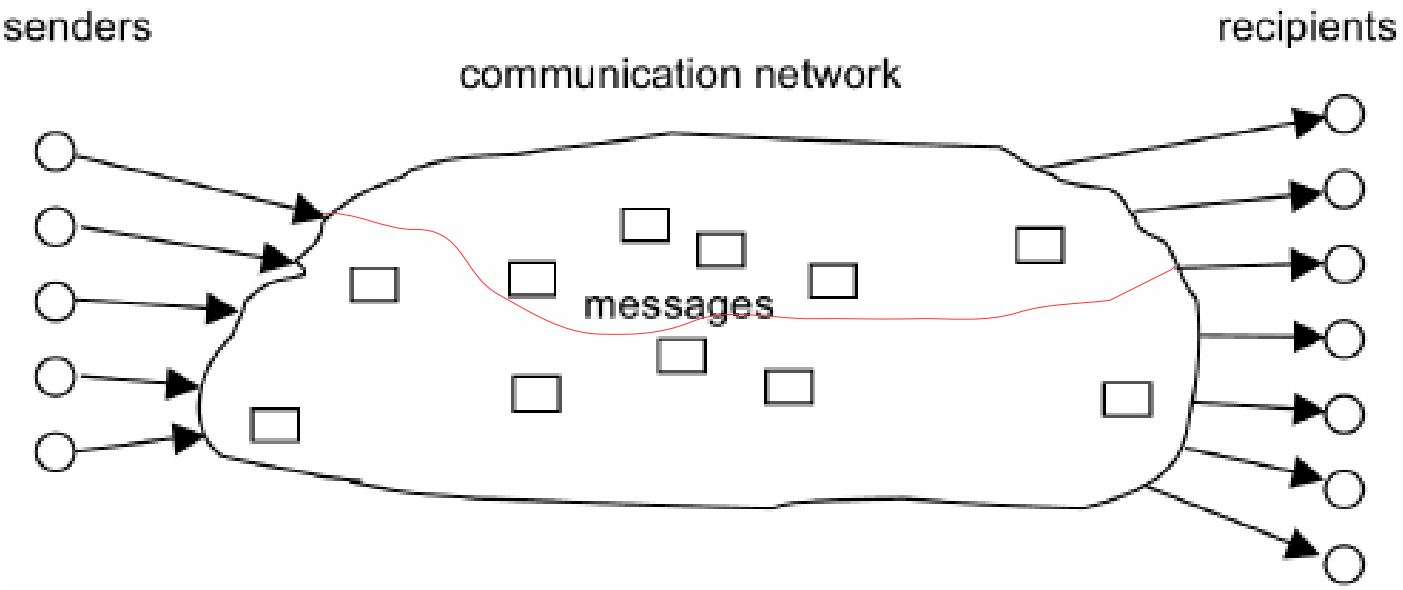
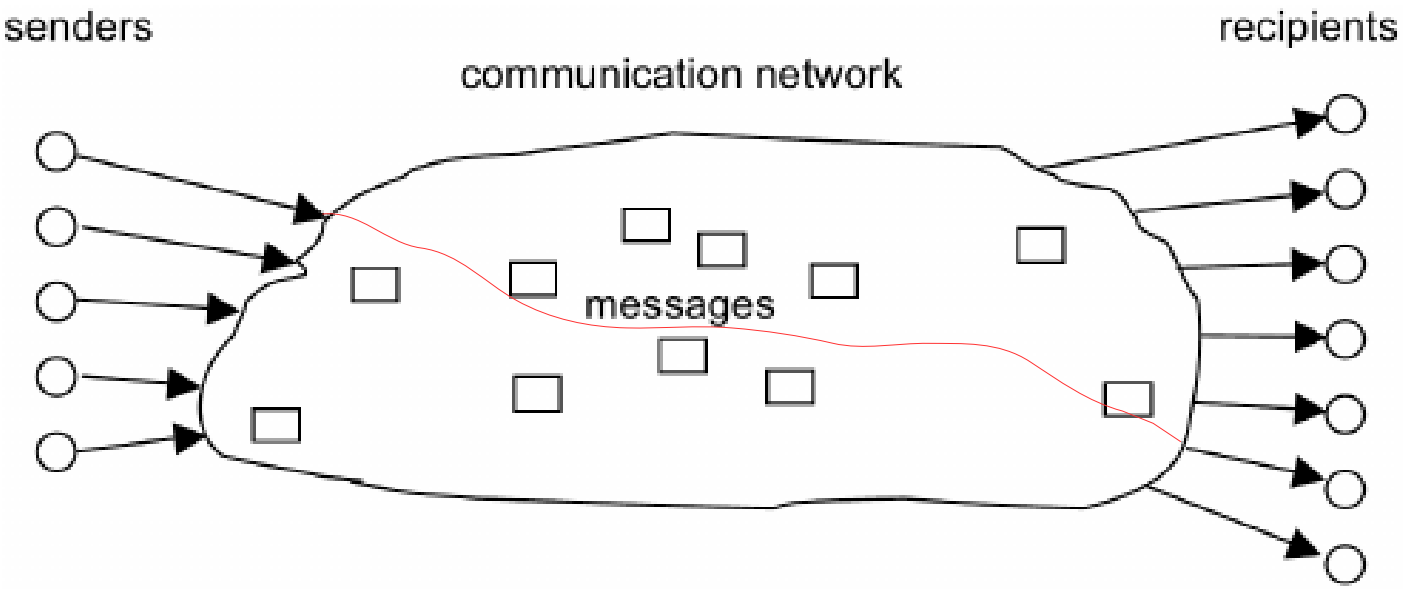
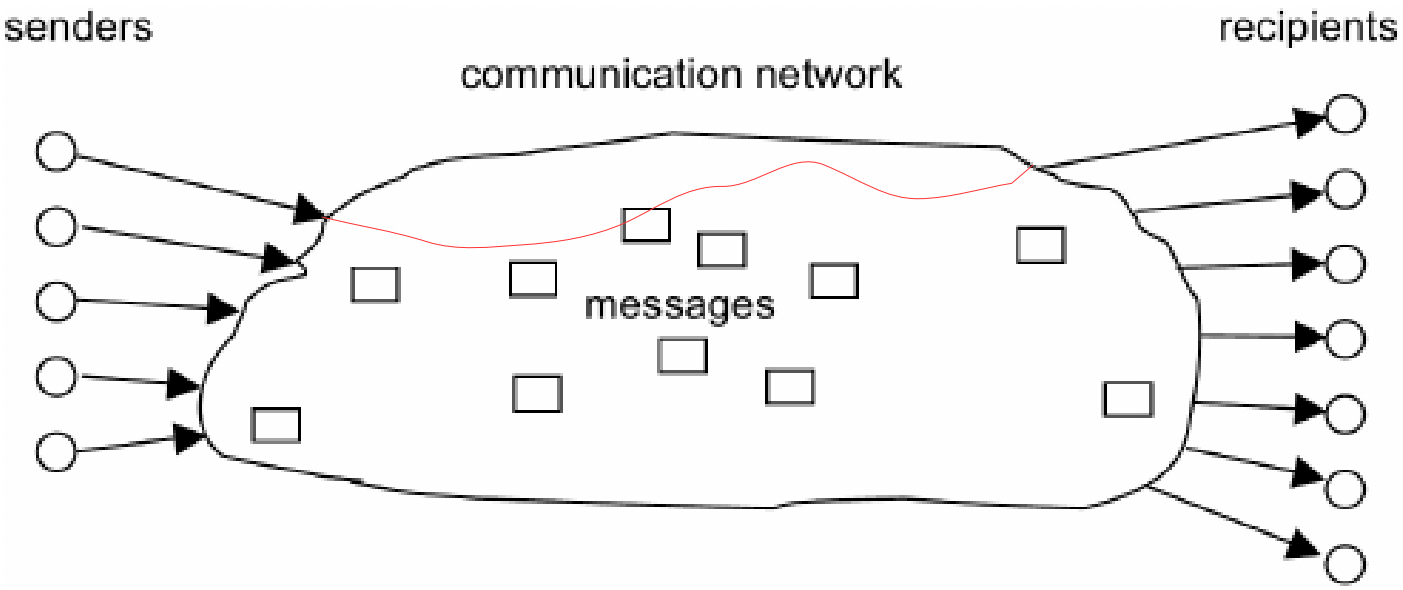
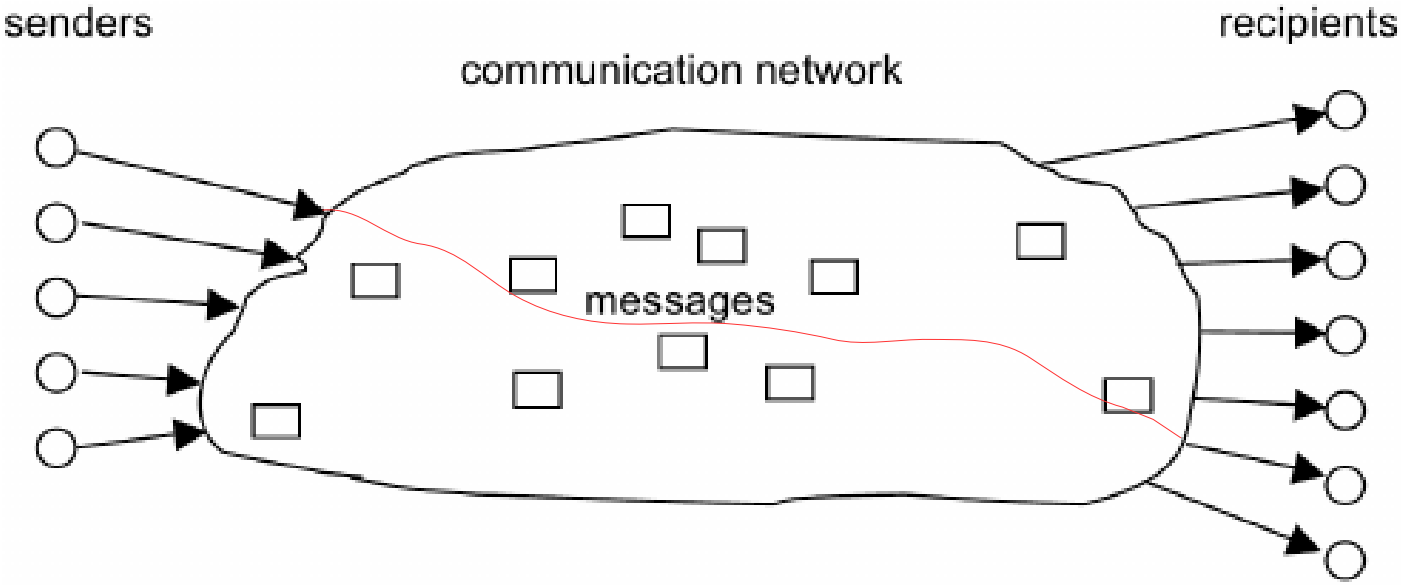


Image credit (before modification):
Christina Pöpper
Ruhr-University Bochum


senders
communication network
recipients
messages

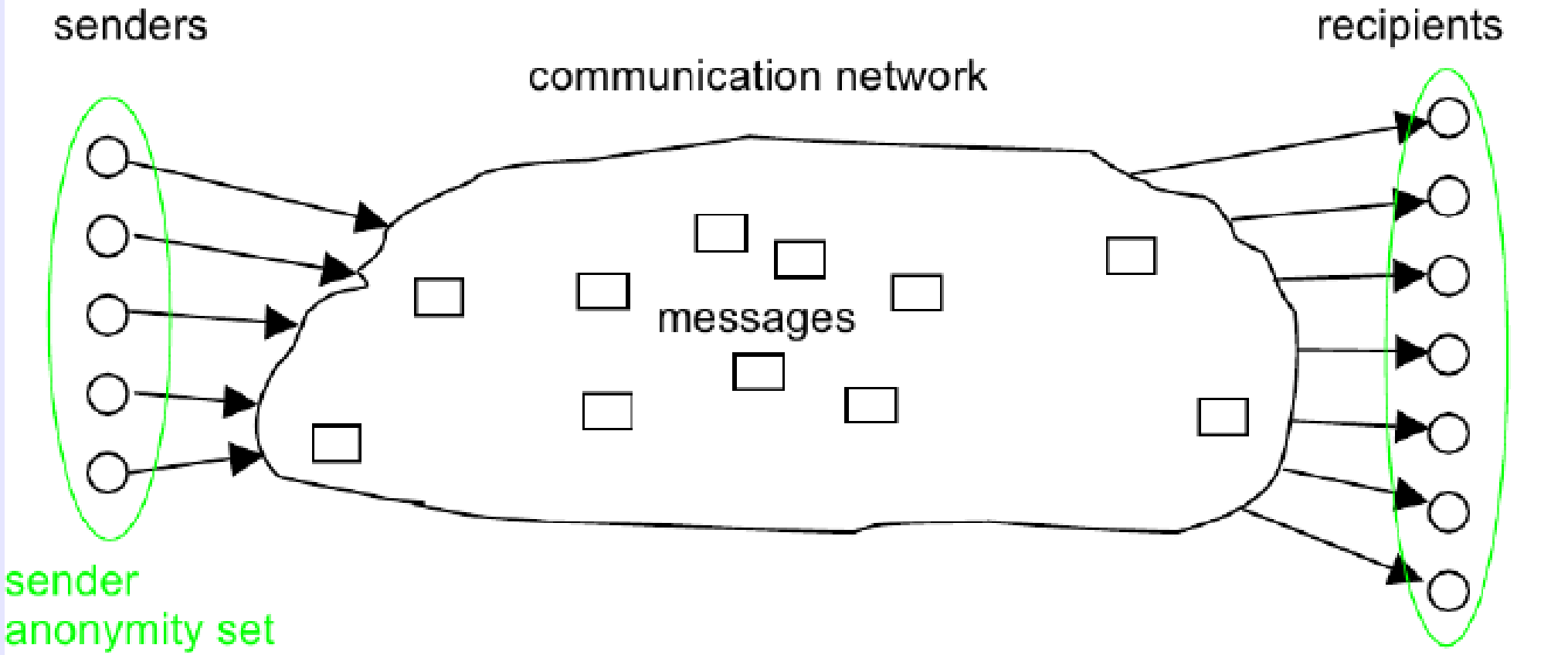

senders
communication network
recipients
messages


senders
recipients
communication network
messages



Anonymity set

- Can you clearly describe the limiting cases for the anonymity set?



largest possible anonymity sets

Image credit Christina Pöpper
Ruhr-University Bochum

	MAC	Browser_fingerprint	IP	Sites_visited
SNDER_1	00:a0:ef:eb:5v:ff	af7f098c39728f8cb676e3df82ced01a149ee3aa92af2b88c20c4948a5fad5fd	172.16.1.5	torproject.org, ischool.arizona.edu, maps.google.com...
SNDER_2	00:c0:ff:dd:ff:ef	a5fad5fdd01a149eeaf7f098c39728f8cb676e3df82ce3aa92af2b88c20c4948	172.16.12.4	nytimes.com, purple.com