Ostra: Leveraging trust to thwart unwanted communication

Alan Mislove[†‡] Ansley Post[†‡] Peter Druschel[†] Krishna Gummadi[†]

[†]MPI-SWS
[‡]Rice University

NSDI 2008

Digital communication

Electronic systems provide low-cost communication

- Email
- VoIP
- Blogs
- IM
- Content-sharing

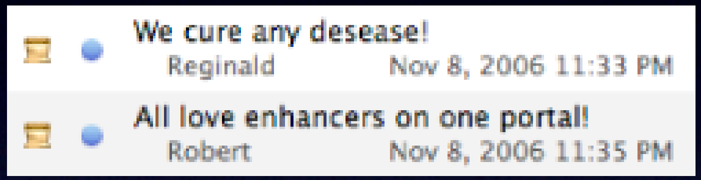
Democratized content publication

- Can make content available to (millions of) users

Unwanted communication

Low cost abused to send unwanted communication

- Spam
- Unwanted Skype invitations



Affecting content-sharing sites

- Mislabeled content on YouTube

Users are not accountable

- Banned users can create new identity

Previous approaches

- Filter based on content
  - Hard for rich media (videos, photos)

Previous approaches

Filter based on content
- Hard for rich media (videos, photos)

Charge money to send
- Requires micropayment infrastructure

Previous approaches

Filter based on content

- Hard for rich media (videos, photos)

Charge money to send

- Requires micropayment infrastructure

Introduce strong identities

- Resisted by users

Ostra

New approach to preventing unwanted communication

- Leverages an (existing) social network

Works in conjunction with existing communication system

- No content filtering
- No additional monetary cost
- No strong identities

Key idea: Exploit cost of maintaining social relationships

- Inspired by trust in offline world

Outline

Inspiration: Hawala

Ostra in detail

Evaluation

Related work

Conclusion

Inspiration: Hawala

Hawala

System for transferring money

- Originated in India, centuries old

Give money to a hawala dealer

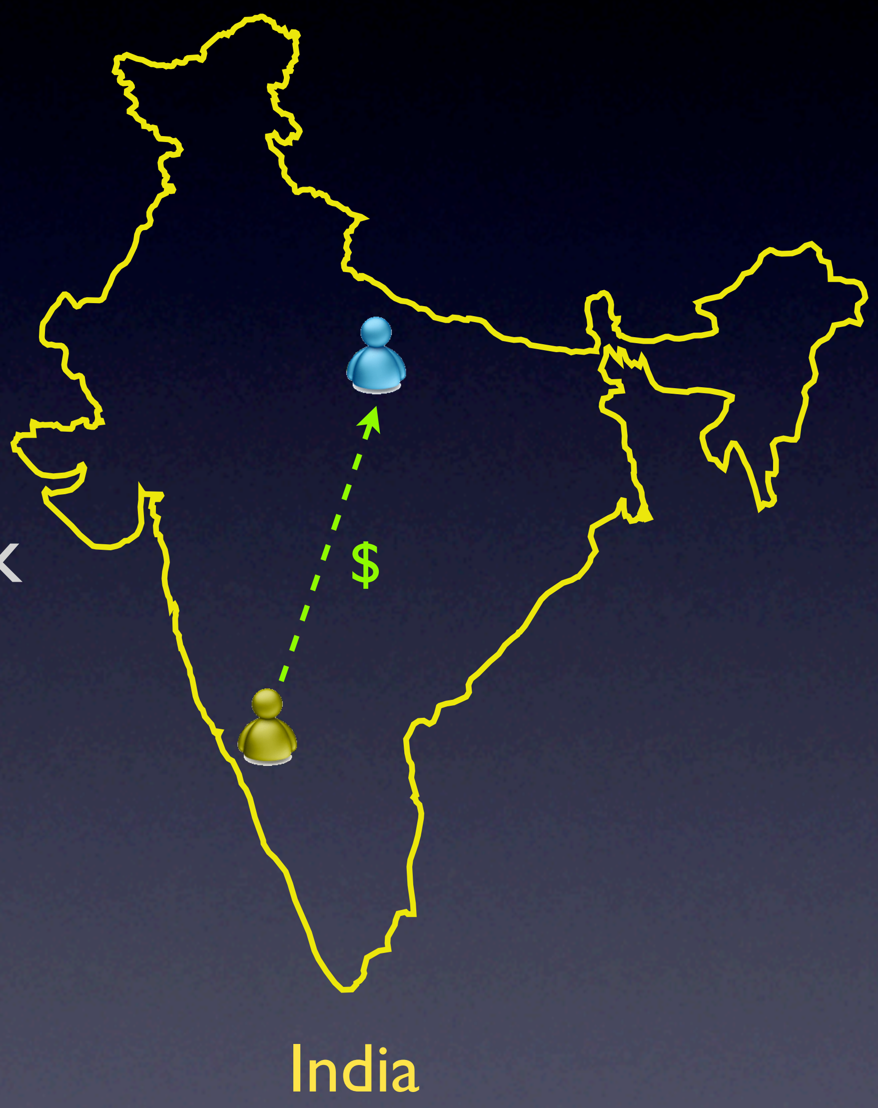
- Often someone you know already
- Transfered via hawala dealer social network

Hawala dealers only exchange notes

- Settle up in the future

Comparable to debt between banks

- But trust is only pairwise



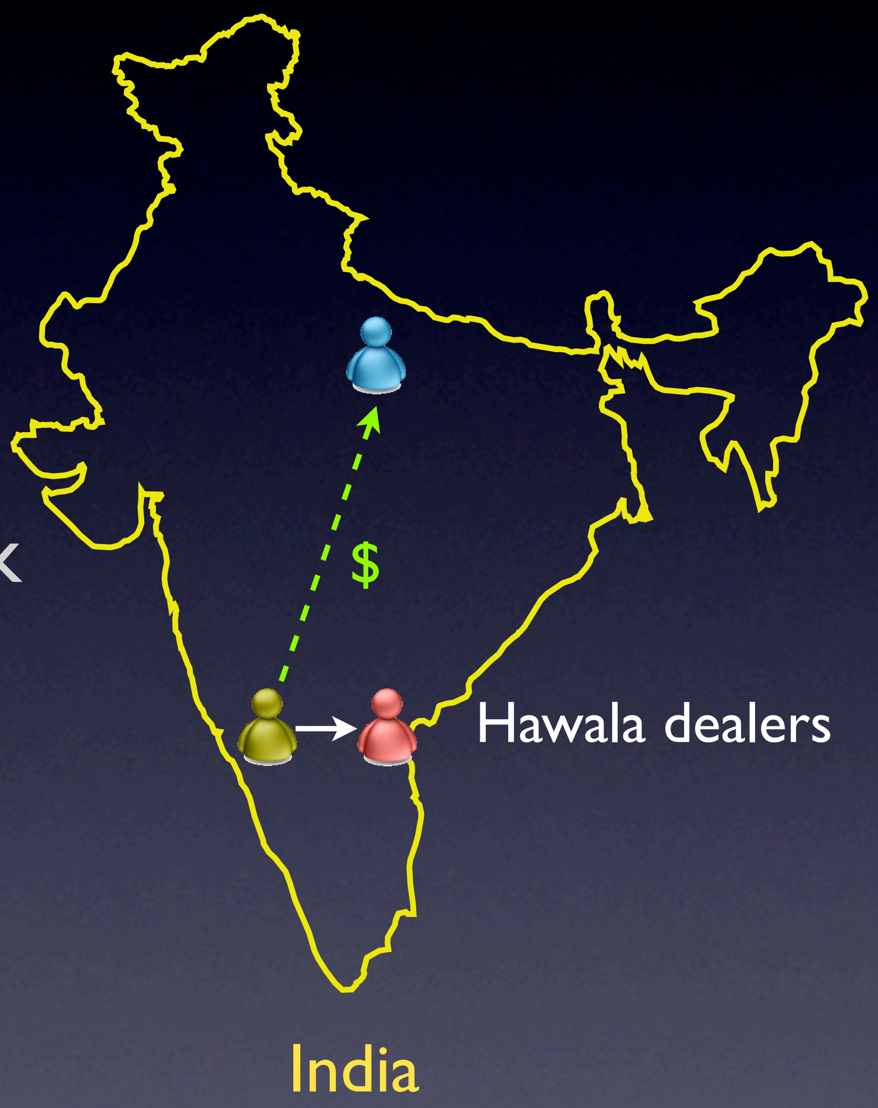
Hawala

System for transferring money
- Originated in India, centuries old

Give money to a hawala dealer
- Often someone you know already
- Transfered via hawala dealer social network

Hawala dealers only exchange notes
- Settle up in the future

Comparable to debt between banks
- But trust is only pairwise



Hawala

System for transferring money

- Originated in India, centuries old

Give money to a hawala dealer

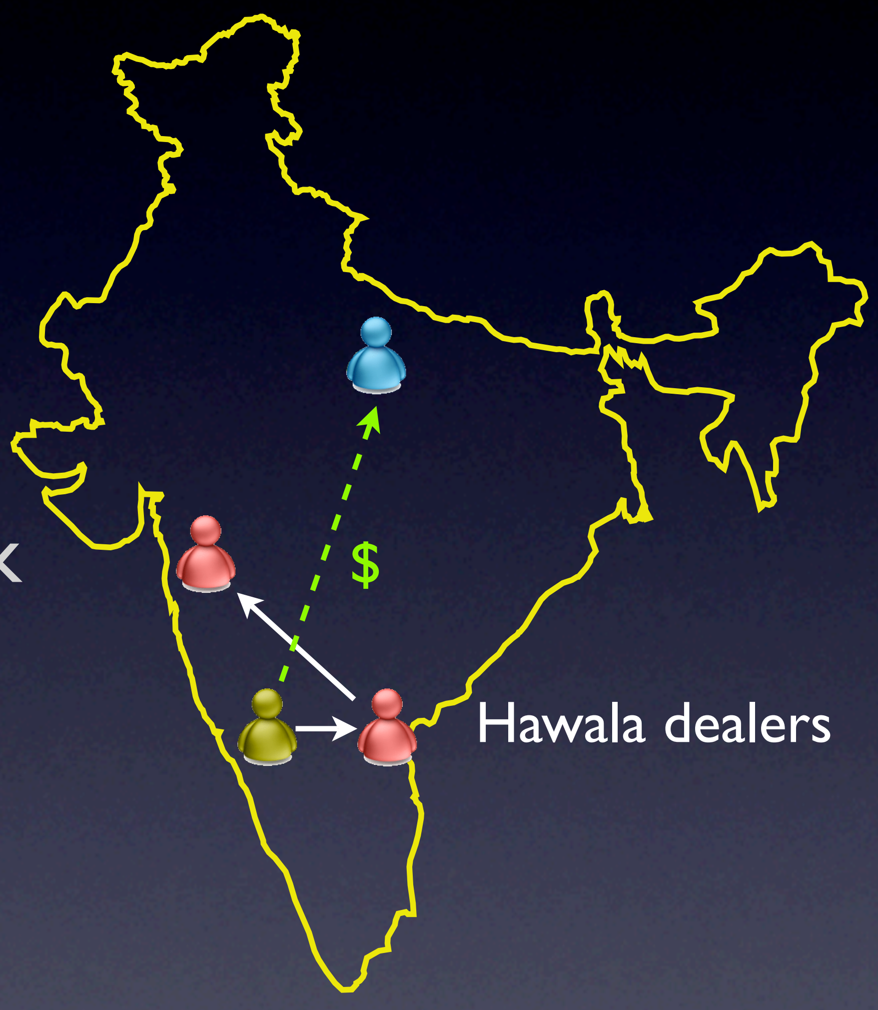
- Often someone you know already
- Transfered via hawala dealer social network

Hawala dealers only exchange notes

- Settle up in the future

Comparable to debt between banks

- But trust is only pairwise



Hawala

System for transferring money

- Originated in India, centuries old

Give money to a hawala dealer

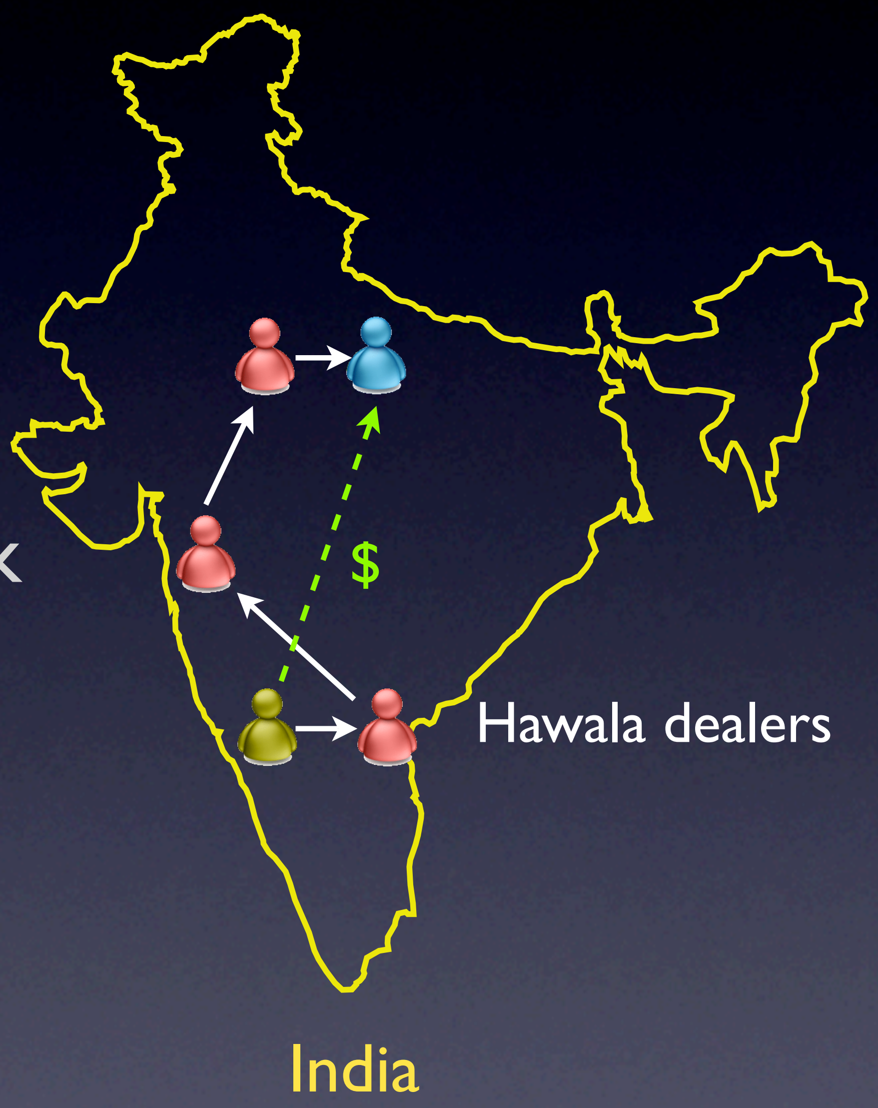
- Often someone you know already
- Transfered via hawala dealer social network

Hawala dealers only exchange notes

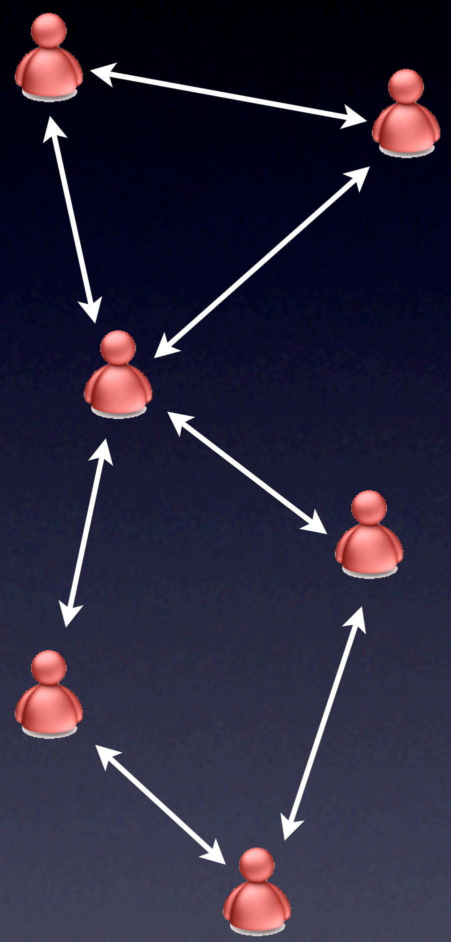
- Settle up in the future

Comparable to debt between banks

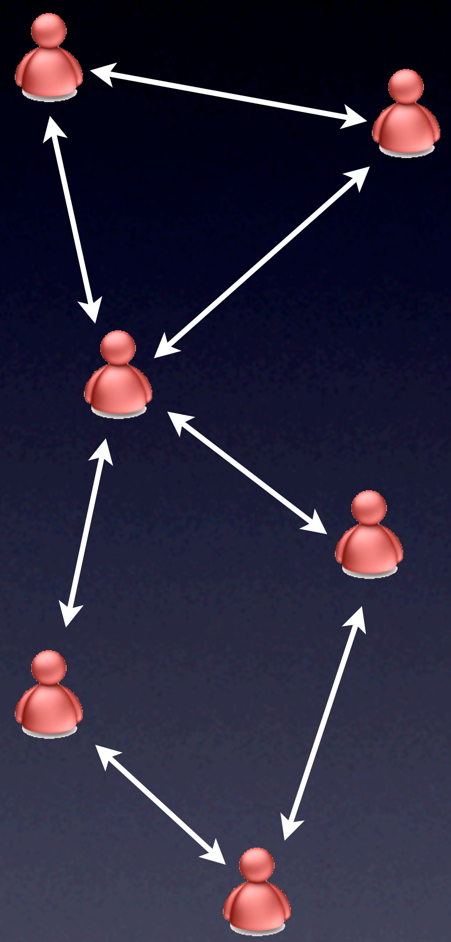
- But trust is only pairwise



Why does hawala work?



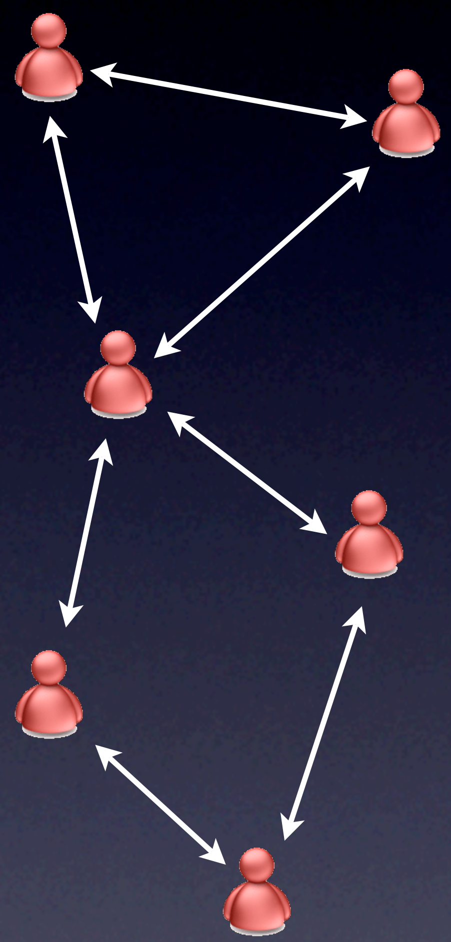
Why does hawala work?



Links take effort to form/maintain

Can't get new links easily

Why does hawala work?



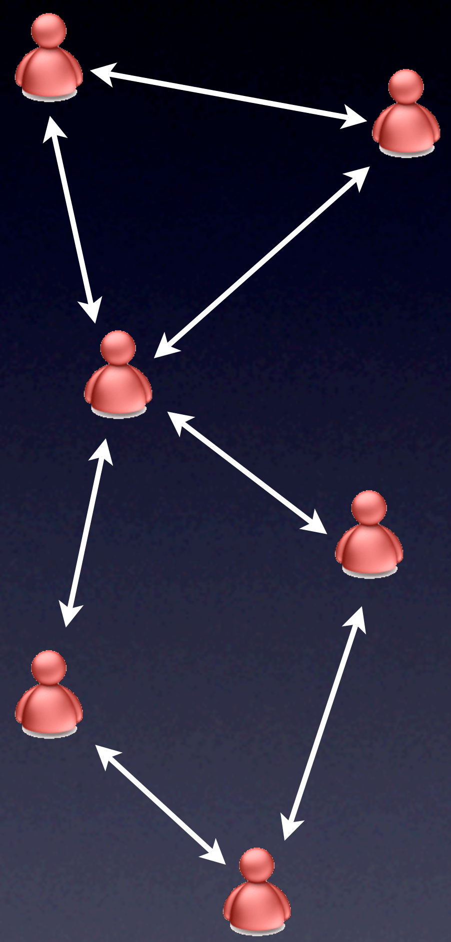
Links take effort to form/maintain

- Can't get new links easily

Misbehavior results in being ostracized

- Short-term gain vs. long-term loss

Why does hawala work?



Links take effort to form/maintain

- Can't get new links easily

Misbehavior results in being ostracized

- Short-term gain vs. long-term loss

Result: Social network used to transfer money

Ostra

Ostra

Uses social network to prevent unwanted communication

- Same mechanism as hawala

Ostra does not need a high level of trust

- Cost of failure in hawala is high → high level of trust needed
- Far less at stake in Ostra

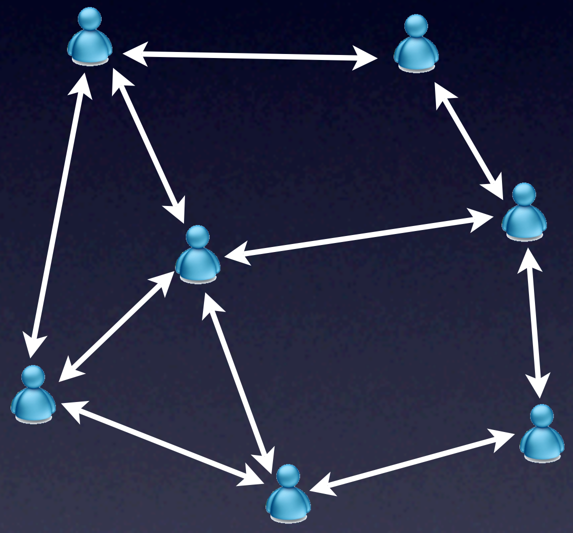
Can be applied to

- Messaging (email, IM, VoIP)
- Group communication (mailing lists)
- Content sharing (YouTube, Flickr)

Ostra's social network

Most communication systems embed social network

- Email contacts
- IM buddies
- Social network friends

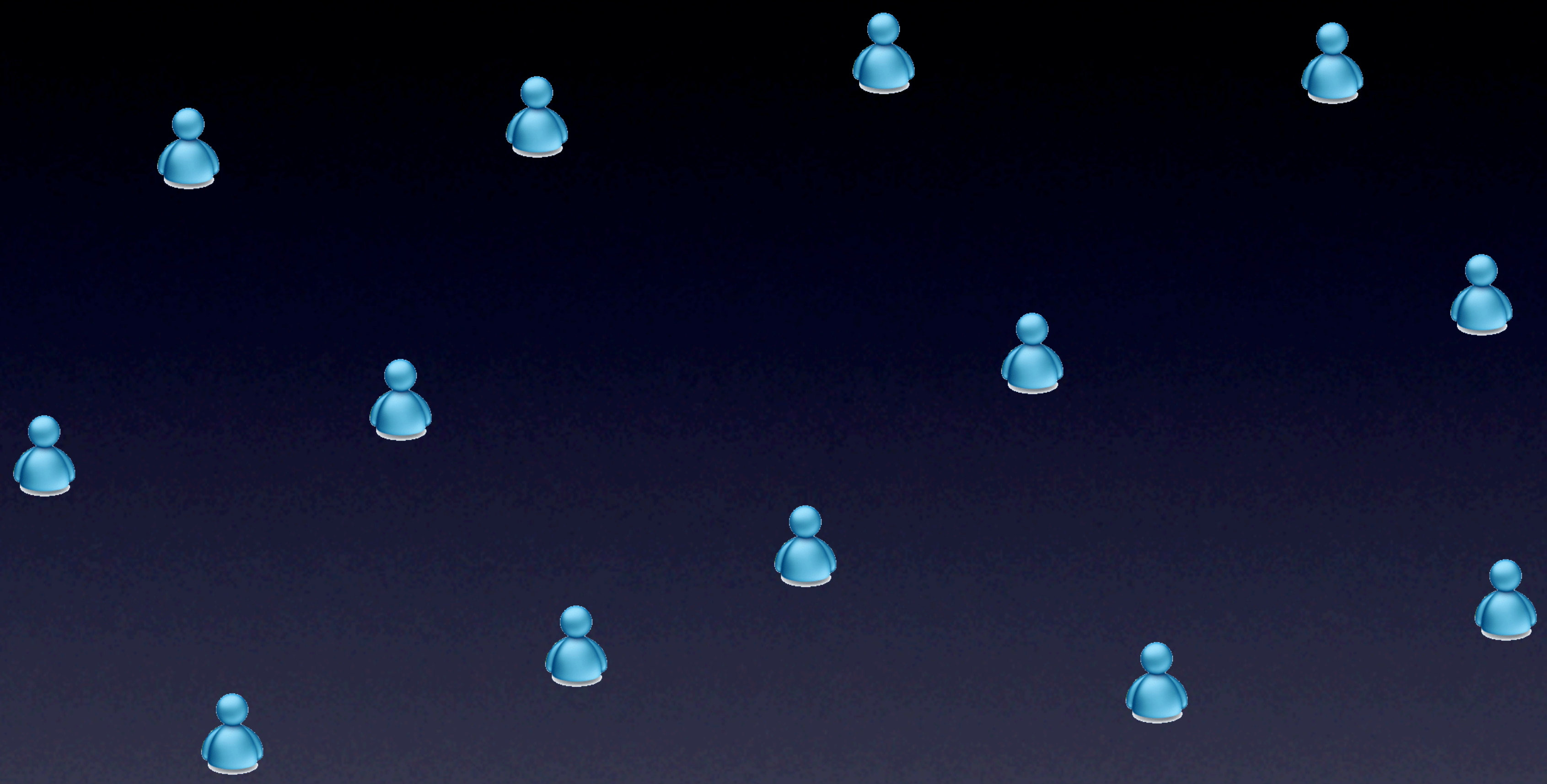


Can be explicit or implicit

Assumptions

- Links take some effort to form and maintain
- Trusted site maintains social network

High-level overview

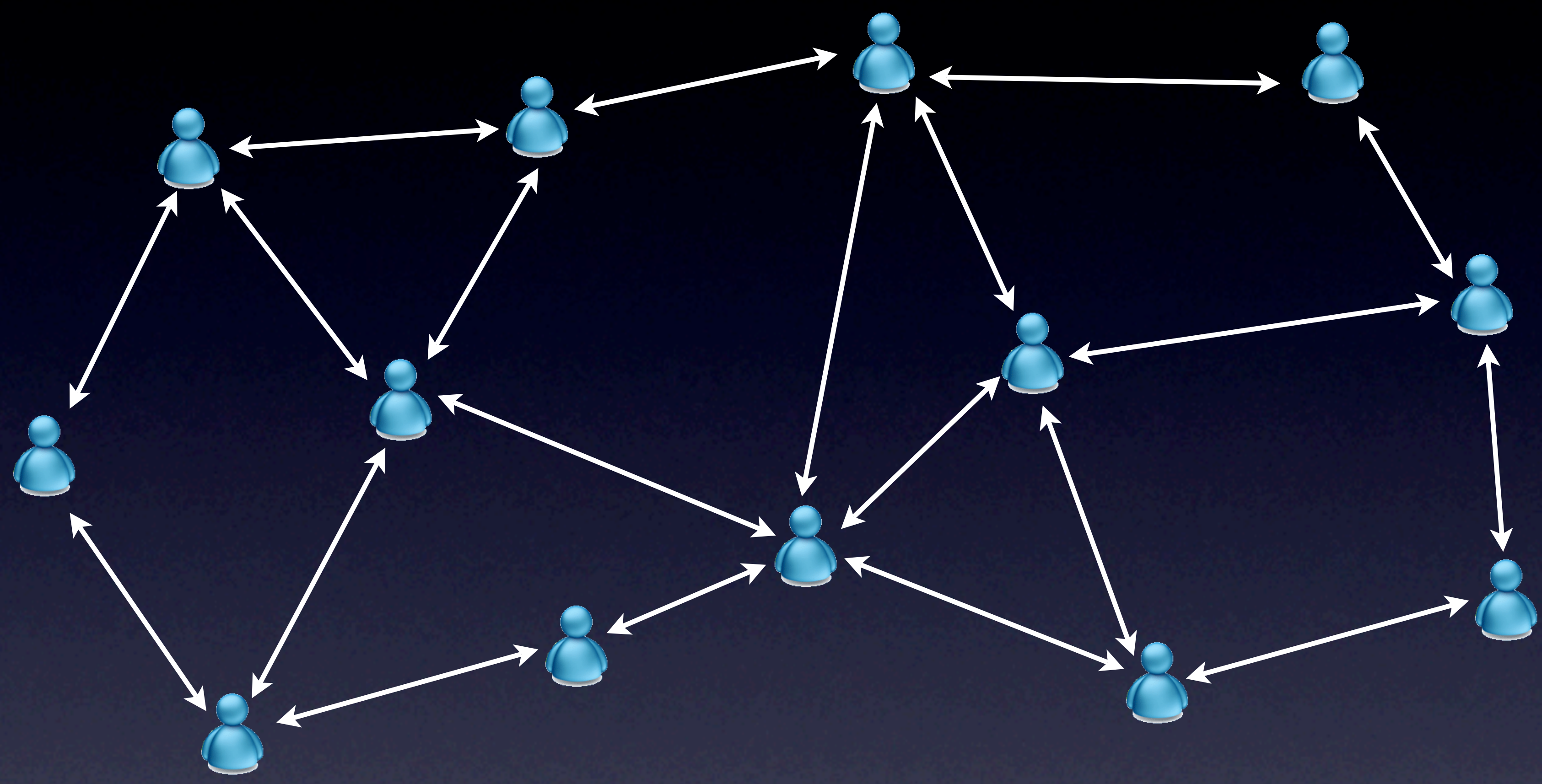


Recipients classify messages

- Can be implicit (e.g., deleting or responding to a message)

Messages are sent directly

High-level overview

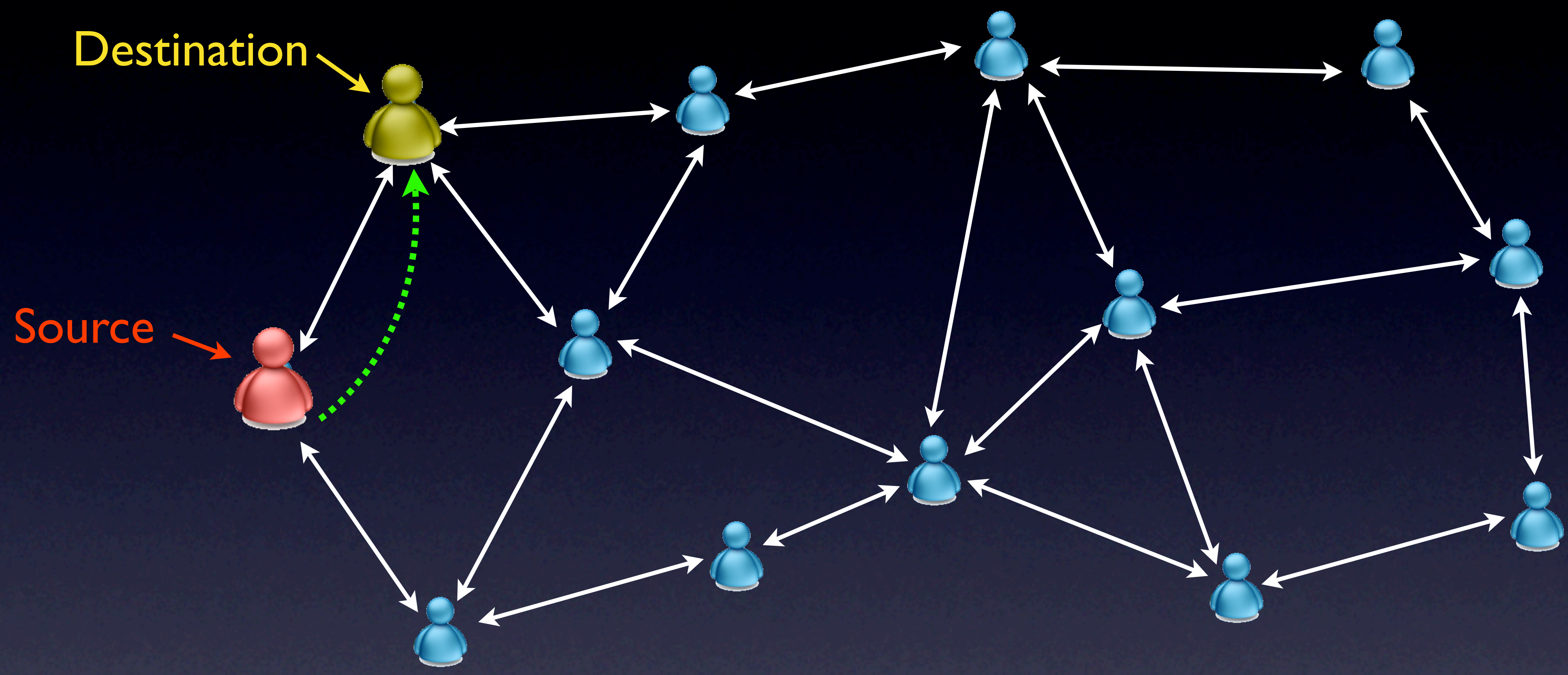


Recipients classify messages

- Can be implicit (e.g., deleting or responding to a message)

Messages are sent directly

High-level overview

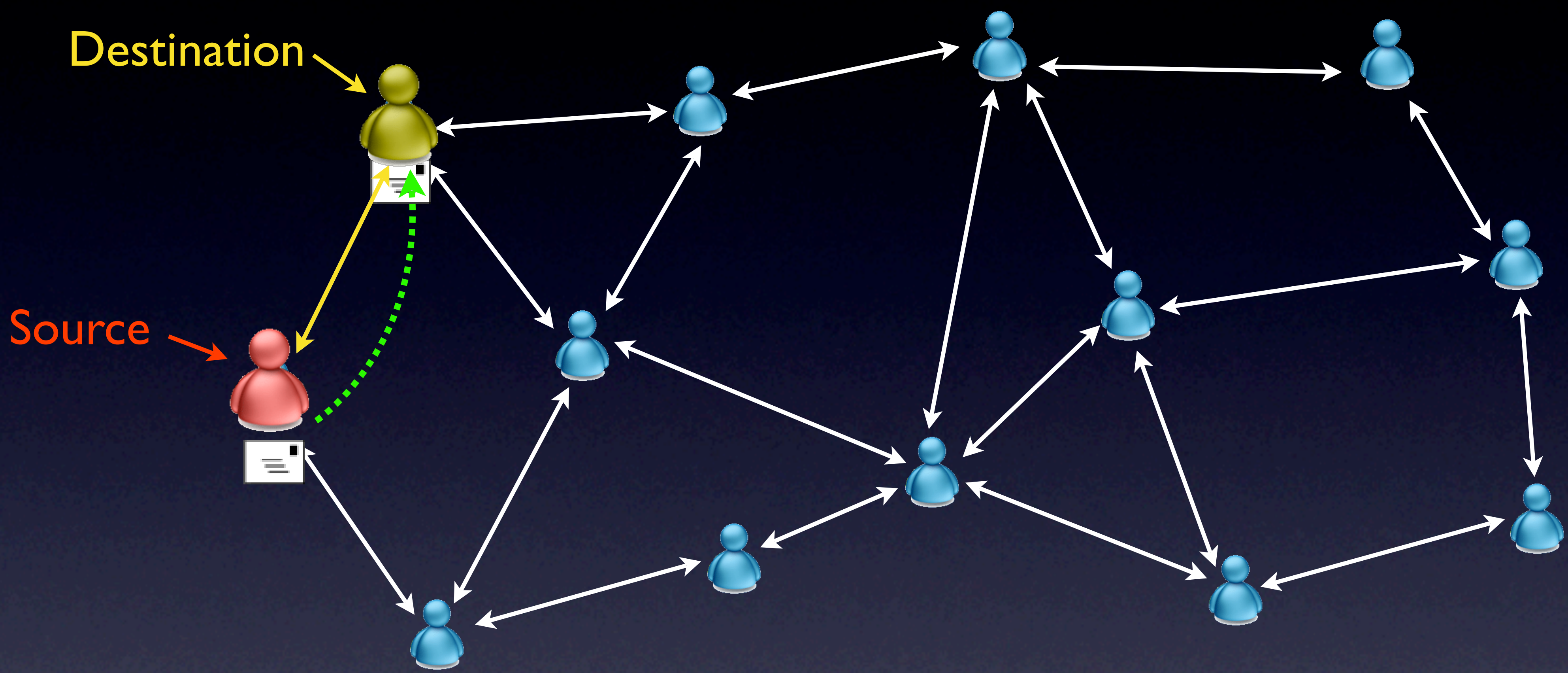


Recipients classify messages

- Can be implicit (e.g., deleting or responding to a message)

Messages are sent directly

High-level overview

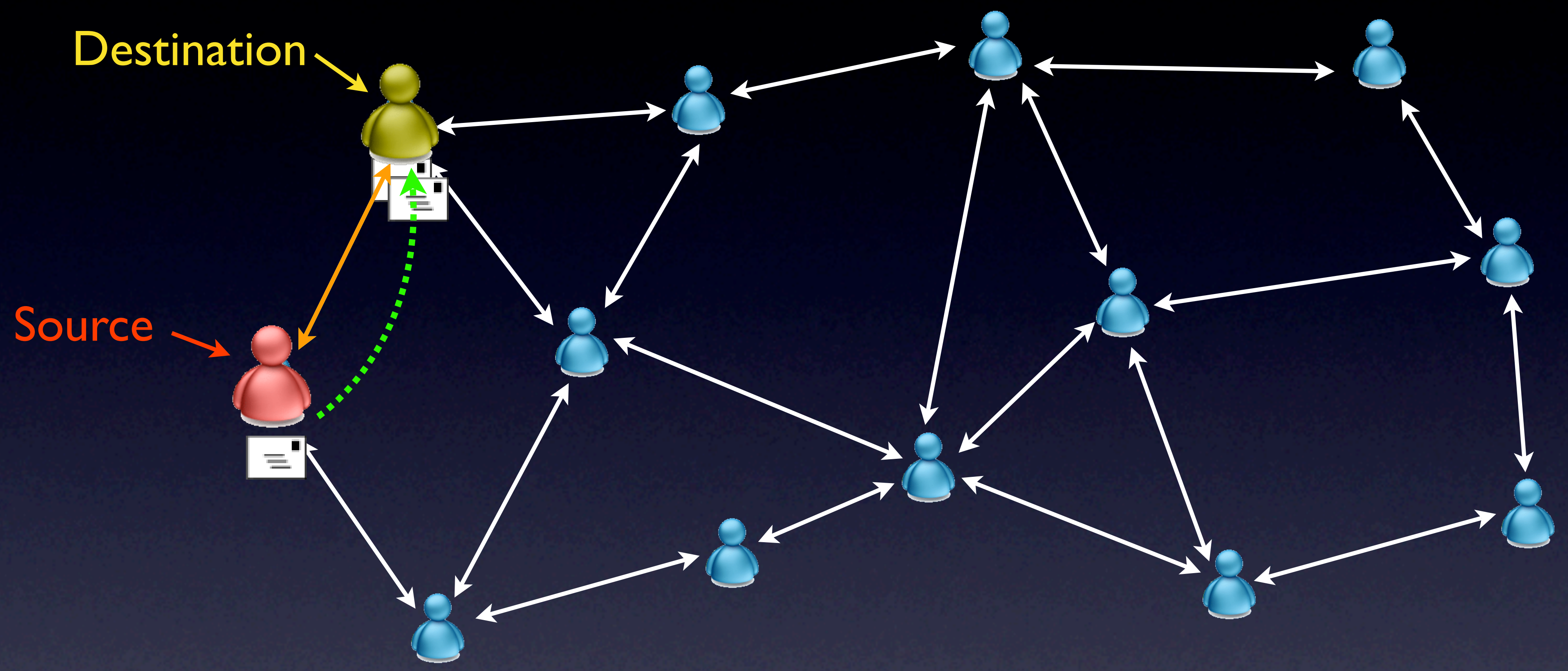


Recipients classify messages

- Can be implicit (e.g., deleting or responding to a message)

Messages are sent directly

High-level overview

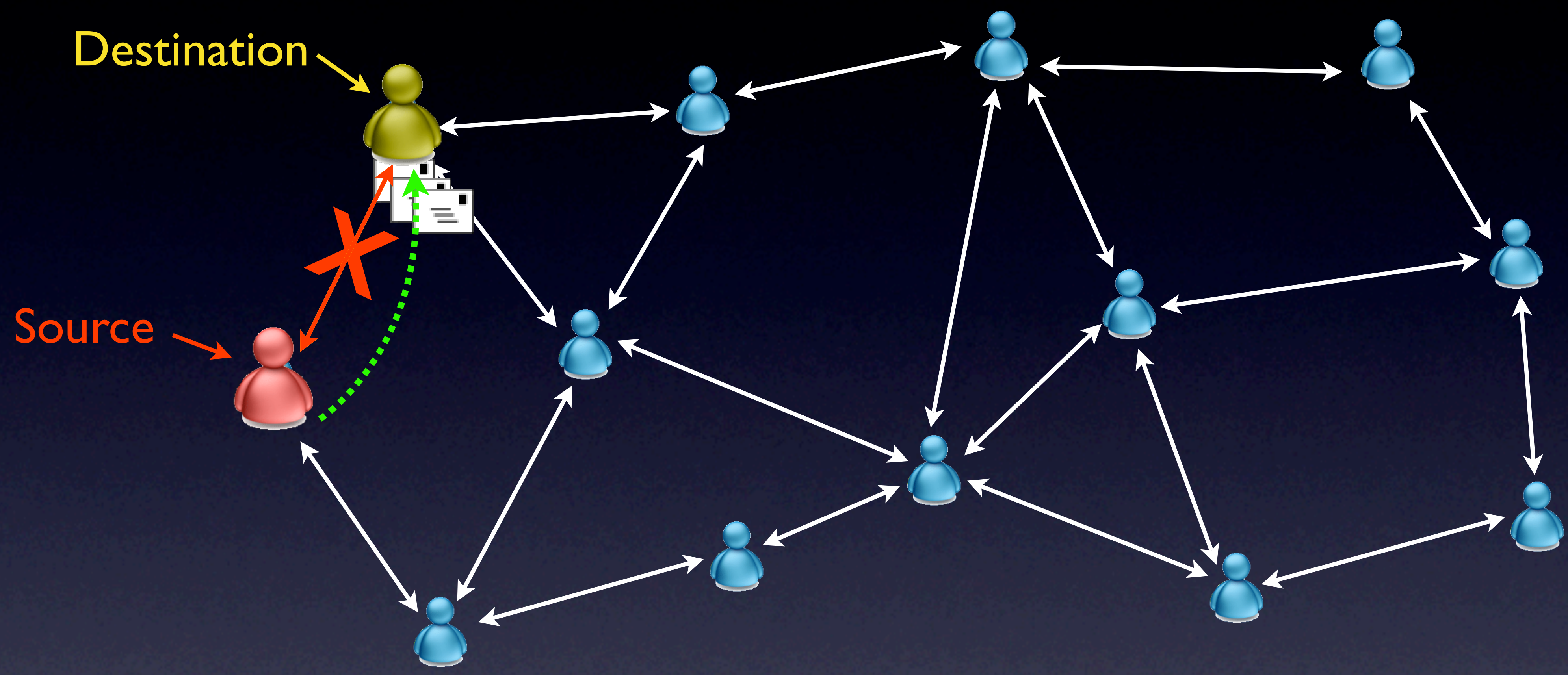


Recipients classify messages

- Can be implicit (e.g., deleting or responding to a message)

Messages are sent directly

High-level overview

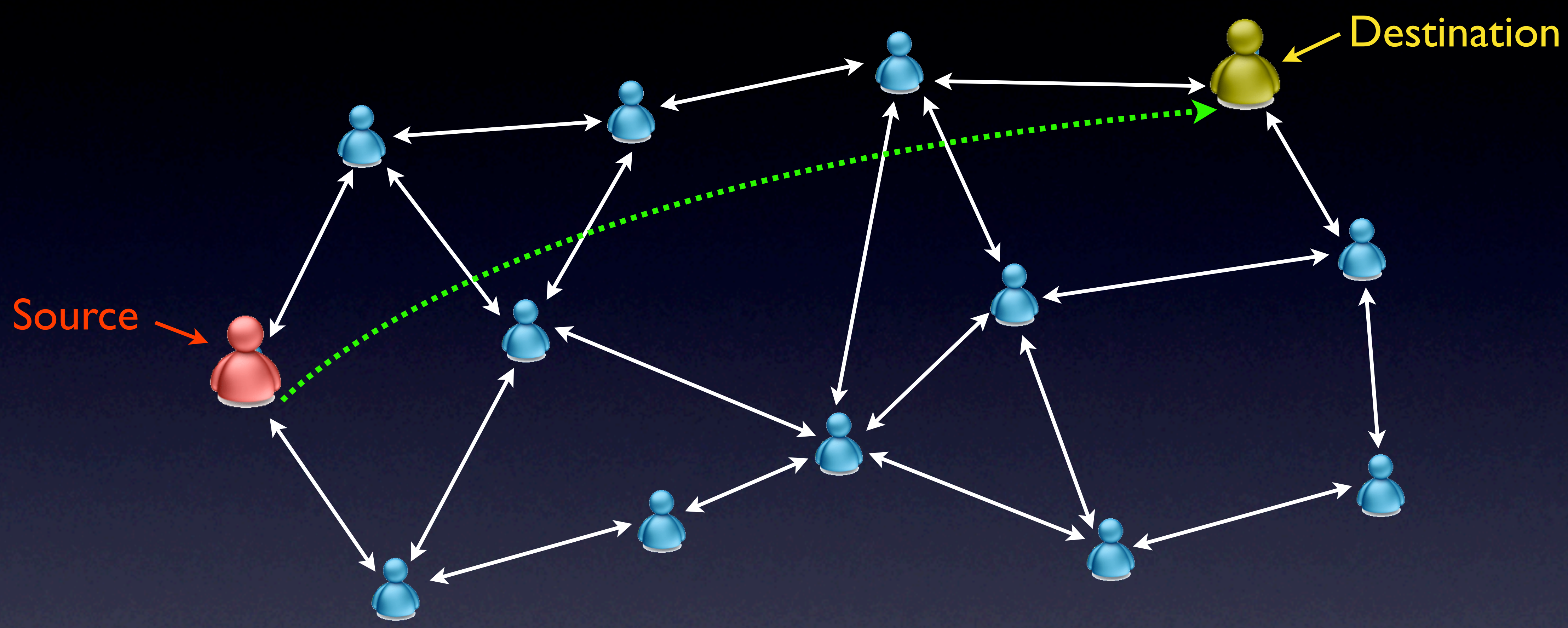


Recipients classify messages

- Can be implicit (e.g., deleting or responding to a message)

Messages are sent directly

High-level overview

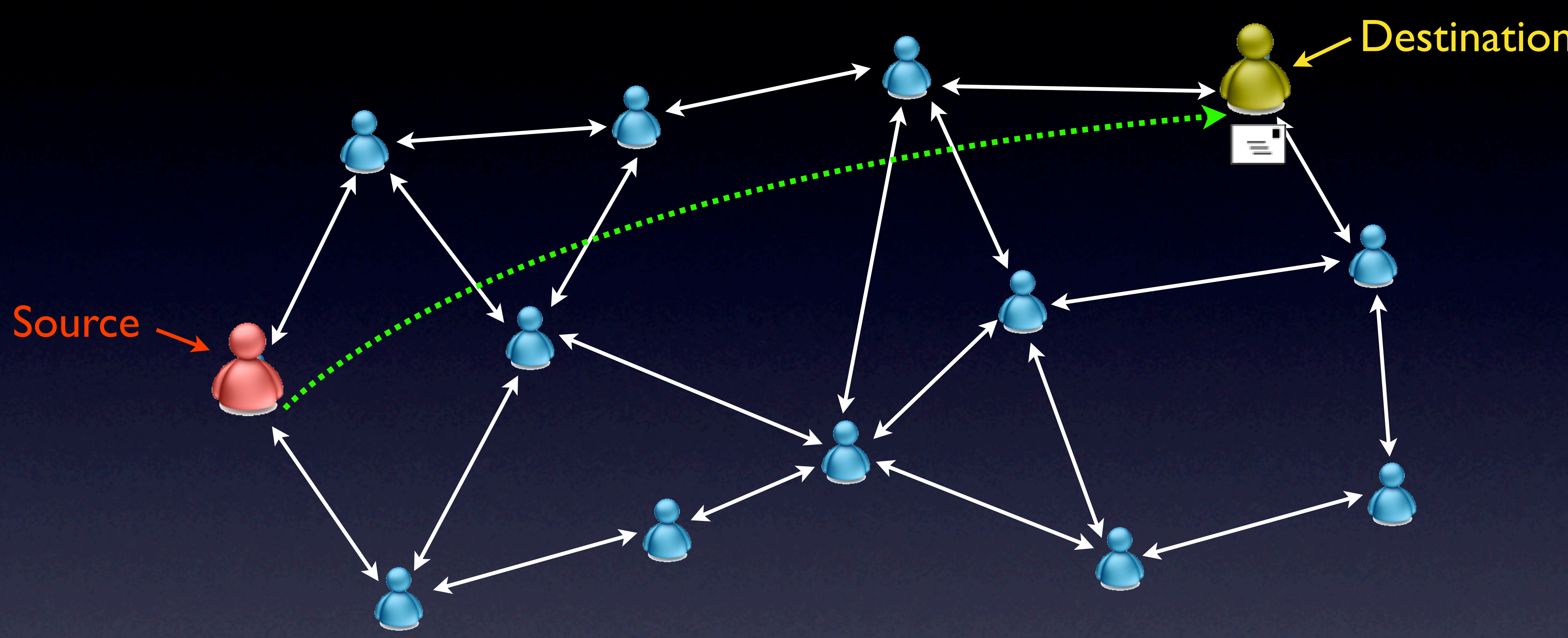


Recipients classify messages

- Can be implicit (e.g., deleting or responding to a message)

Messages are sent directly

High-level overview

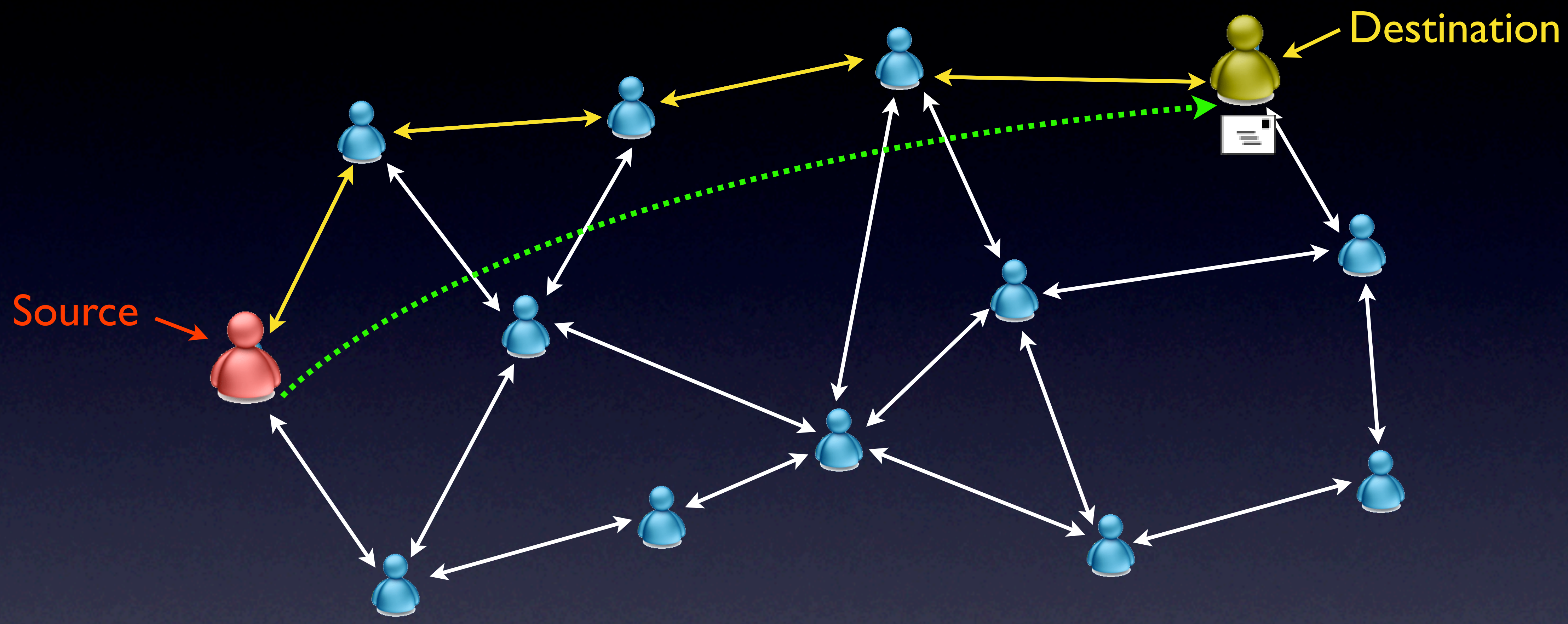


Recipients classify messages

- Can be implicit (e.g., deleting or responding to a message)

Messages are sent directly

High-level overview

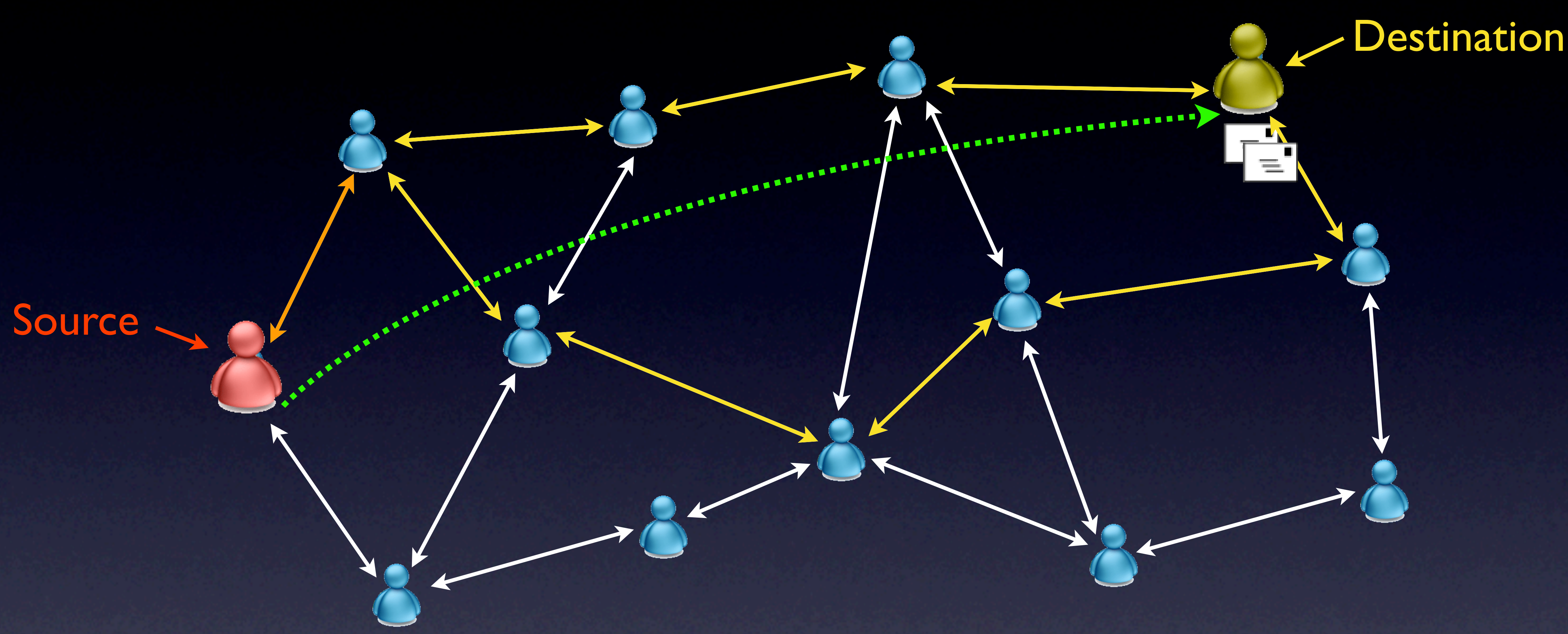


Recipients classify messages

- Can be implicit (e.g., deleting or responding to a message)

Messages are sent directly

High-level overview

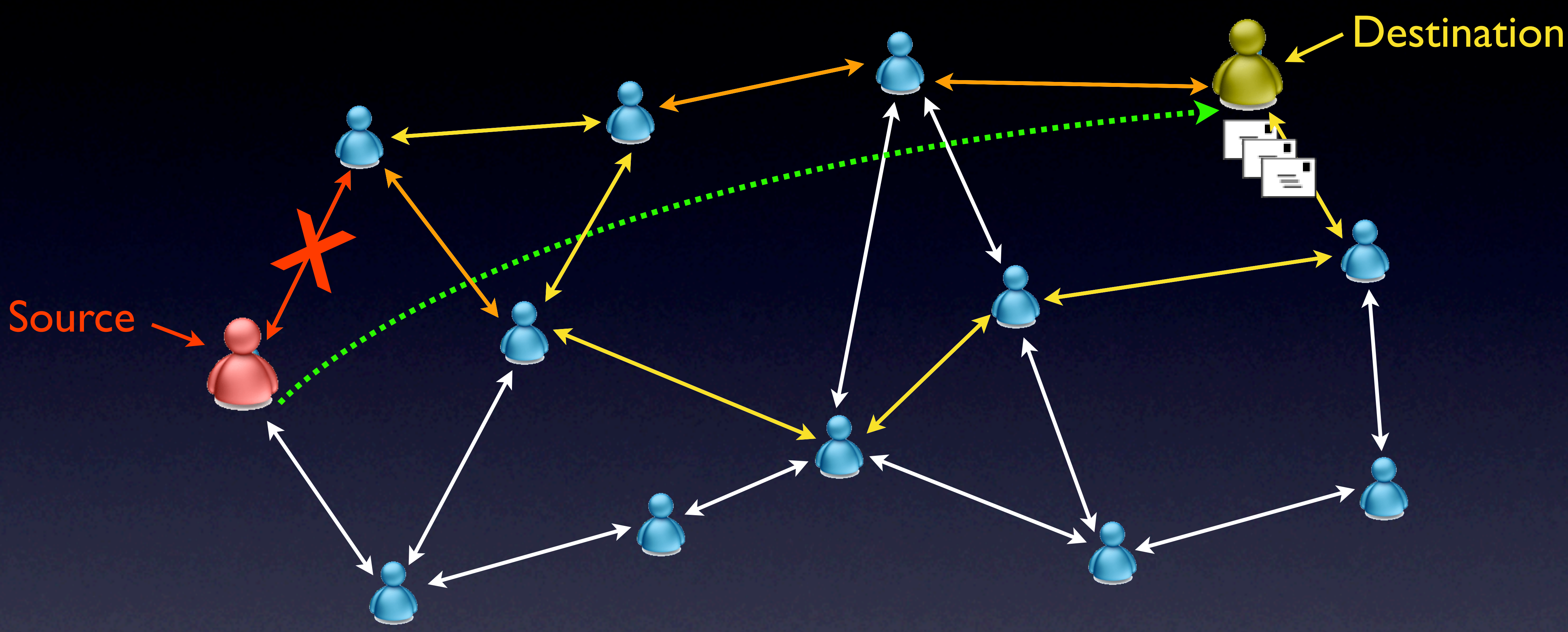


Recipients classify messages

- Can be implicit (e.g., deleting or responding to a message)

Messages are sent directly

High-level overview

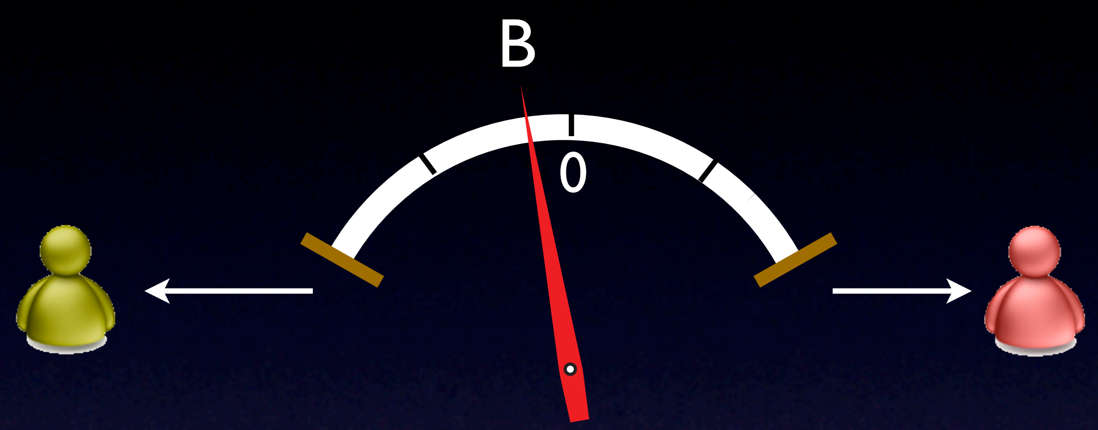


Recipients classify messages

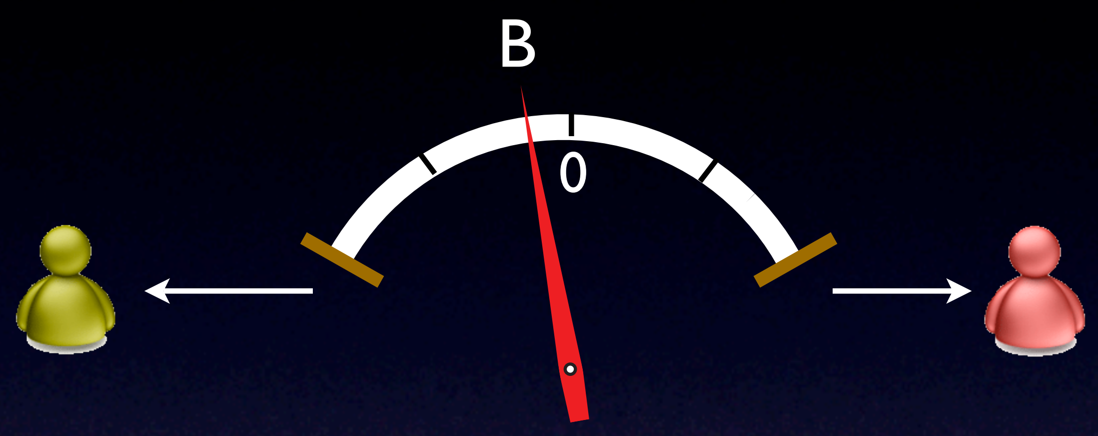
- Can be implicit (e.g., deleting or responding to a message)

Messages are sent directly

Link accounting



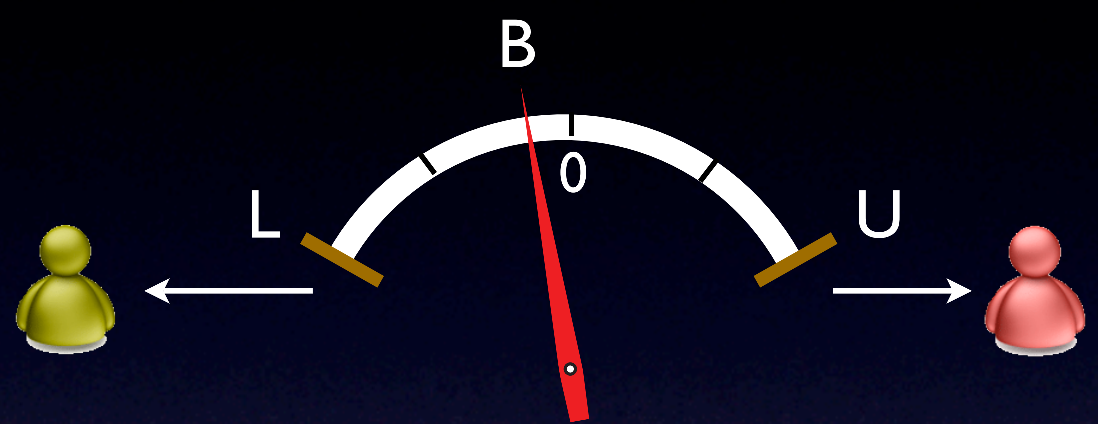
Link accounting



Each link has a credit balance B

- How much one user is "in debt" with the other

Link accounting



Each link has a credit balance B

How much one user is "in debt" with the other

Link also has credit bounds [L,U]

Maximal debt each user is willing to accept ($L \leq B \leq U$)

Link accounting

Each link has a credit balance B

- How much one user is "in debt" with the other

Link also has credit bounds [L,U]

- Maximal debt each user is willing to accept ($L \leq B \leq U$)

Sending a message

When message is sent, lower bound is temporarily adjusted

- Reset once message classified
- If adjustment cannot be made, message is delayed

If recipient marks message unwanted, balance is adjusted

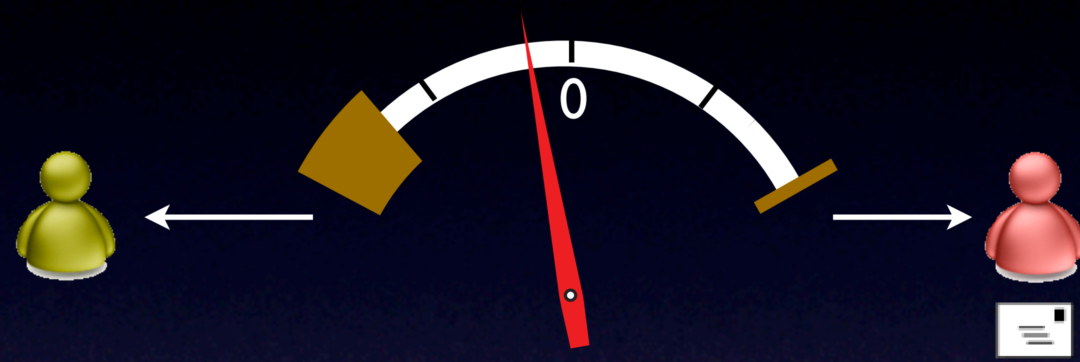
Sending a message

When message is sent, lower bound is temporarily adjusted

- Reset once message classified
- If adjustment cannot be made, message is delayed

If recipient marks message unwanted, balance is adjusted

Sending a message

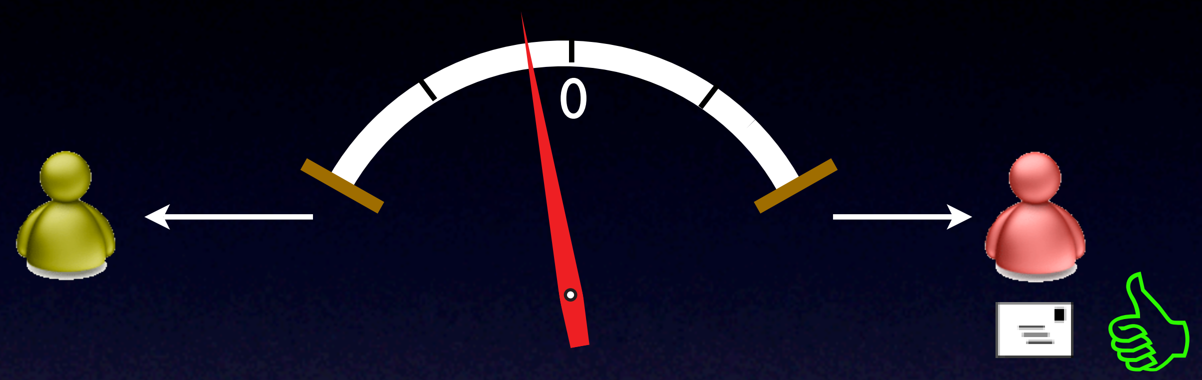


When message is sent, lower bound is temporarily adjusted

- Reset once message classified
- If adjustment cannot be made, message is delayed

If recipient marks message unwanted, balance is adjusted

Sending a message



When message is sent, lower bound is temporarily adjusted

- Reset once message classified
- If adjustment cannot be made, message is delayed

If recipient marks message unwanted, balance is adjusted

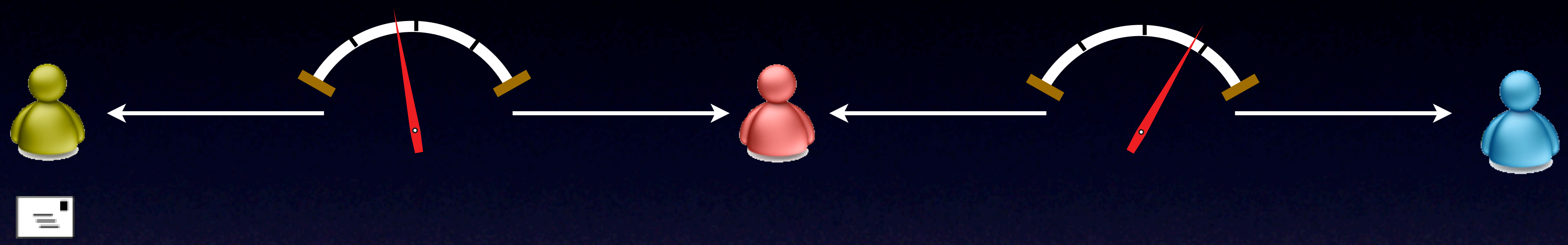
Sending a message

When message is sent, lower bound is temporarily adjusted

- Reset once message classified
- If adjustment cannot be made, message is delayed

If recipient marks message unwanted, balance is adjusted

Sending to non-friends



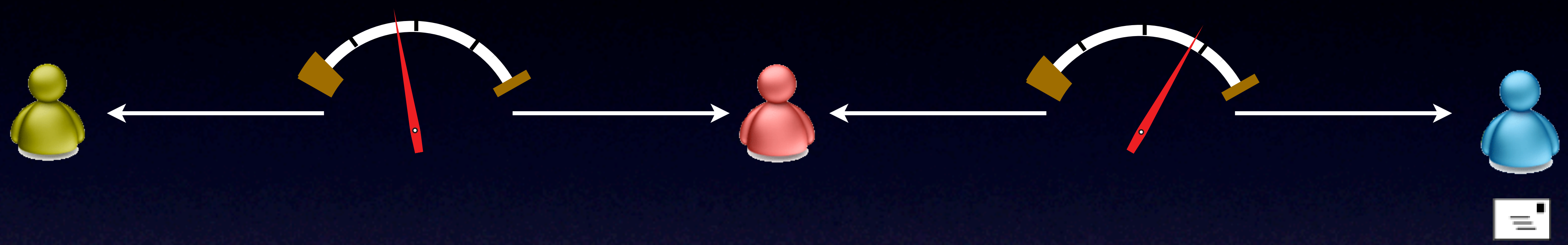
Process iterates for sending to non-friends

- Find any path from source to destination

Intermediate users indifferent to outcome

- In either case, total credit is the same

Sending to non-friends



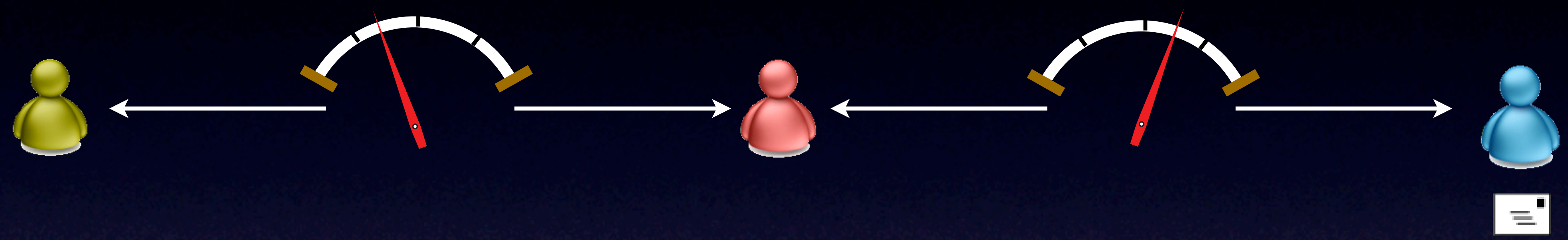
Process iterates for sending to non-friends

- Find any path from source to destination

Intermediate users indifferent to outcome

- In either case, total credit is the same

Sending to non-friends



Process iterates for sending to non-friends

- Find any path from source to destination

Intermediate users indifferent to outcome

- In either case, total credit is the same

Guarantees

What is the per-user bound on sending spam?

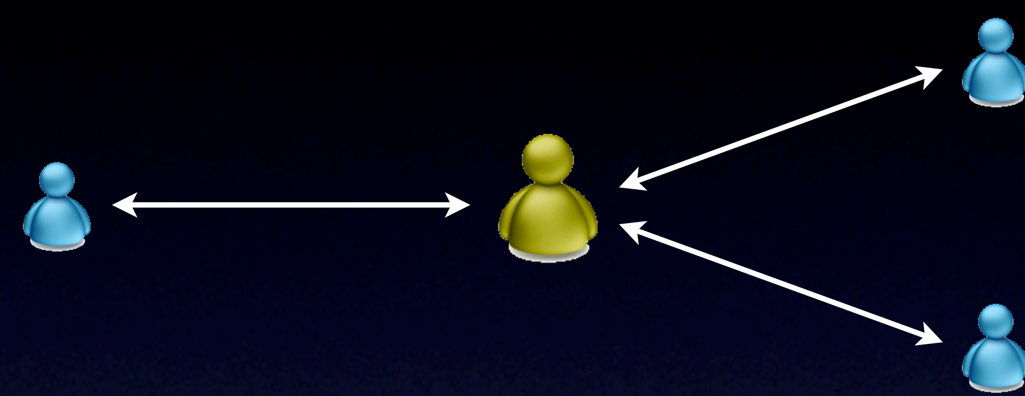
$$\underbrace{S}_{\text{Received spam}}$$

Guarantees

What is the per-user bound on sending spam?

$$\overbrace{|L|}^{\text{Lower bound}} + \underbrace{S}_{\text{Received spam}}$$

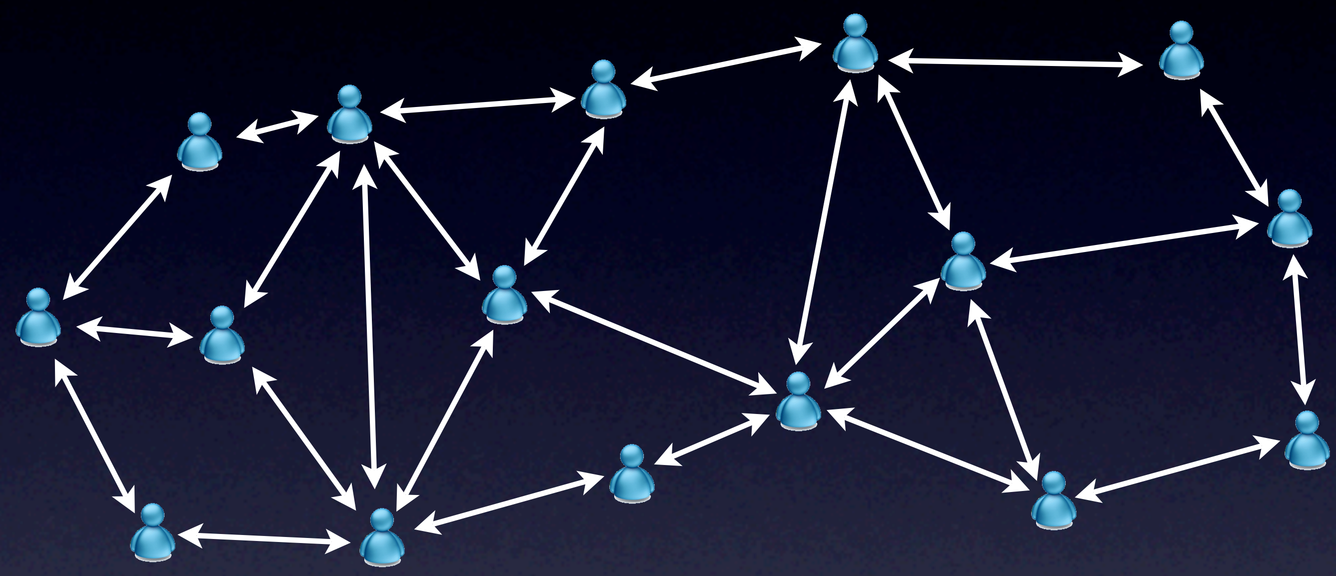
Guarantees



What is the per-user bound on sending spam?

$$\underbrace{N}_{\text{Number of links}} * \overbrace{|L|}^{\text{Lower bound}} + \underbrace{S}_{\text{Received spam}}$$

Guarantees for groups



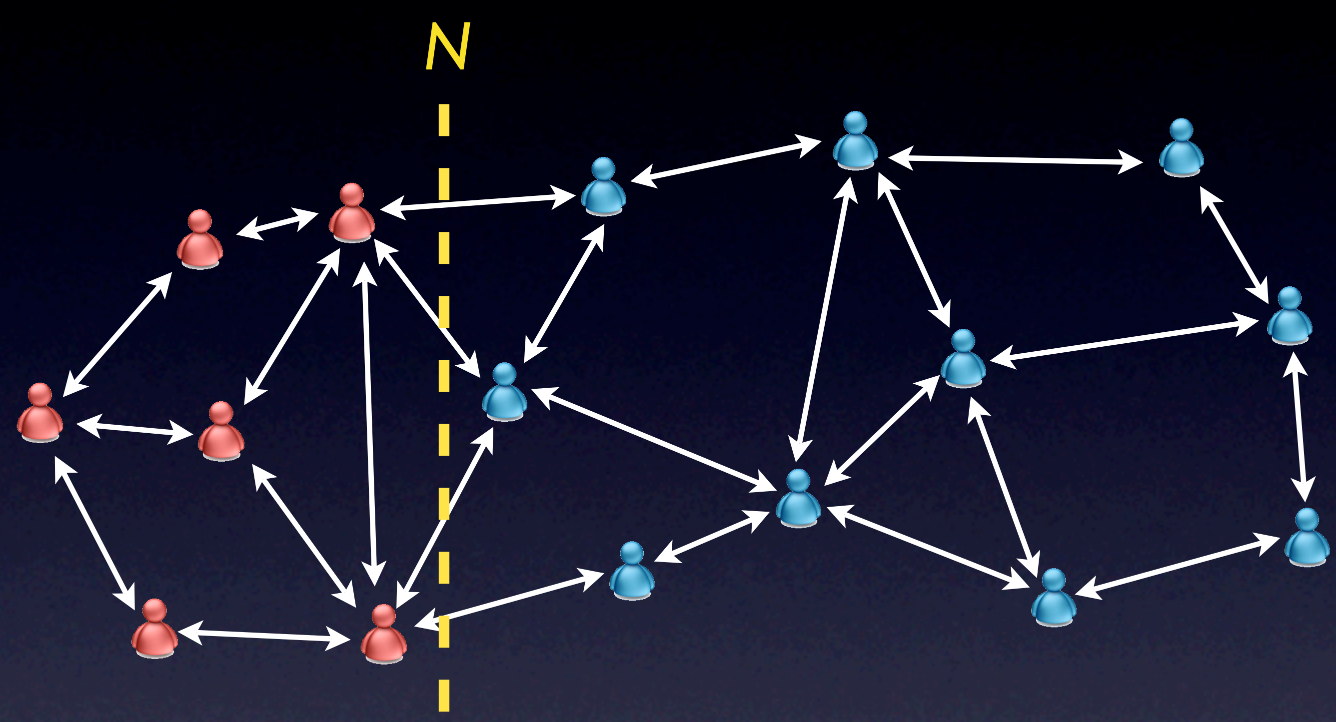
$$N * |L| + S$$

Analysis is same for any subgraph

Conservation of credit: Credit can neither be created nor destroyed

Result: Collusion doesn't help attackers

Guarantees for groups



$$N * |L| + S$$

Analysis is same for any subgraph

Conservation of credit: Credit can neither be created nor destroyed

Result: Collusion doesn't help attackers

Adjustments

An average user will occasionally
- Receive an unwanted message (receive credit)
- Send mail marked as unwanted (lose credit)

May cause user's balance to hit bounds
- If (B = L), cannot send
- If (B = U), cannot receive

Introduce credit decay *d*
- Outstanding balance (+ or −) decays (e.g., *d*=10% per day)

Preserves conservation of credit

Adjustments (cont.)

Offline users may cause credit reservation

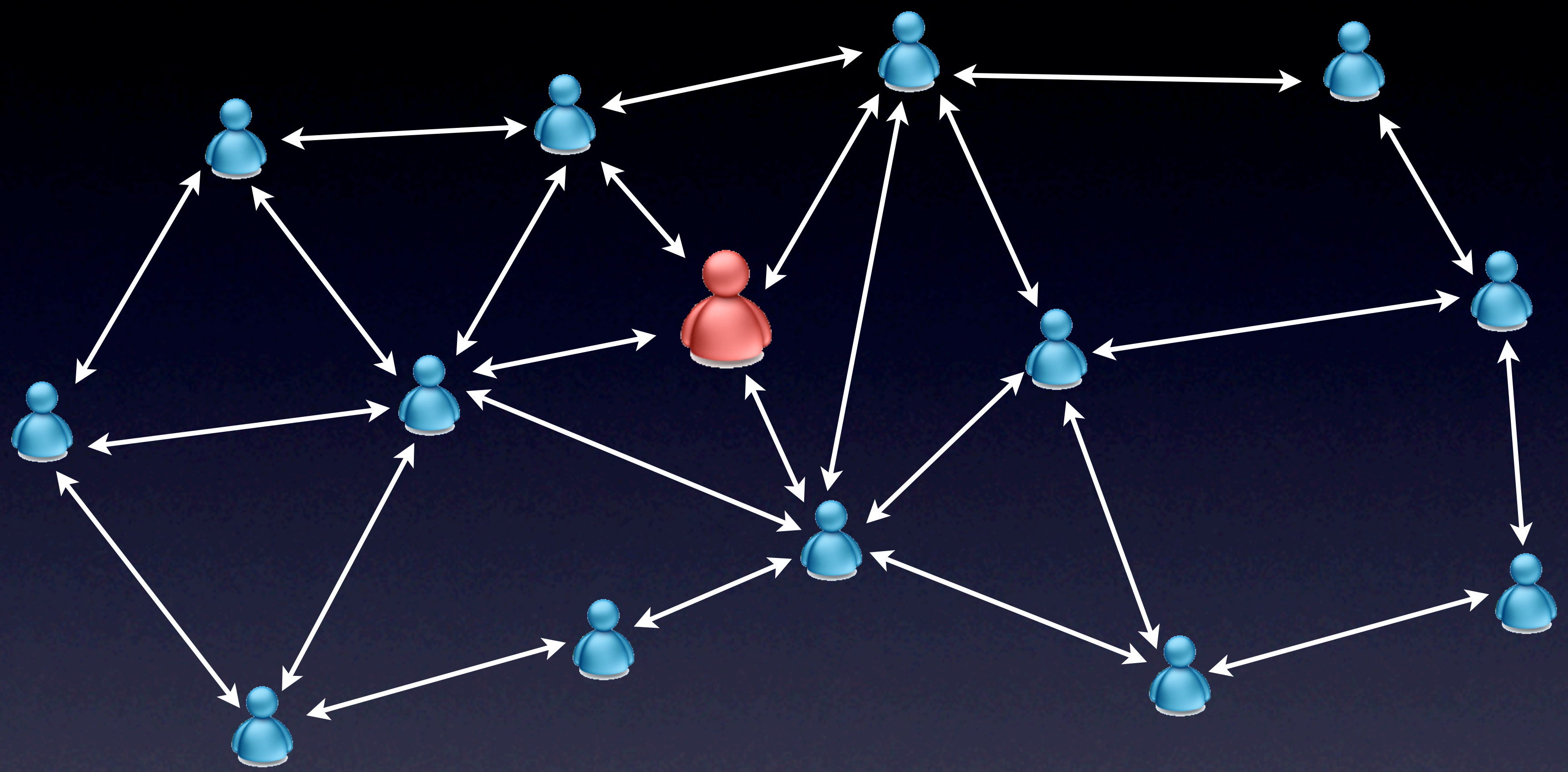
- Bounds adjusted until message classified

Introduce classification timeout *T*

- Message treated as "wanted" if unclassified after *T*

Also offers plausible deniability of receipt

Applying Ostra to content-sharing



- Create "virtual" identity for content-sharing site
  - Uploads are message to this identity
- Site uses existing mechanisms to determine if unwanted

Ostra security

Can conspiring users “create” credit?

Could Ostra reach starvation?

What about users with multiple identities?

Can attackers disconnect the network?

Ostra security

~~Can conspiring users “create” credit?~~

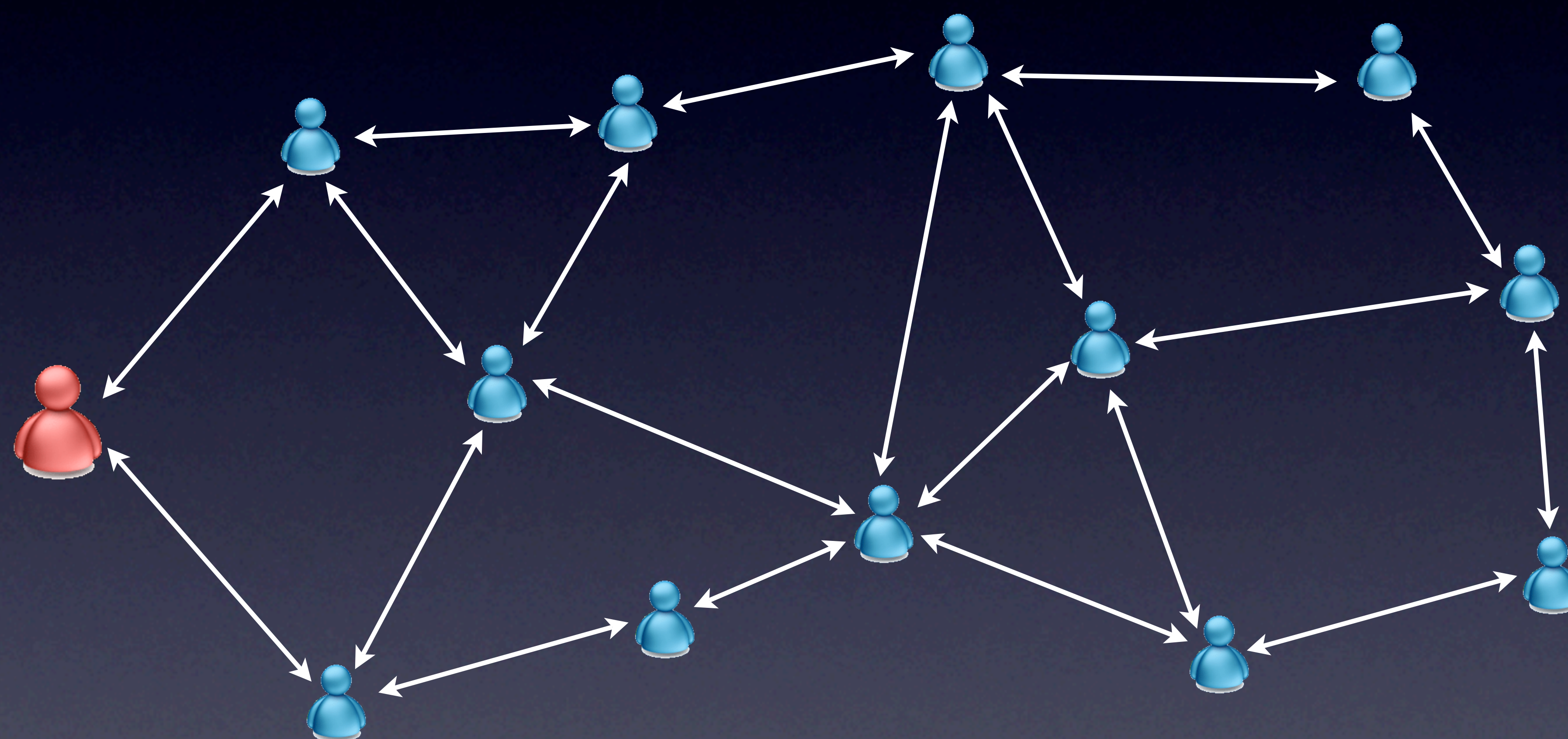
~~Could Ostra reach starvation?~~

In paper

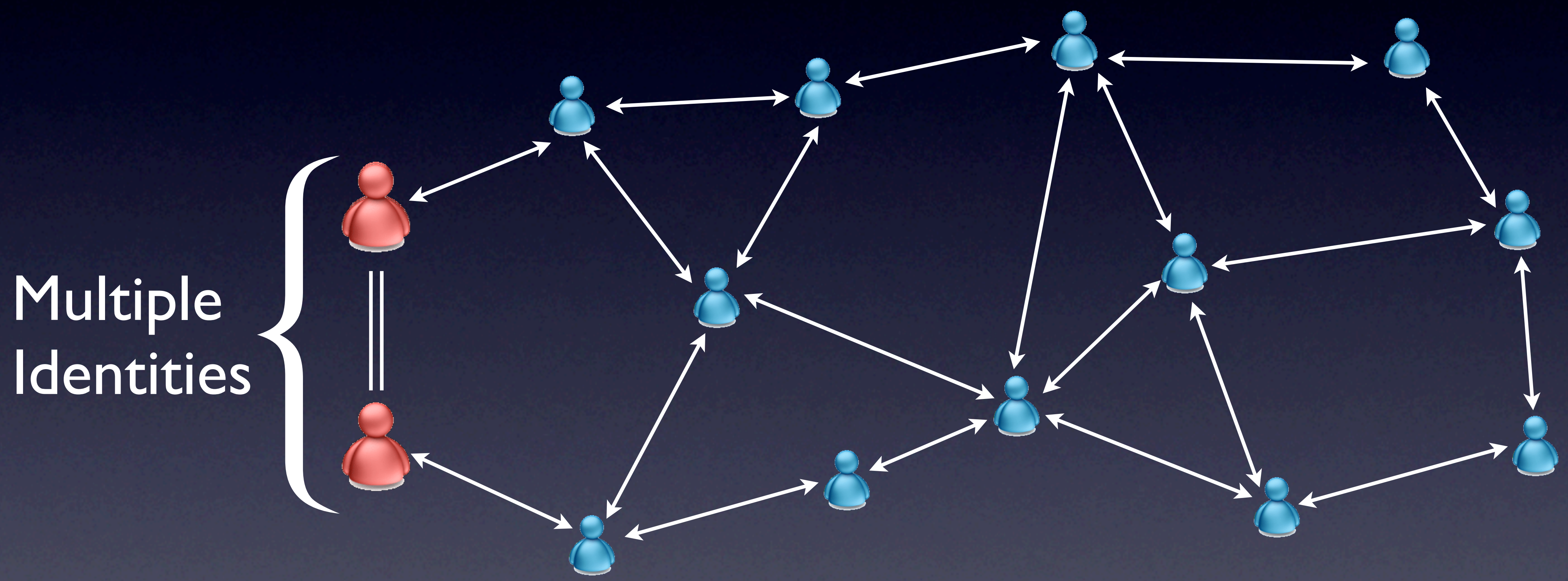
What about users with multiple identities?

Can attackers disconnect the network?

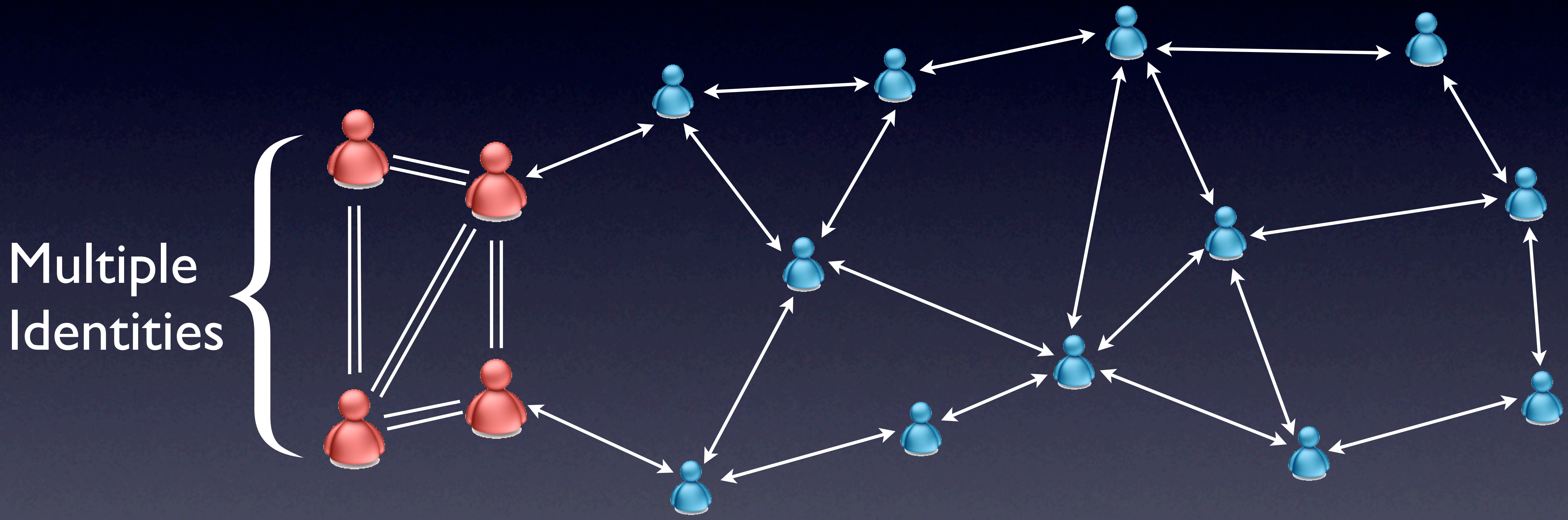
What about multiple identities?



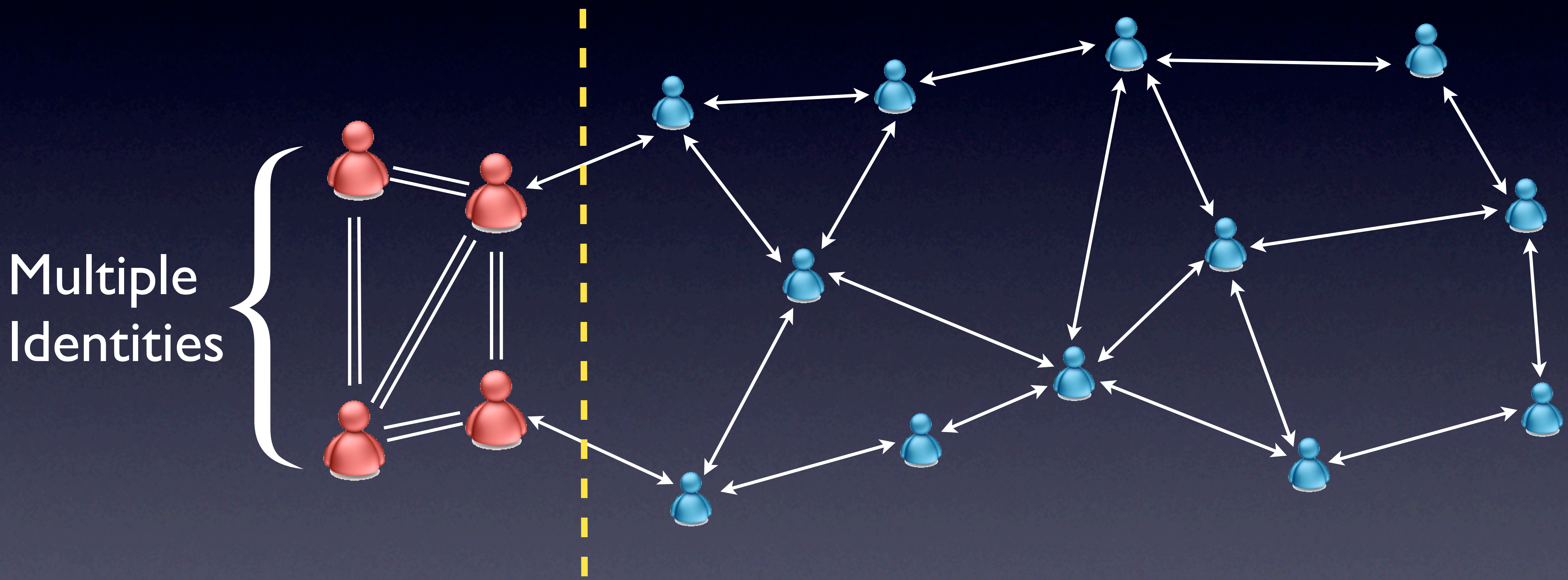
What about multiple identities?



What about multiple identities?



What about multiple identities?



Can attackers target vital links?

Social networks tend to have dense core [IMC'07]

Min-cut is almost always at source or destination (see paper)

Can attackers target vital links?

Social networks tend to have dense core [IMC'07]

Min-cut is almost always at source or destination (see paper)

Evaluation

Is Ostra effective in blocking unwanted communication?

Does Ostra delay message delivery?

What is the complexity of finding paths?

How do parameter settings affect performance?

Does incorrect message classification break Ostra?

Are there vulnerable links in social networks?

Evaluation

Is Ostra effective in blocking unwanted communication?

Does Ostra delay message delivery?

~~What is the complexity of finding paths?~~

~~How do parameter settings affect performance?~~

~~Does incorrect message classification break Ostra?~~

~~Are there vulnerable links in social networks?~~

In paper

Simulating Ostra

Need a social network and a message trace

- Social network trace from YouTube (446K users, 1.7M links)
- Email trace from MPI (150 users for 3 months, 13K messages)

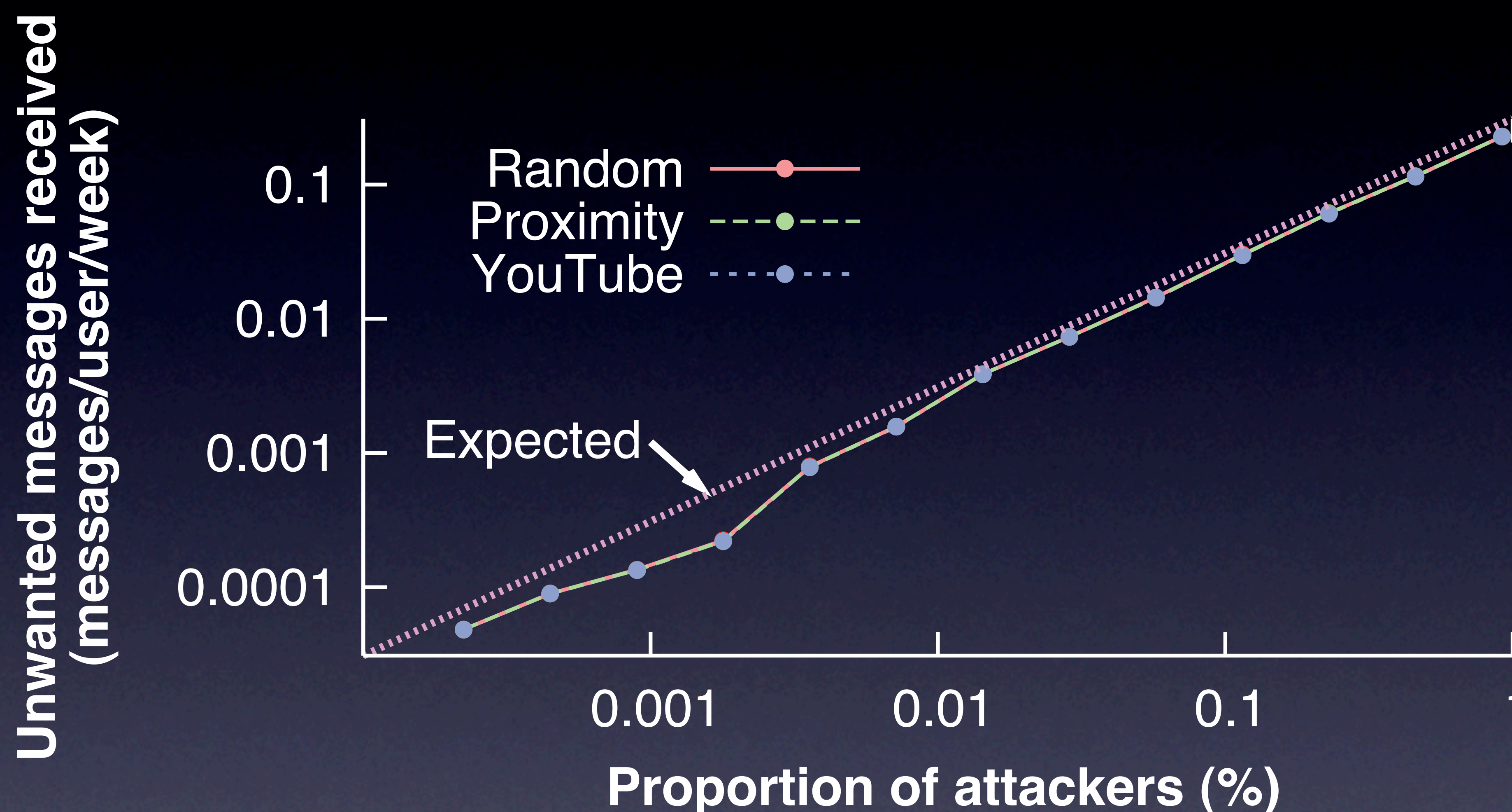
Simulated Ostra in three scenarios

- Messaging with random traffic
- Messaging with proximity-biased traffic
- Centralized content-sharing site

Simulation parameters

- Selected random attacking users
- Bounds of [-3,3] and d=10% per day

Does Ostra block spammers?



Ostra limits amount of unwanted communication

Even with 20% attackers, only 4 messages/good user/week

Do messages get delayed?

Classification delay (h)	Fraction delayed	Average delivery delay (h)
2	1.3%	4.1
6	1.3%	16.6

Very few messages get delayed

Related work

Preventing unwanted communication

- Content filtering: DSPAM, SpamAssassin
- Whitelisting: LinkedIn, RE: [NSDI'06]

Using social networks

- PGP Web of Trust
- SybilGuard [SIGCOMM'06]
- SybilLimit [Oakland'08]

Conclusion

Ostra: a new approach to preventing unwanted communication

- Inspired by offline trust

Leverages social network that often already exists

Desirable properties

- Does not require global user identities
- Does not rely on automatic content classification
- Respects recipient's notion of unwanted communication

Can be applied to messaging, as well as content sharing

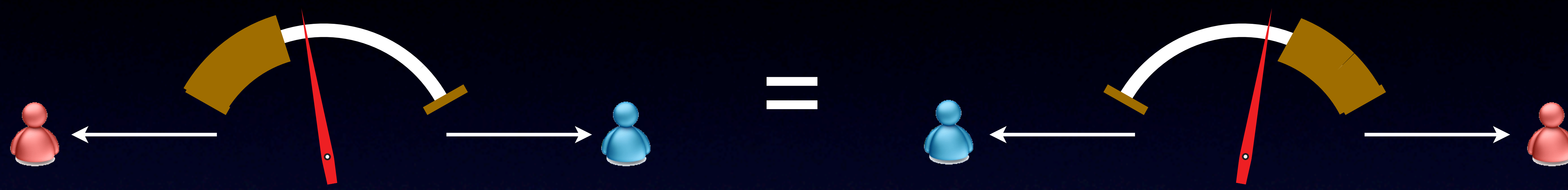
Questions?

Updated guarantees

$$\underbrace{d}_{\text{System decay}} * \underbrace{N}_{\text{Number of links}} * \underbrace{|L|}_{\text{Lower bound}} + \underbrace{S}_{\text{Rate of incoming spam}}$$

Bound on amount of spam becomes bound on rate of spam

What's up with U and L?



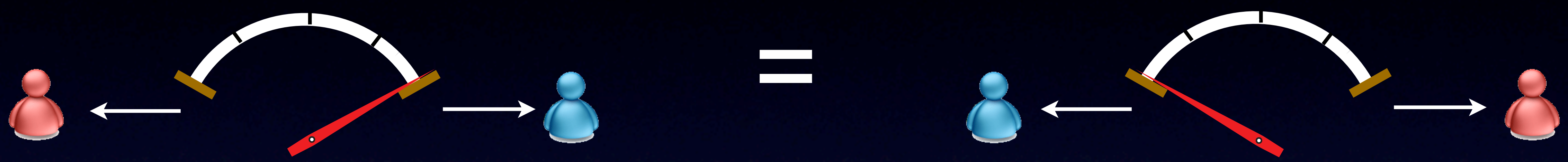
Link balance *B* and bounds [*L*,*U*] are from one user's perspective

- Link can be viewed from from other's perspective, too

For link X ↔ Y, all values symmetric

- $B_X = -B_Y$
- $L_X = -U_Y$
- $U_X = -L_Y$

Why can't I receive when B=U?



When B=U on a link

- For other user, B=L

Thus, other user can't send

- So you can't receive

Full decentralization

So far, assumed a centralized site

- Keeps link state
- Finds paths

In paper, sketch of decentralized design

- Routing using techniques from MANETs
- Link state is kept decentralized

Work in progress

Can attackers target users?

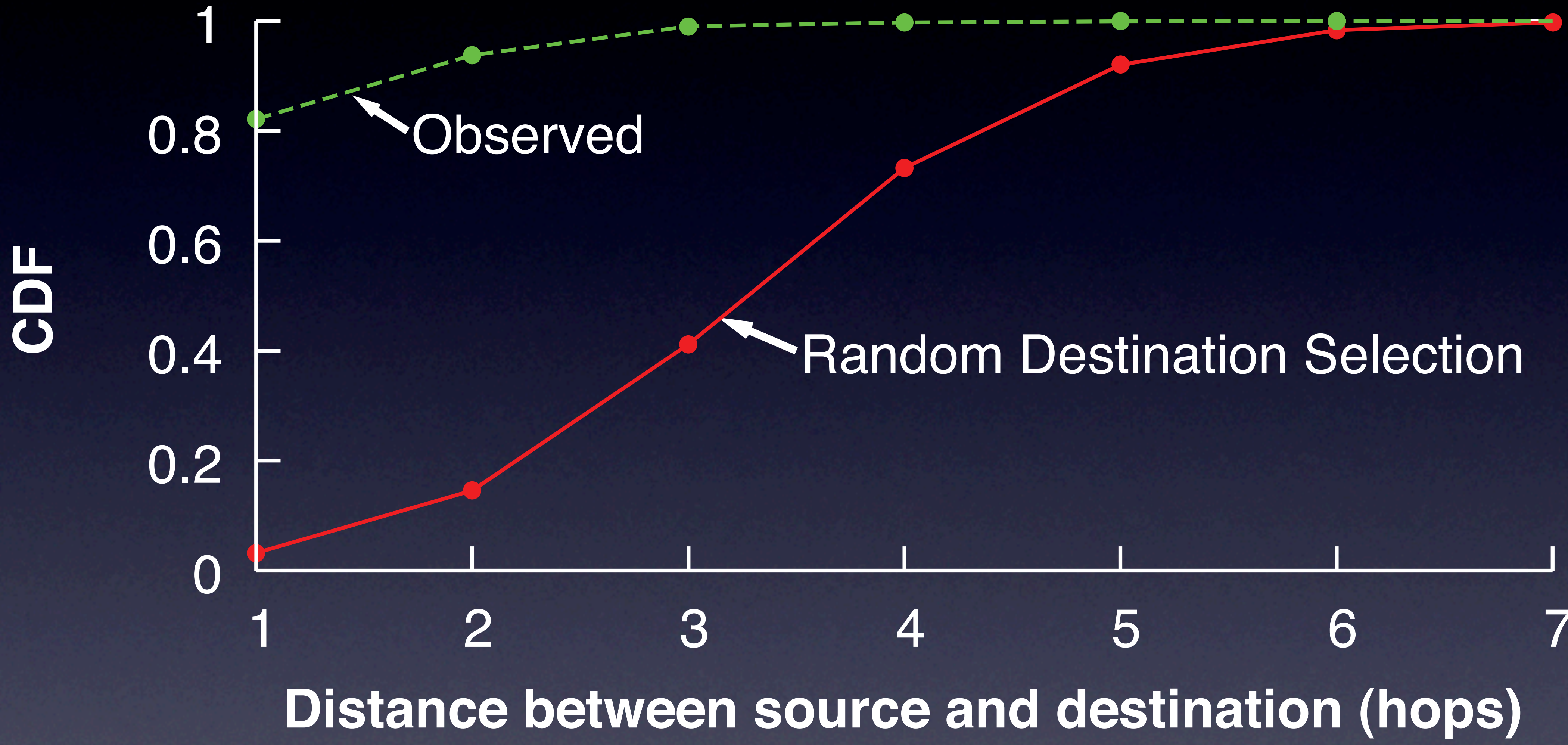
Can attackers prevent users from receiving messages?

- Send victim lots of unwanted communication
- Victim has too much credit to receive

But, victim has simple way out

- Can "donate" credit to friends
- And attackers quickly run out of credit

Who do people talk to?



Users communicate with close users

Reduces path computation complexity

More on content sharing

Why don't links in the YouTube graph run out of credit?