Towards "Dark" Social Networking Services

ICSI, 27. 3. 2013

Thorsten Strufe

Joint work with Stefanie Roos,

Andreas Höfer, Benjamin Schiller, Antonio Cutillo

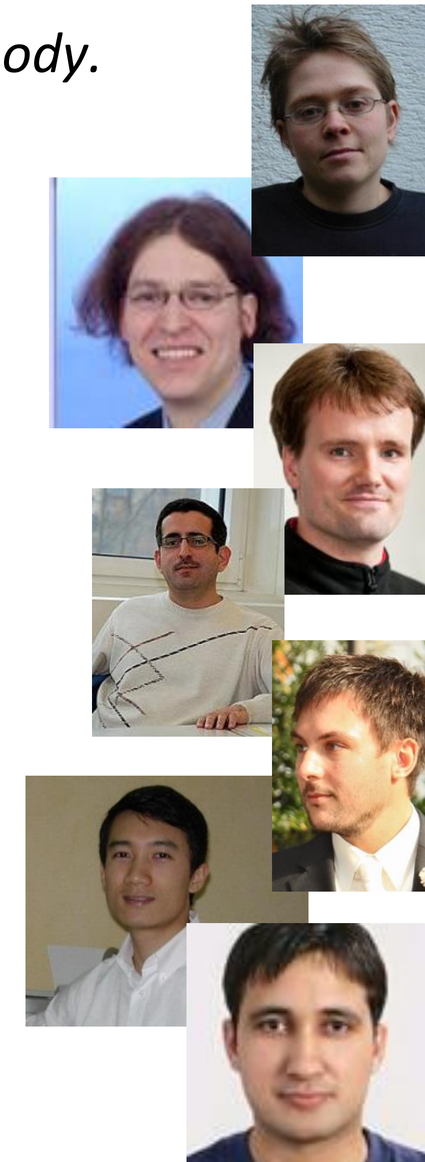
The P2P Group – What we are trying to do

- *Aim: Private and fair services (communication) for everybody.*

- Topics
  - Privacy preserving social networking
  - Robust and *resistant* means of communication

- Tools: distribution of data, processing, and control
  - Measurements
  - Analysis and modelling
  - Protocol design and simulation
  - Prototyping and measurements

- Scopes
  - Short term: Immediate remedies
  - Longer term: Paving the way
  - Vision: bullet-proof privacy/resistance



Why “dark” social networking services?

- Corporations and governments suppress individuals in plenty of ways
- ***Network effects***, quasi monopolies, and perfect ***observability*** aggravate this situation with ***digital markets***
  - individuals are incapable of understanding/checking ***what happens with their data***
  - Using and ***losing data*** to selected systems is ***not a free choice*** anymore
  - Corporations/governments abuse their power for discrimination, commercialization, and enforcing terms of use

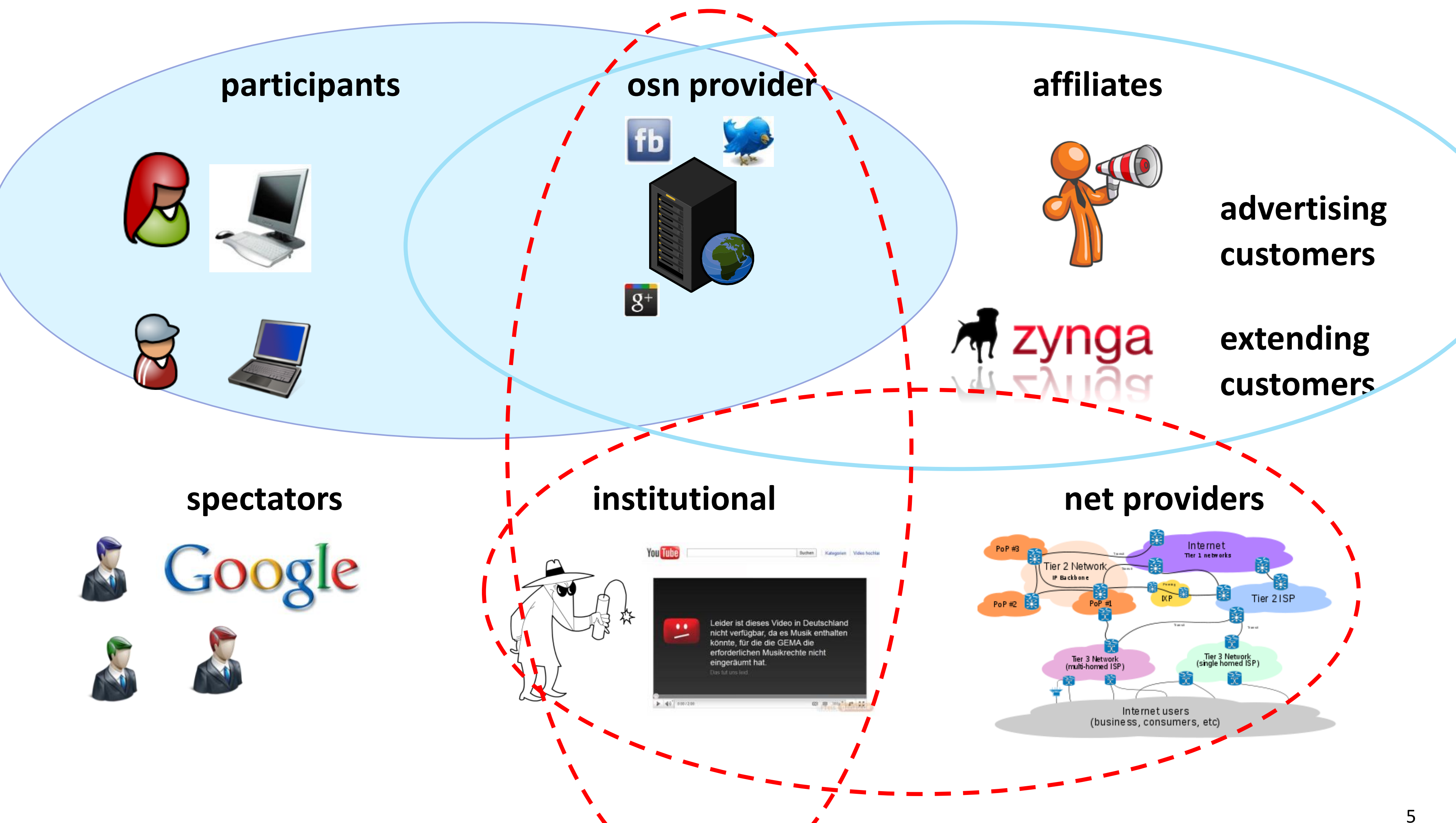
- Comprehensive identity concealment required for freedom of speech
- Way to publish information without fear of retribution necessary

- Requires a system that enables individuals to
  - Communicate anonymously/under pseudonyms
  - Publish information reliably, and anonymously
  - Conceal their participation to untrusted parties (anybody)

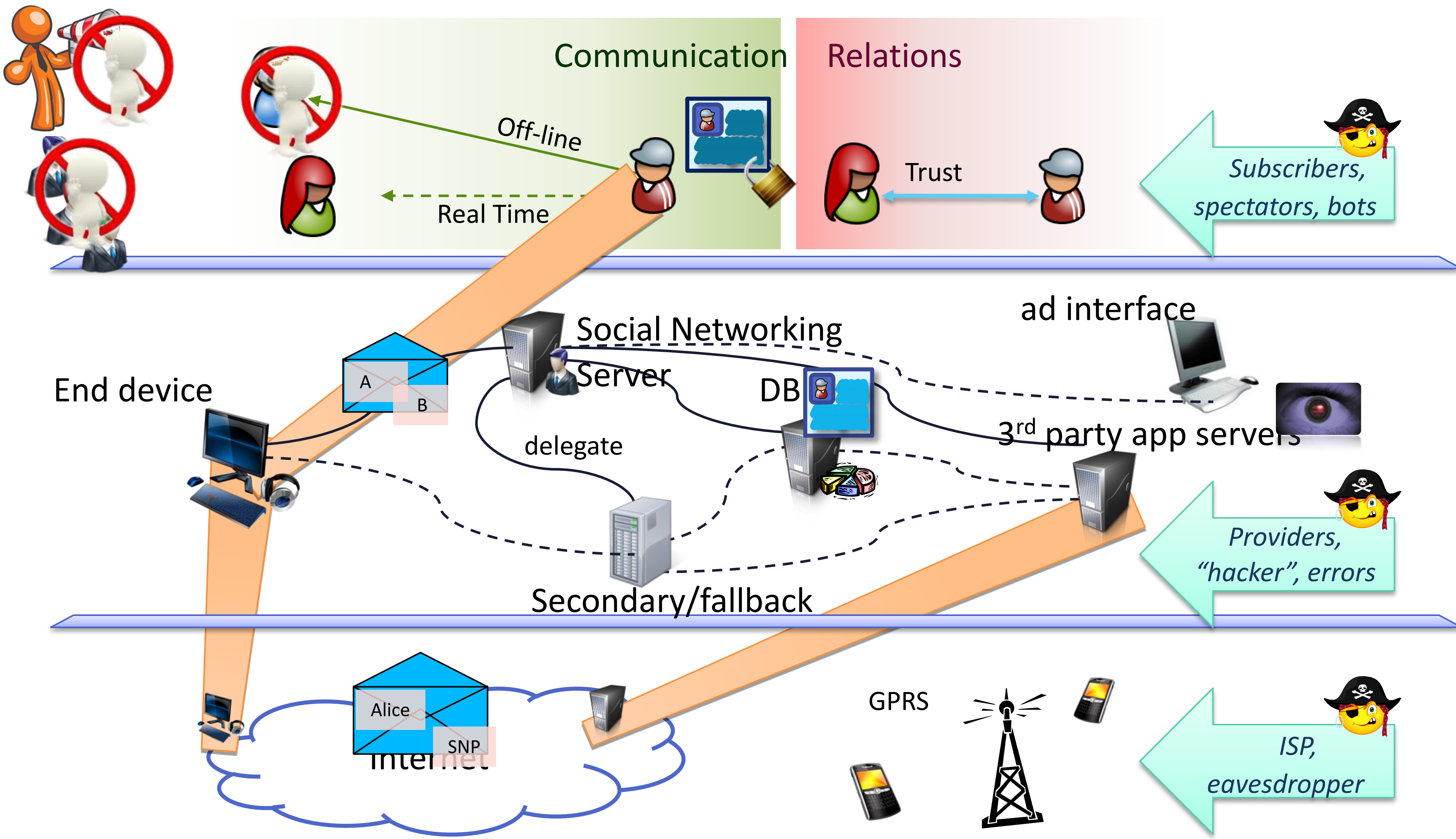
Today‘s means of communication

Quelle: www.wsj.com

Stakeholders in Communication Services

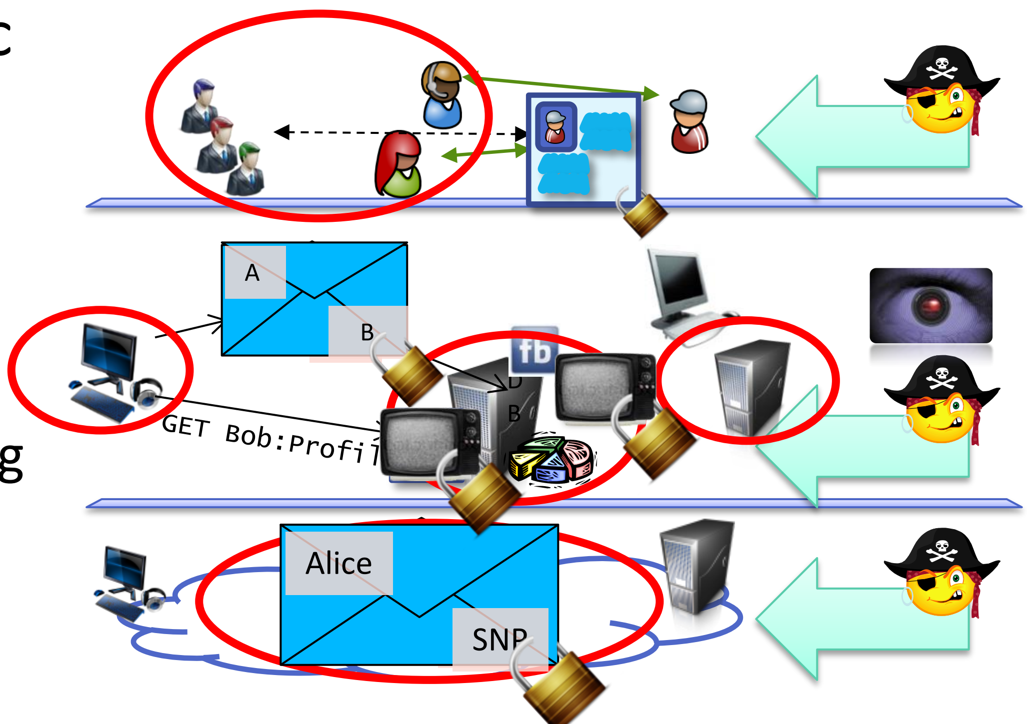


Model and Potential Adversaries



Solution Classes

- Trust „everybody“
- Suspect Network
  - Transport Layer Security
- Suspect subscribers, public
  - Trust provider (& affiliates)
  - ➢ Apply OSN Access Control
- Suspect affiliates/browser
  - Access abuse, unsolicited msg
  - ➢ Web security, Sandboxing..
- Suspect provider & affiliates
  - ***Aim:*** *Content confidentiality*
  - ➢ *Crypto Schemes* (Scramble, NOYB)



Solution Classes – ctd.

- Suspect provider and affiliates
  - Objective: anonymity, behavioral privacy
  - ➢ Decentralization
    - Distribute data and control

diaspora*

PeerSoN

Safebook – Privacy through Decentralization

- ***Centralized*** service identified as vulnerability

- *Safebook: Secure Social Networking through decentralization*
  - Remove centralized instance
  - Distribute storage and control

  - Decentralization requires: controlled access, trust, availability, discovery

  - Friends in social networking services trust each other in the real world
    - Leverage existing „social trust“ to encourage ***cooperation***
    - ***Data replication*** at trusted nodes to facilitate availability
    - Suspect all other service providers: encrypt everything (PKC)

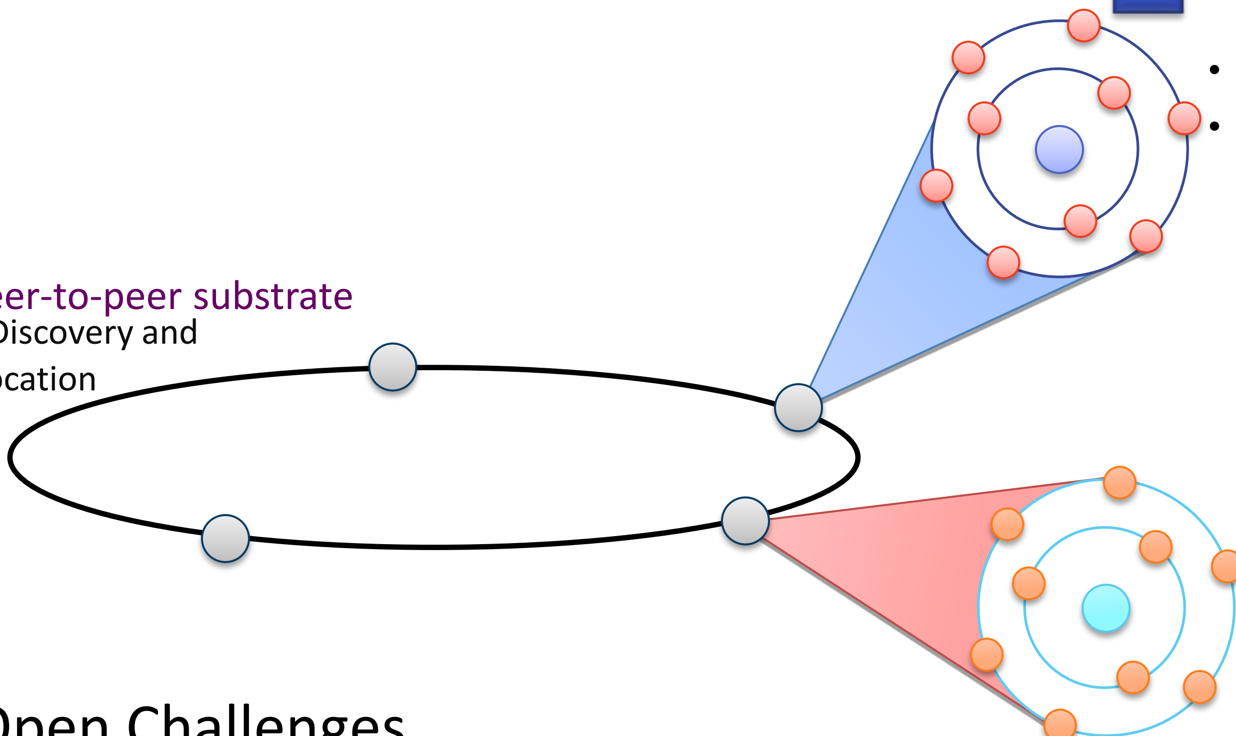
Safebook

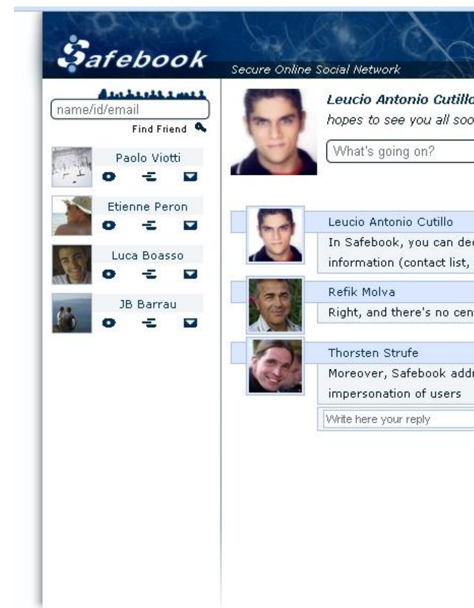
1 Matroshka

- Storage of data
- Cooperative Anonymization

2 Peer-to-peer substrate

- Discovery and Location





- Open Challenges
  - ***Performance*** is insufficient
  - ***Availability*** questionable (correlated churn)
  - ***Concealed participation*** impossible

Social Overlays (“Darknets”)

- *Decentralized OSN don‘t achieve what we want…*

- Stricter requirements
  - Anonymity/ Pseudonymity (sender and receiver)
  - Hidden participation (no 3rd party disclosure: hidden „friendships“)
  - Efficient discovery and interactive communication

- Concepts
  - Connectivity constraints: mutual trust in RL
    - Overlay reflects social trust graph, topology is fixed
  - Information containment: source rewriting, mixing
  - Addressing and routing
    - log / polylog expected routing length required
    - Structured overlays: ***(1) choose ID, (2) choose neighbors***
    - ***(2) is restricted .. adapt (1)***

Network Embedding

A ***network embedding*** on an undirected graph $G = (V, E)$ is a function

$$ID : V \rightarrow M$$

to a metric space M equipped with a distance

$$d : M \times M \rightarrow \mathbb{R}+ .$$

For a node u ∈ V, *ID(u)* is the identifier of *u*.

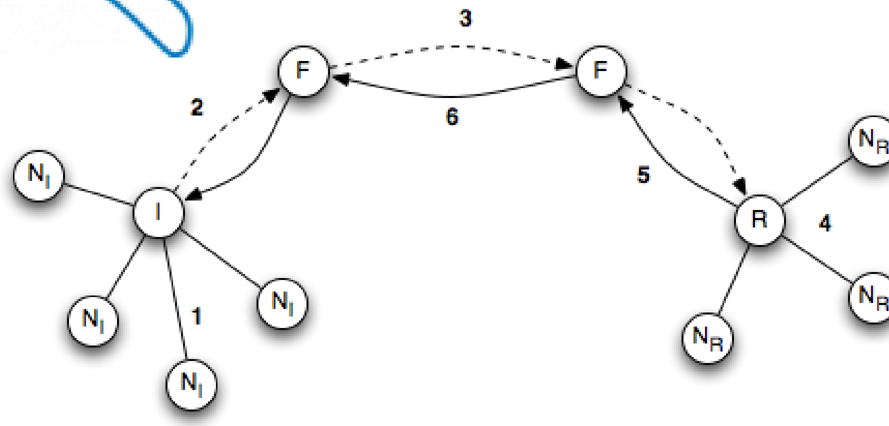
- ***Greedy embeddings***

  guarantee greedy routing success (for every distinct node pair *s*,*t*: *s* is connected to or has a neighbor that is closer to *t*).

- ***Goal:***

  find a decentralized algorithm that approximates a greedy network embedding

The Dark Freenet



- Only deployed (used) darknet

- Assumptions:
  - Social graphs are small world, power law
  - Kleinberg

- Approach:
  - Find embedding of nodes into Kleinberg-like topology (namespace: [0,1))
  - Simulated annealing to approximate lattice with additional long-range neighbor L_u for each node *u:* $P(L_u = v) \propto \frac{1}{d(u,v)^d}$
    - Periodic random sampling of node pairs
    - Comparison of neighborhoods: $c(u,v) = \frac{\prod_{i \in N(u)} d(ID(u),ID(i)) \ \prod_{i \in N(v)} d(ID(v),ID(j))}{\prod_{i \in N(u)} d(ID(v),ID(i)) \ \prod_{i \in N(v)} d(ID(u),ID(j))}$
    - ID swap with probability: $min\{1,c(u,v)\}$
  - Embedding not greedy, adapted routing (DDFS)

Embedding: Attacking Freenet

- Vulnerabilities: Unattested
  - Request period, source of random walk, TTL
  - ID, neighborhood (arbitrarily bad)
- Ad-hoc attacks:
  - Randomize (all IDs constantly)
    - Pretend having random ID, distant neighbors
  - Contract (all to target ID)
    - Pretend having target ID, distant neighbors

- Simulate
  - 10k users
  - 1% adversaries

single adversary

- Results:
  - Hit Ratio

Attack Type	Immediate attack		Attack after convergence	
	R	H	R	H
Randomize	24%	21%	32%	22%
Contract	27%	22%	32%	31%

No adversary: 60%

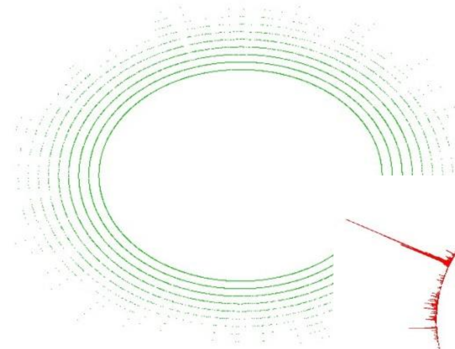
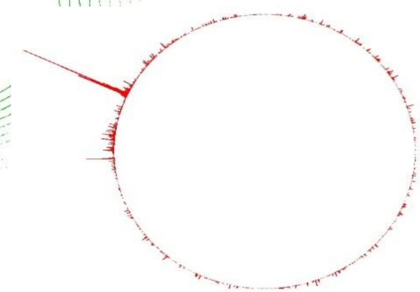
random embedding: 21%

Embedding: A Defense – LMC

- Aim: minimize influence of adversaries:
  - Initiating/faking swap requests
  - Impact of neighborhood

- Adapt own ID based on trusted neighbors only
  - Node *v* selects new ID at random
  - New ID accepted with probability $min\{1,c(v)\}$

- Adversary: only fake own ID
- Reduces diversity, yields slow collaps

- Results
  - Hit Ratio

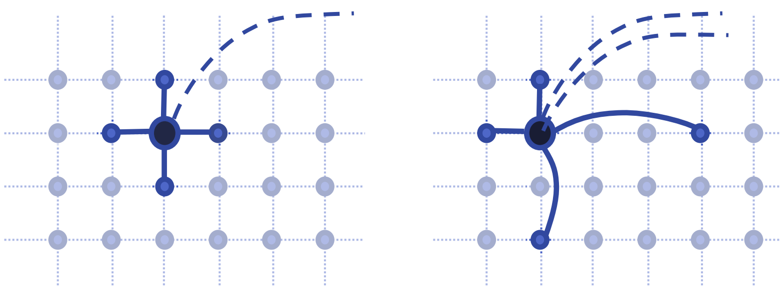
Attack Type	Immediate attack		Attack after convergence	
	R	H	R	H
Randomize	59%	59%	62%	62%
Contract	60%	57%	60%	59%

No adversary: 60%

random embedding: 21%

Routing: Extending Kleinberg‘s Model

- Observe: *Perfect lattice not achieved*
- Extend Kleinberg:
  - Max. distance to closest neighbor ≠ 1
  - Multitude of long range neighbors

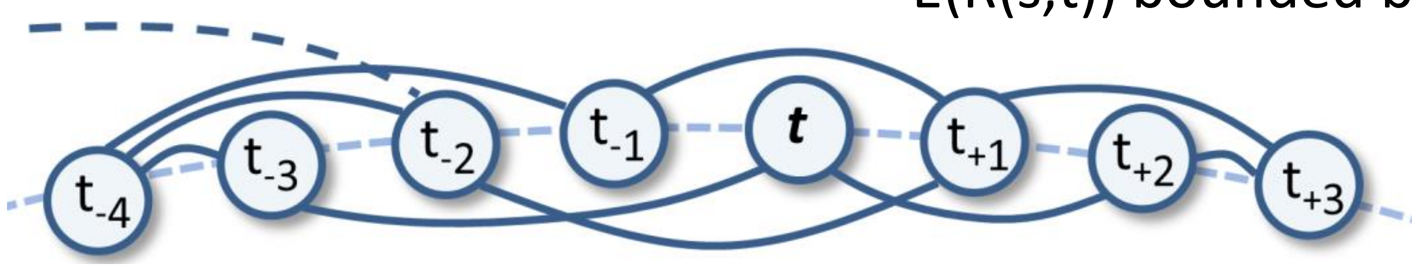


  - *K‘(n,d,C,L)*
    - n^d nodes in d dimensional lattice
    - $C \in \mathbb{N}$: max distance to any node‘s closest neighbor
    - *L:* distribution of long-range links

Routing: Freenet not polylog

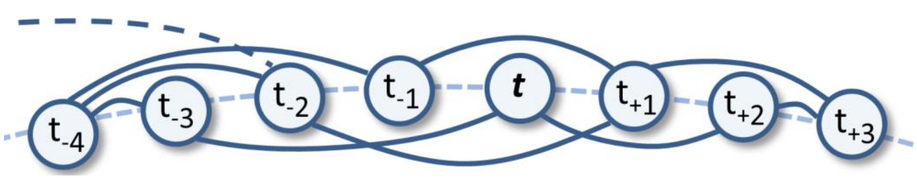
- Routing: *Distance-directed depth first search*
  - Forward to neighbor closest to t *that has not received the message before*
  - Backtrack when no neighbor left
  - „On backtrack“: *next closest* neighbor
- *„Try best node that has not received the message before…“*

- *Proof idea* ($C>2$, bounded L):
  1. Adverse scenario: local routing unsuccessful, long range link taken
  2. Success only on backtrack or other long-range link
  3. P_1 linear, P_2 in polylog steps negligible
- Result:
  - E(R(s,t)) bounded by $\log^{\rho}$ n



Routing: Achieve polylog – NBO

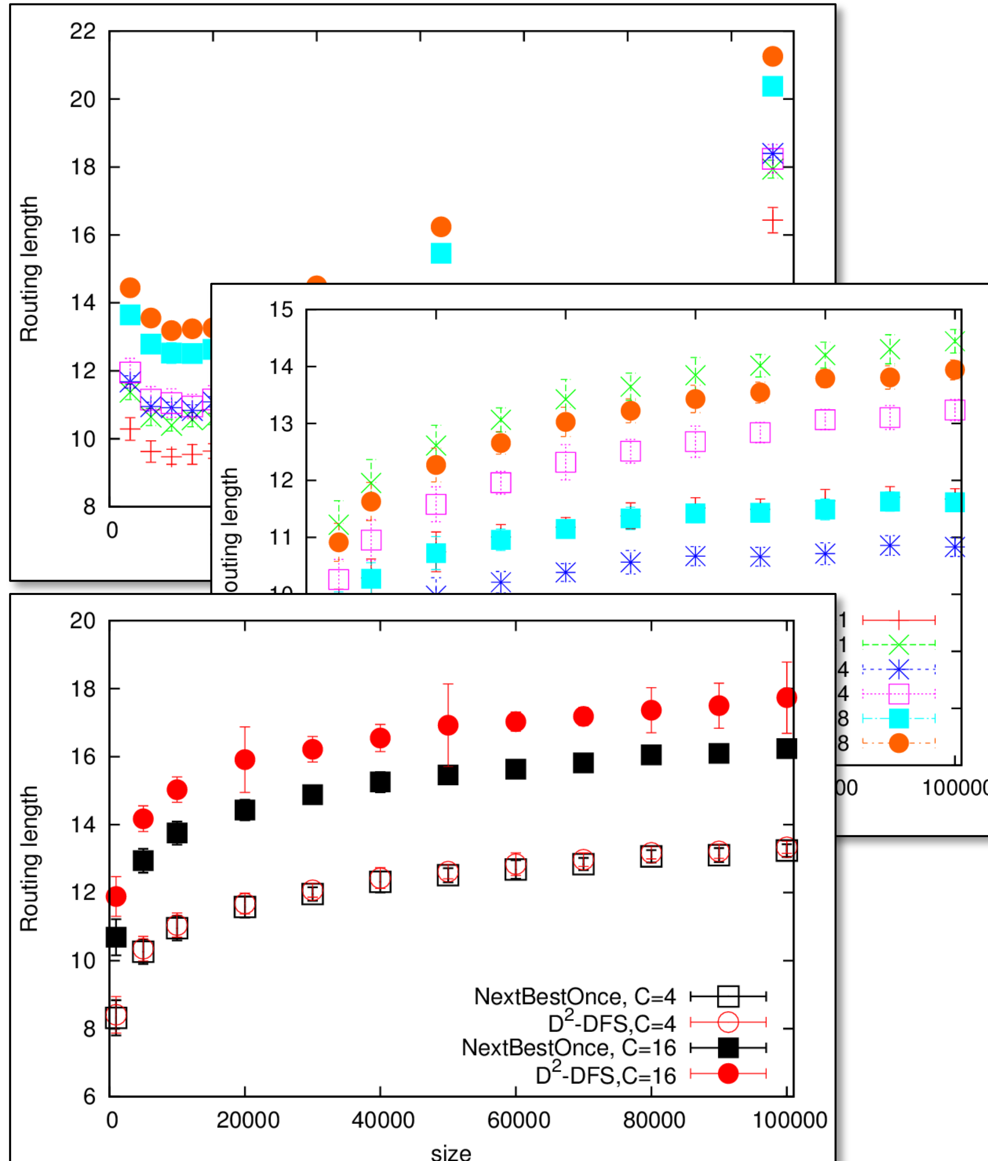
- Rationale: stick to *C-neighborhood* of *t*
- Idea:
  - Revisit nodes until *all neighbors closer* to *t* visited
  - (Signal exhausted nodes in Bloom Filter)



- *Proof idea*:
  - R_1, get „close“
  - R_2, get within C-neighborhood
  - R_3, get to *t*
- R_1, R_2: polylog, halve distances in each step
- R_3: message not passed to long distance node
  - *(Proof rather technical cf. paper)*
- Result:
  - E(R(s,t) = $O(\max\{\log^{\alpha-1} n \log\log n, C^2 \log^{\alpha-1} n\})$

Asymptotic results.. Check with Simulations

- *„Are we nearly there?“*
- G ∈ K‘ (n ∈ {1k, 1mio}, C =[1..10,16,32])
- $R^{DDFS}(s,t)$, $R^{NBO}(s,t)$
- 30 runs each



GTNA
Graph-Theoretic Network Analyzer

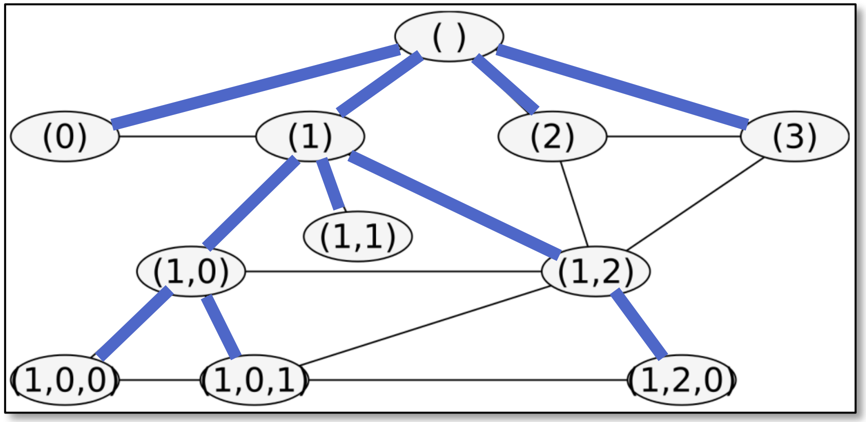
- *We‘re not. :-)*

Embedding Revisited: Trees

- Large C yields too long paths
- Recall: *greedy embedding*
- Highly connected graphs cannot be greedily embedded, but:
- ***Trees*** can:
  - Hyperbolic space
  - High dim. euclid. space
  - Max-norm space (Herzen ’11)

A tree embedding

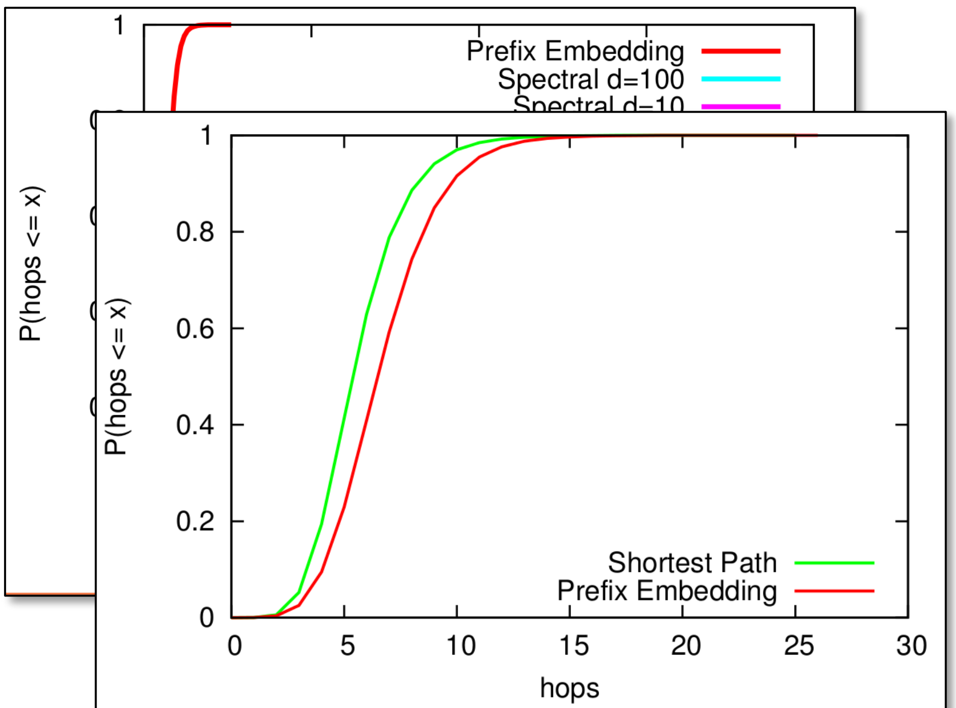
1. Find spanning tree
2. Enumerate children



- $d(s,t) := |s| + |t| - 2|(\text{matchingprefix}(s,t)|$
- (|.|: length of coordinate)

Preliminary Results

- TE achieves greedy embedding
- PGP-WoT; DDFS, Greedy



- Issues:
  - Content Addressing
  - Vulnerabilities:
    - Spanning Tree, embedding
    - „Friendship“ disclosure
- Advantages:
  - Fast enough

Outlook

- ***Need for private communication is evident.***

- Social Overlays represent ***one*** solution class
  - Approximate embedding w. adapted routing
    - Better privacy, low performance
  - Greedy embeddings of spanning trees
    - High performance, lower privacy

- Towards Dark Social Networking Services

there's a long road ahead of us