Decentralized Online Social Networks

Advanced Computer Networks

Summer Semester 2012

Online Social Networks (OSN)

- Online service that allows humans to express social relations, e.g.:
  - Friendship
  - Recommendations
    - Restaurants
    - Movies
  - News
  - Business Contacts

- Typically web-based, but sometimes e-mails, instant messaging, …

Popular Examples

- Facebook
  - 800M users as of Sep. 2011
- Twitter
  - 500M users as of 2012, 340M tweets daily
- LinkedIn
  - 150M users as of Feb. 2012

All still trying to acquire more users!

Potential Problems?

- Technically most platforms work stable
- BUT:
  - Networks already have insights to the “life” of approx. 10% of the global population
  - This data might be misused
    - Facebook Beacon that leaked shopping information
    - Password leakage of millions of passwords of LinkedIn users
    - Business model based on advertisement
    - Data not encrypted

Research Vector

- Removal of the central control unit

- Decentralization of control and data storage!
- New questions:
  - Where to store the data?
  - How to perform lookup?
  - How to secure the data?

General Concepts

- Most solutions are based on insights from the classical P2P filesharing world
  - DHTs
  - Data replication
  - Encryption

- Considering replication for constant data availability:
  - Where to store the data? How to select the replica nodes or storage in a network?

Persona – A Secure OSN

- Decentralized online social network with accessible webspace for data storage
- Key element of the proposal: data privacy by advanced encryption
- **Attribute based encryption and traditional public key cryptography**

R. Baden, A. Bender, N. Spring, B. Bhattacharjee, and D. Starin, “Persona: An Online Social Network with User-defined Privacy,” in SIGCOMM ’09. ACM, 2009.

Persona – Security Model

- Each user generates public/private key pair
  - Distribution of public key is “out-of-band”

- Persona allows to encrypt to groups.
  - All members should be able to encrypt and decrypt data in the group:
  - Traditional way: generate symmetric key for the group and distribute via public keys.
    - But: Problems with colluding group members. Encryption specification to match group “neighbor” and group “football” is impossible.

Persona – Security Model

- Attribute based encryption:
  - Each user requires two keys: an ABE public key (APK) and a master secret key (AMSK).
  - ABE allows to generate a ABE secret keys that incorporate multiple attributes, e.g., being a “co-worker” and a “football-fan”. Bob automatically joins the two groups.
  - Allows to encrypt to “co-worker” OR “football-fan” and to “co-worker” AND “football-fan”.
- ABE is about 100-1000 times slower than RSA

Persona – Key Management

- Each user is identified by self chosen public key.
- ABE groups easily implement the attribute “friend”: Alice computes

$$C = \mathsf{TEncrypt}(Bob.TPK, K)$$

with $K = Alice.ASK_{\text{‘friend’}}$ and

$\mathsf{TEncrypt}(K, m)$	RSA encrypt m with key K

C can be uploaded to any webspace, retrieved and only decrypted by Bob. With K he joins Alice’s friend group!

Persona – OSN

- OSN application using a wall (based on a file system called “Doc”), chat and status update
- Primarily based on writing all info to file for ABE encrypted groups
- Users periodically check the “Doc” files for updates
- Chat is just a document with continues updates (appended chat text)

Persona – OSN

- Persona is implemented as a keystore in Firefox (considered a trusted component)
- The Firefox component performs all relevant encryption/decryption methods
- Browser allows integration with Facebook
  - Use Persona to allow Persona-Friends to access encrypted data

Persona – Summary

- Data storage only of minor interest in the paper
  - Assumption of user’s to provide webspace might be overly optimistic

- But: The security level is state of the art. Most OSN follow up papers consider that aspect as solved (or not hot anymore).

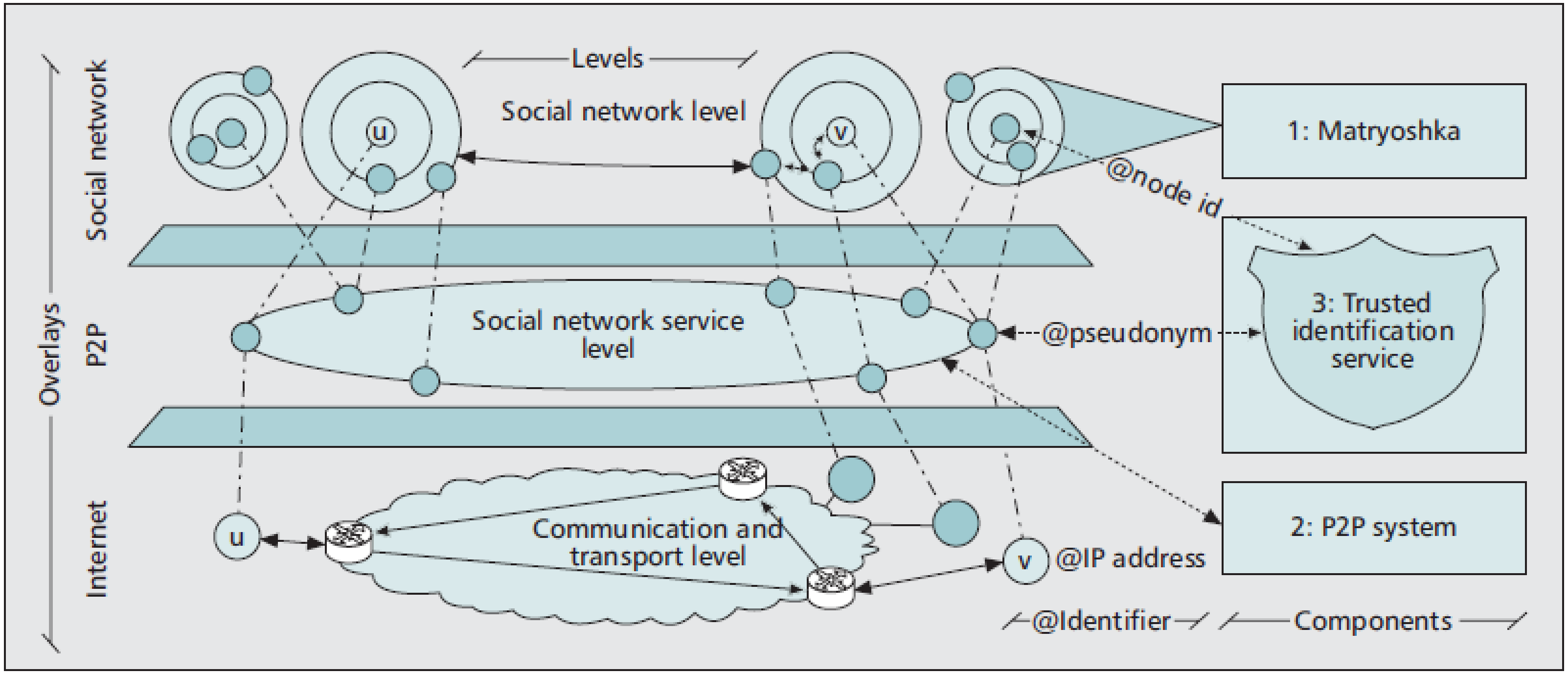
SafeBook

- Based on three objectives:
  - Privacy
  - Data Integrity
  - Availability

- “Safebook: Security Based on Real-Life Trust”
  - Three tiers: The OSN layer, the P2P substrate and the Internet

L. Cutillo, R. Molva, and T. Strufe, “Safebook: A Privacy-preserving Online Social Network Leveraging on Real-life Trust,” Com. Mag., IEEE, vol. 47, no. 12, pp. 94–101, 2009.

Safebook – Overview



- A set of matryoshkas, concentric structures in the OSN layer provide data storage and communication privacy!

L. Cutillo, R. Molva, and T. Strufe, “Safebook: A Privacy-preserving Online Social Network Leveraging on Real-life Trust,” Com. Mag., IEEE, vol. 47, no. 12, pp. 94–101, 2009.

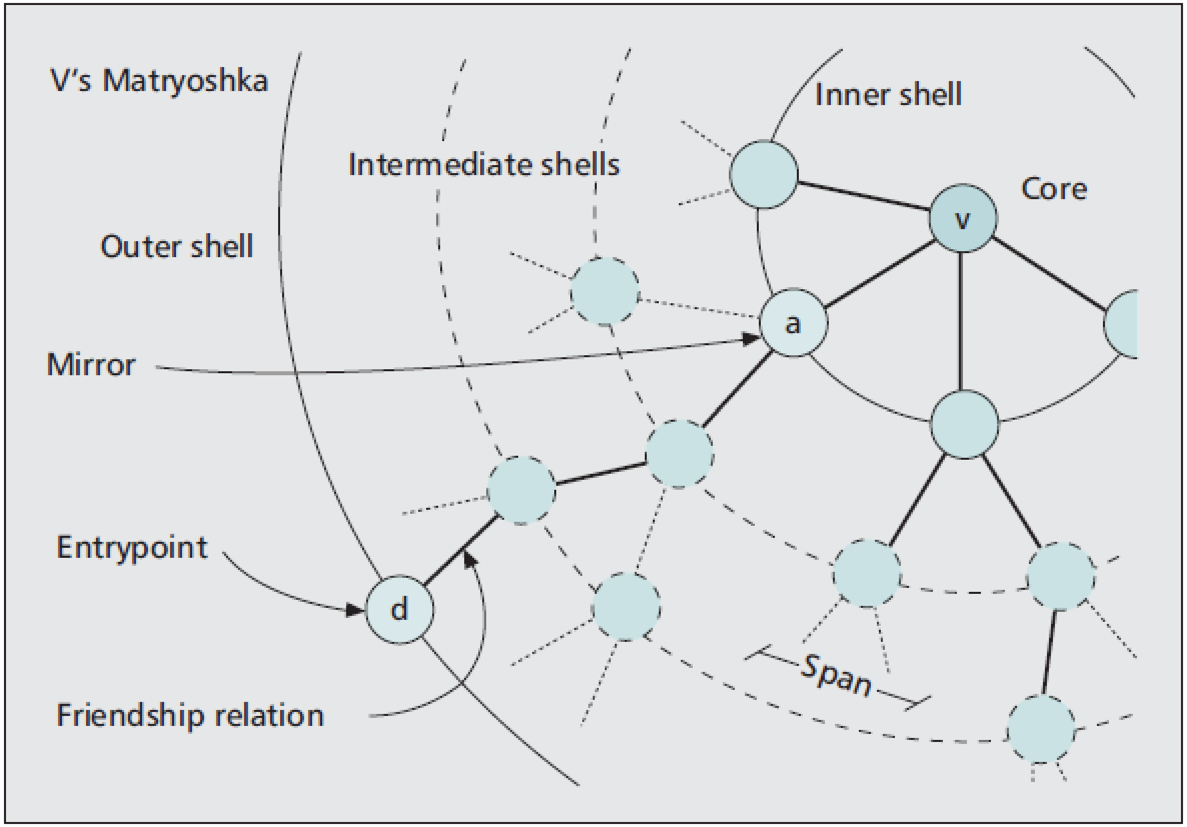
Matryoshkas

- Each matryoshka protects the core of the ring
- Messages to the core (the node) traverse through the "shells"
- From the paper: *"The innermost and outermost shells of a matryoshka have a specific role: the innermost shell is composed of direct contacts of the core, and each of them stores the core's data in an encrypted form."* -> These are data mirrors

http://en.wikipedia.org/wiki/File:Russian-Matroshka.jpg

Data Access

1. U requests data from pseudonym v.
2. Lookup returns outmost shell



3. Each shell forwards request and hides the origin (like Onion Routing)

Safebook – Data Availability

- The data access is well protected, BUT:
  - Requires traversal of the shells along a path of simultaneously online nodes that befriend each other
  - Replicas are stored at user’s friends
- This limits the data availability with increasing shells
  - With 3 shells 13 replica nodes were required to achieve 90% data availability
  - With 4 shells 23 replica nodes were required to achieve 90% data availability

PeerSoN

- An early P2P based solution employing:
  - *Encryption* to guarantee user data privacy
  - *Decentralization* to be infrastructure independent
  - Direct exchange of data tackling problems of Delay Tolerant Networks (DTNs) or challenged networks

Aimed at leveraging real-life social links

S. Buchegger, D. Schiöberg, L.-H. Vu, and A. Datta, “PeerSoN: P2P Social Networking: Early Experiences and Insights,” in SNS ’09. ACM, 2009.

PeerSoN – Security Model

- Assumption of public-key infrastructure (PKI)
  - Includes the possibility of key revocation
  - Public keys of the friends are available for encryption
  - Peers vouch for each other using “certificates”
- For additional identity theft prevention a challenge-response protocol with friends is proposed

PeerSoN - Architecture

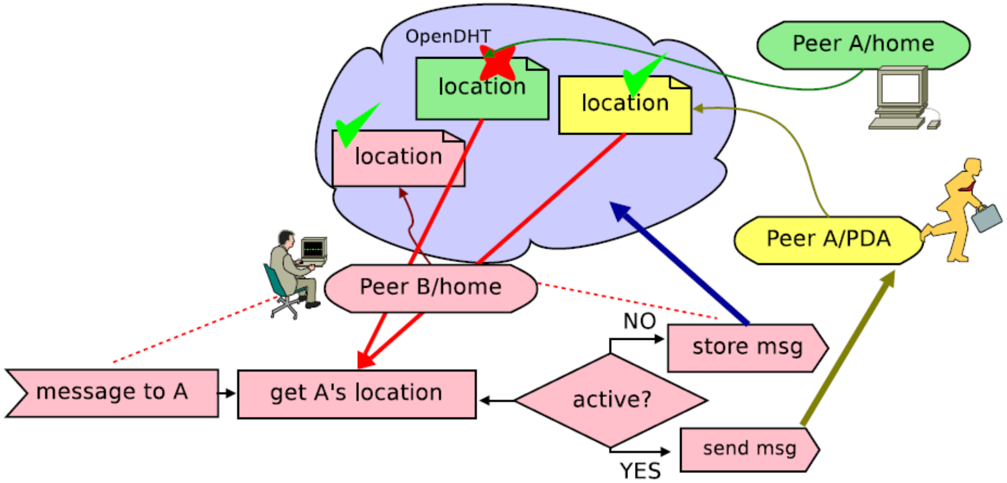
- No central nodes with access to all (even encrypted) data items.

- Data is ordered in a “Digital Personal Space” for wall, pictures etc.

- Strongly incorporating data forwarding elements from the DTN world. Direct contacts allow data exchange.

PeerSoN - Architecture

- Lookup service stores meta-data (find users and data)
- DHT as key/value pair lookup
- Peers directly interconnect

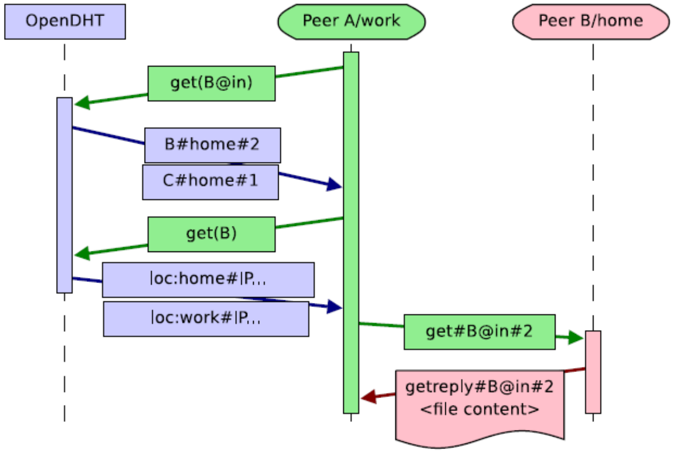


Message flow from B to A

PeerSoN - Architecture

- Data retrieval:
  1. Location lookup of latest index file
  2. Determination of method to establish connection

- It is not required that the data is stored at the data owner!



Basic file request of B@in

PeerSoN - Storage

- Data availability should be maxed in a system where peers store data for each other

- Idea: Place data to cliques of peers that replicate for each other.

- Key: “Firstly, the peers are sorted by non-increasing availabilities av(i).”

K. Rzadca, A. Datta, and S. Buchegger, “Replica Placement in P2P Storage: Complexity and Game Theoretic Analyses,” in IEEE ICDCS’10, June 2010.

PeerSoN - Storage

- Cliques are formed between peers with very similar data availability
- Peers are selected from a mixture of a “random pool” and a “metric pool”
- The metric pool contains potential candidates with matching availability. If a current replica gains a low “score”, a switch is initiated

- Achieved availability: <90 to 100% based on the nodes own availability.

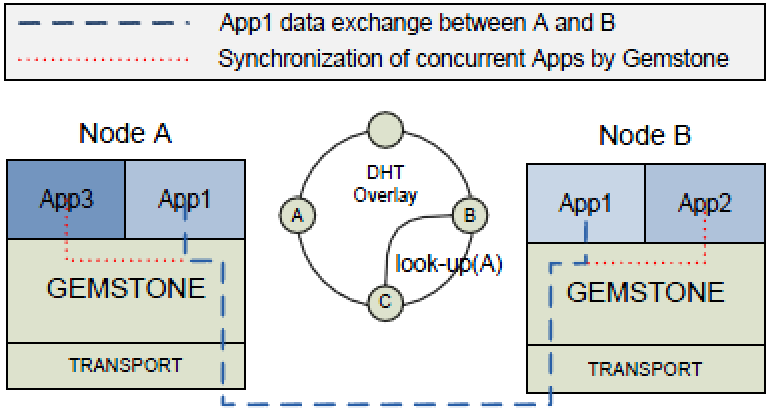
Gemstone

- Focuses on data availability optimization
- Basic assumptions:
  - Users store data in a peer-to-peer manner or on altruistically provided space
  - Storage is unstable
  - Online patterns follow typical social network behavior instead of file-sharing behavior (as e.g. in PeerSoN)

F. Tegeler, D. Koll, and X. Fu, Gemstone: Empowering Decentralized Social Networking with High Data Availability. ;In Proceedings of GLOBECOM. 2011, 1-6.

Gemstone – Basic Structure

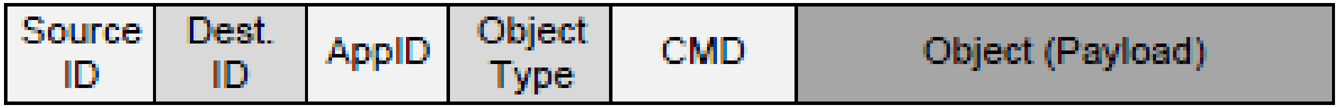
- Gemstone is an overlay system for Online Social Networking applications
- It provides
  - Social graph mngtmnt
  - Data delivery (msg)
  - Profile storage



- The key element is the efficient storage of data replicas

Gemstone - Security

- IDs are public keys
- All data items are stored in Gemstone objects with encrypted payloads:



Source ID	Dest. ID	AppID	Object Type	CMD	Object (Payload)

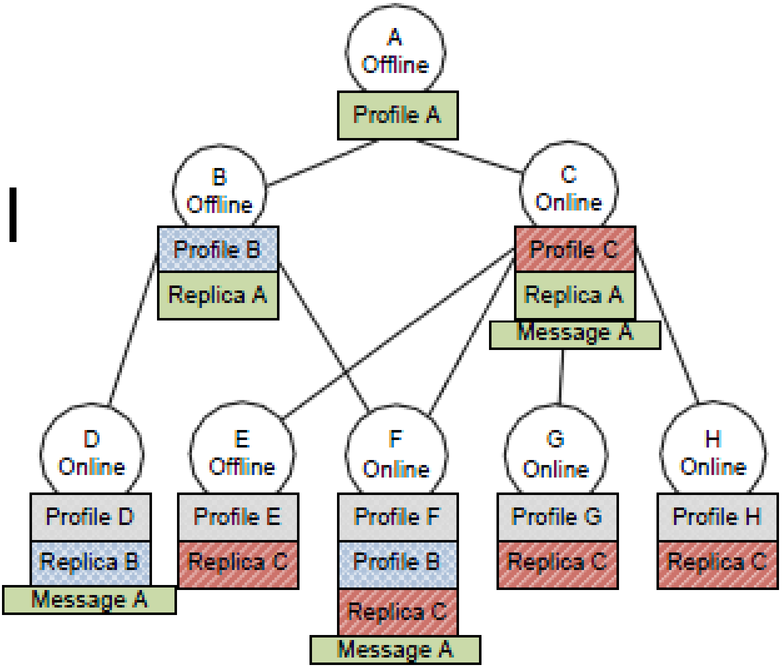
- Encryption is done via Attribute Based Encryption and follows Persona

Gemstone – Store and Retrieve

- The profile information is a Gemstone object as well

- If the user is online, all information is directly retrieved

- If the user is offline, the replica nodes are requested to deliver the profile.



Gemstone – Replica Selection

- The problem now shifts to an intelligent selection of replica nodes

- Three key factors of a candidate replica node are currently considered:
  - The normalized online time (the more the better)
  - The social relation to that node (friends are more likely to not drop the data)
  - The previous, personal user experience (positive past behavior is a good sign)

Gemstone – Replica Selection

- Choosing the optimal nodes with a fixed n would be an NP-complete (Knapsack) problem.

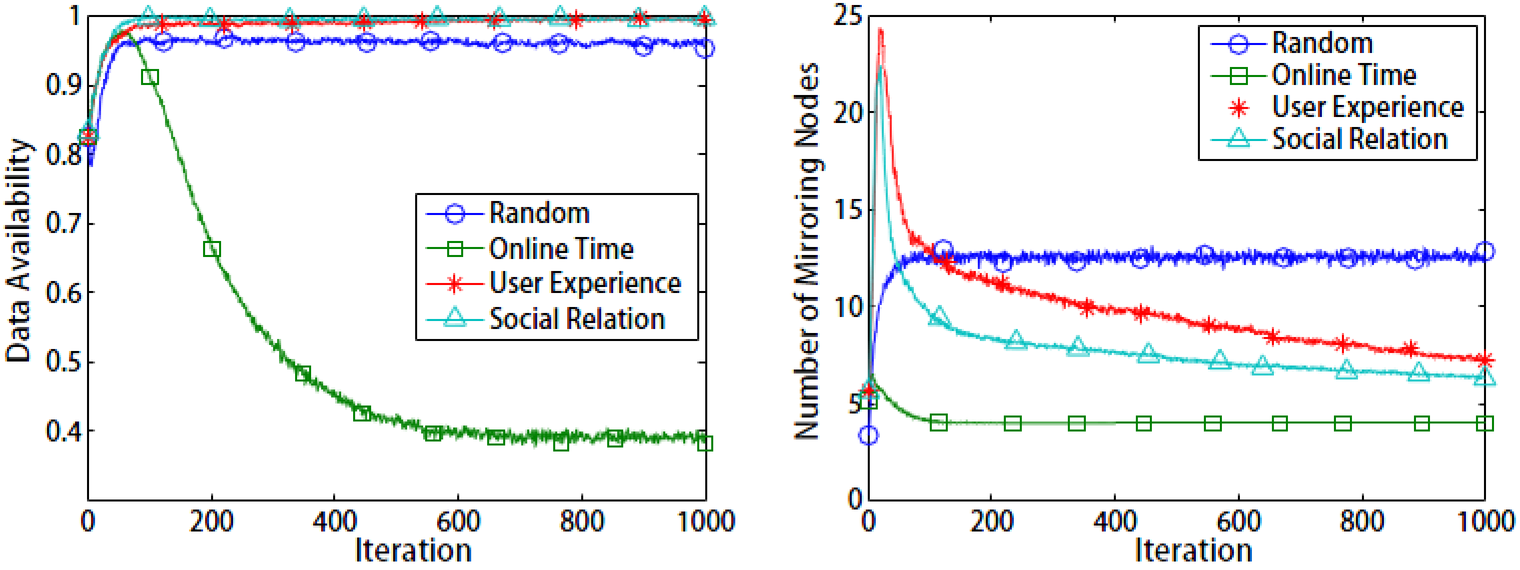
- The system estimates how many nodes to choose to achieve 99% availability:

$$\prod_{i=1}^{n}(1-p_i) < 0.01$$

- Each node selects an individual number of replicas!

Data Availability

- Fast converges to data availability >99% with 6-8 replicas…



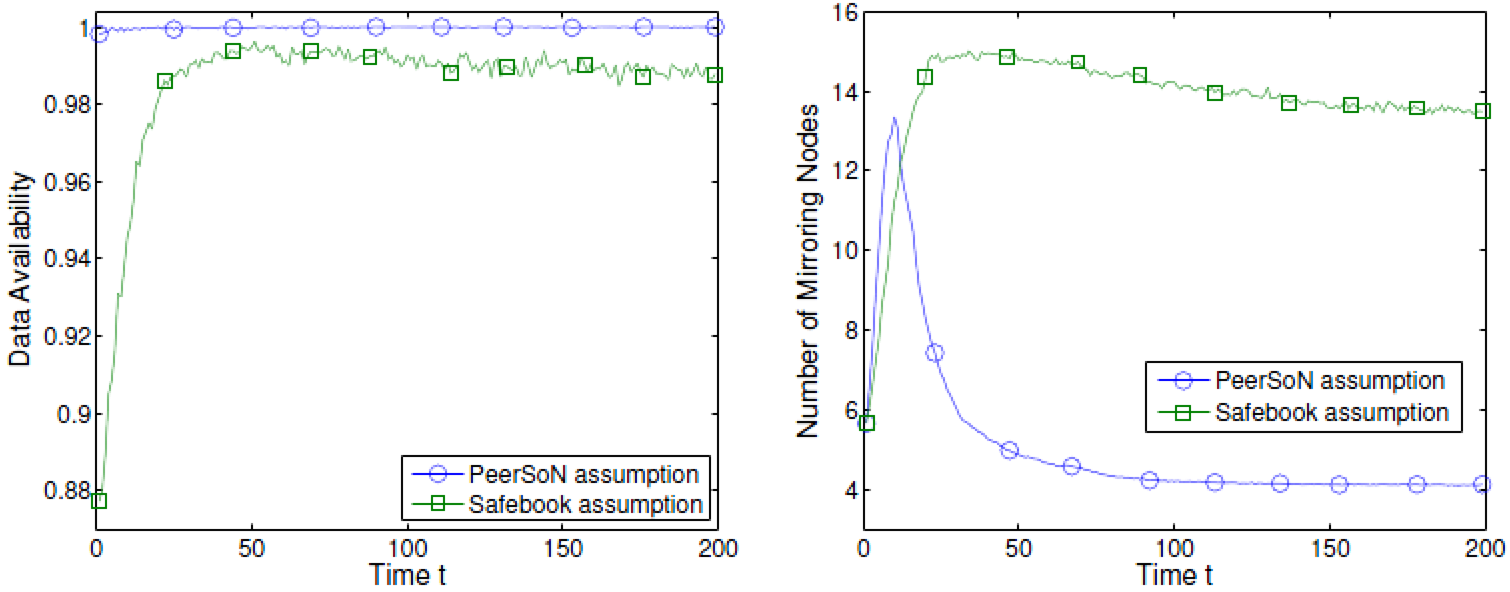
(a) Data availability by strategy

(b) Number of DHAs by strategy

Fig. 4: Simulation Results

Online Time Assumption Impact

- Remember: Safebook required 13 replicas for 3 shells to achieve 90% availability…



(a) Data availability under PeerSoN and Safebook assumptions

(b) Number of DHAs under PeerSoN and Safebook assumptions

Fig. 5: Simulation results under different assumptions

Gemstone – Recent Advances

- The replica node selection algorithm was significantly improved
  - Friend’s report on the performance of mirrors
  - Candidate set considers diurnal patterns
  - Improved security against attacks on the scheme
- Gemstone advantages:
  - Performs well even for highly inactive nodes
  - Uses all storage available
  - Achieves high availability with low number of replicas

Summary

- Decentralizing social networks has significant advantages:
  - Data privacy and control over data remains at the users level

- Security issues can be solved using Persona like Attribute Based Encryption
- Data storage and replica selection is tricky, PeerSoN, Safebook and Gemstone provide ideas