Peer-to-Peer Networks

Chapter 5: Applications on P2P, Online Social Networks

Thorsten Strufe

Chapter Outline

- Overview
- Distributed Backup and Storage
- Gaming
- Communications (Online Social Networks)
- Content Distribution Revisited (P2P Streaming)

P2P: More Than Filesharing?

- P2P so far focussed on finding stuff
- Main application: File Sharing (exchange of data objects between end hosts)
- It works well and is very popular, since
  - Resources (bandwidth/selection of data) in client-server is limited, scales better with P2P
  - Users have an incentive to participate (download sth fast, it‘s free as in beer)
  - Almost pervasive broadband access, „always on“

  → Difference between clients and servers decreases (again…)

- So what about other applications of the P2P paradigm?

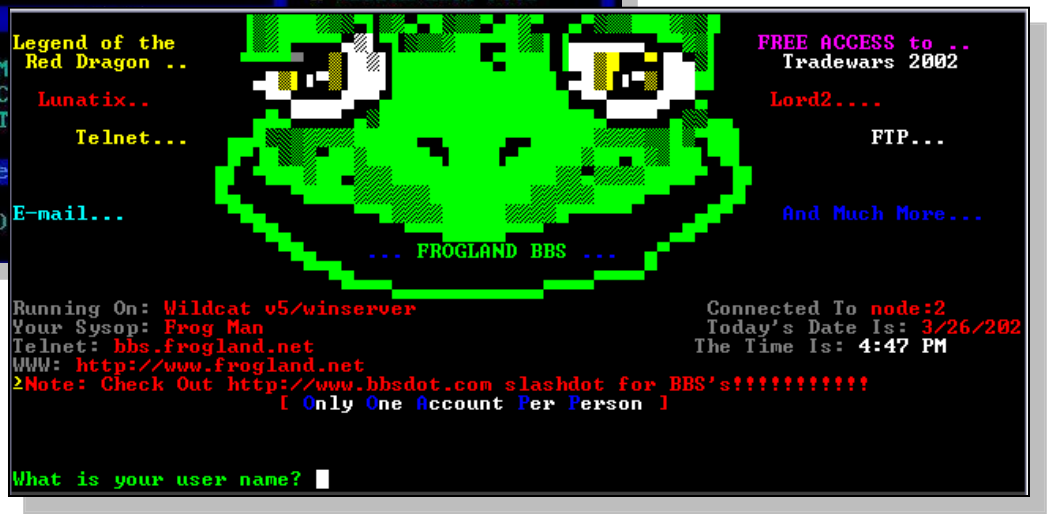
More than File-Sharing!

- Reversing the paradigm:
  - Distributed backup and storage
- Pushing content distribution a bit further:
  - Overlay- / P2P streaming, Application Layer Multicasting
- Distribution of control
  - Groupware / Collaborative work
  - Decentralized online social networks
- Centralized use of P2P
  - Enhancing centralized systems using P2P

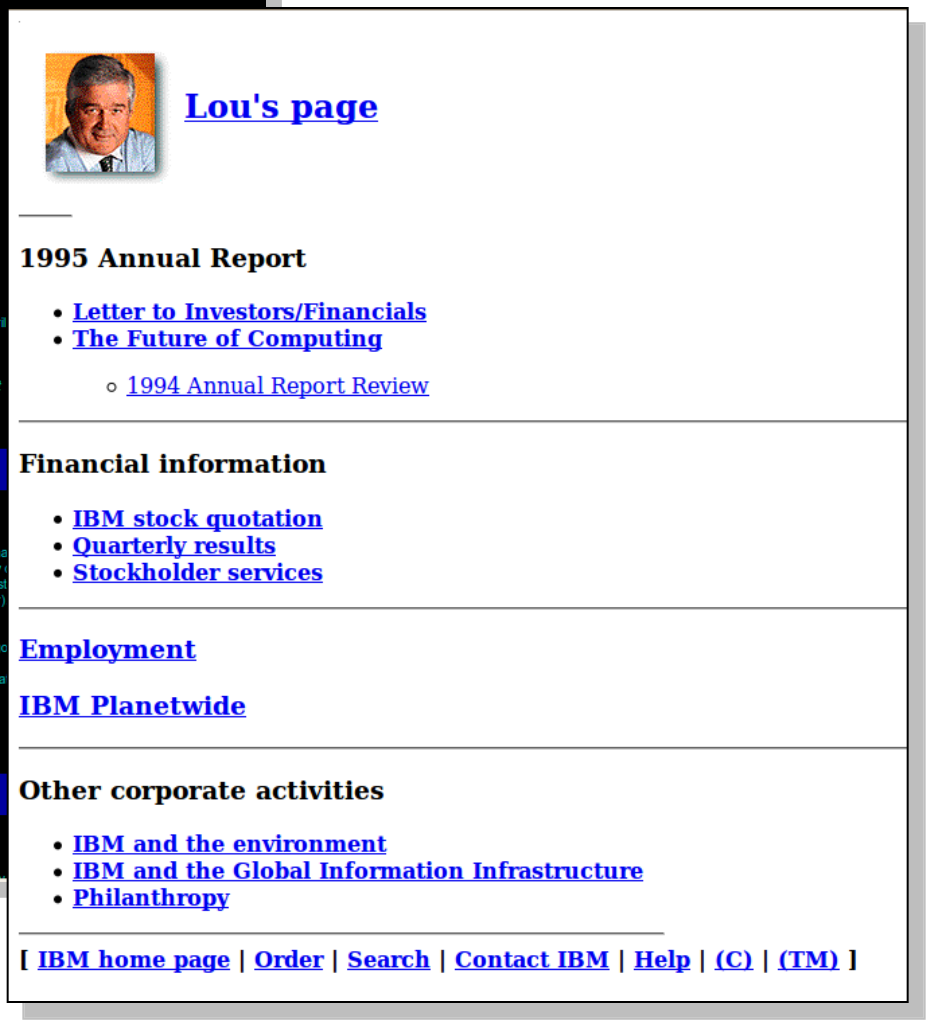
P2P Online Social Networking

- What are Online Social Networks and how have they evolved?
  - A short overview of the history of networked services
- Why are they interesting for research?
  - A peek at their prevalence
- How do people actually use them?
  - Characterizing user behavior in OSN
- The privacy problem
- Movements to decentralize
- Safebook

A Short History of Networked Services…

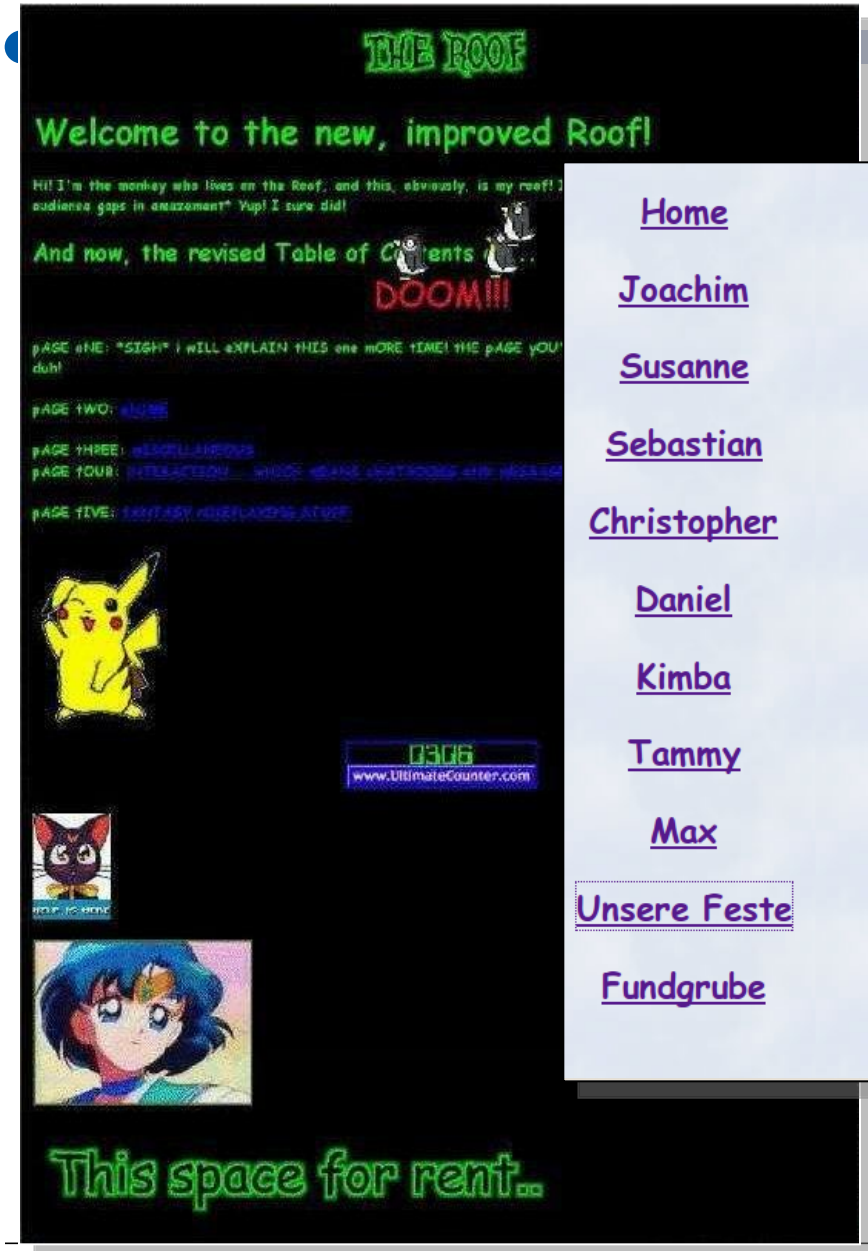


…around came the World Wide Web…



http:www.ibm.com/IBM/ last visited: 1996-10-21

…with more info than we really wanted…



Home

Joachim

Susanne

Sebastian

Christopher

Daniel

Kimba

Tammy

Max

Unsere Feste

Fundgrube

Unsere Feste in Bildern …

Da es sich hierbei um ganz private Feste handelt, sind diese Seiten nur für Familienmitglieder und Freunde zugänglich.

An alle Verwandten und Freunde!
Bitte schickt mir eine Mail, falls ihr eine Zugangsberechtigung wollt.
Ich werde euch dann umgehend die Daten zumailen.

JA!!!
Ich will ein Passwort!!!

Momentan könnt ihr Bilder folgender Feste betrachten:

Sebastian's Konfirmation

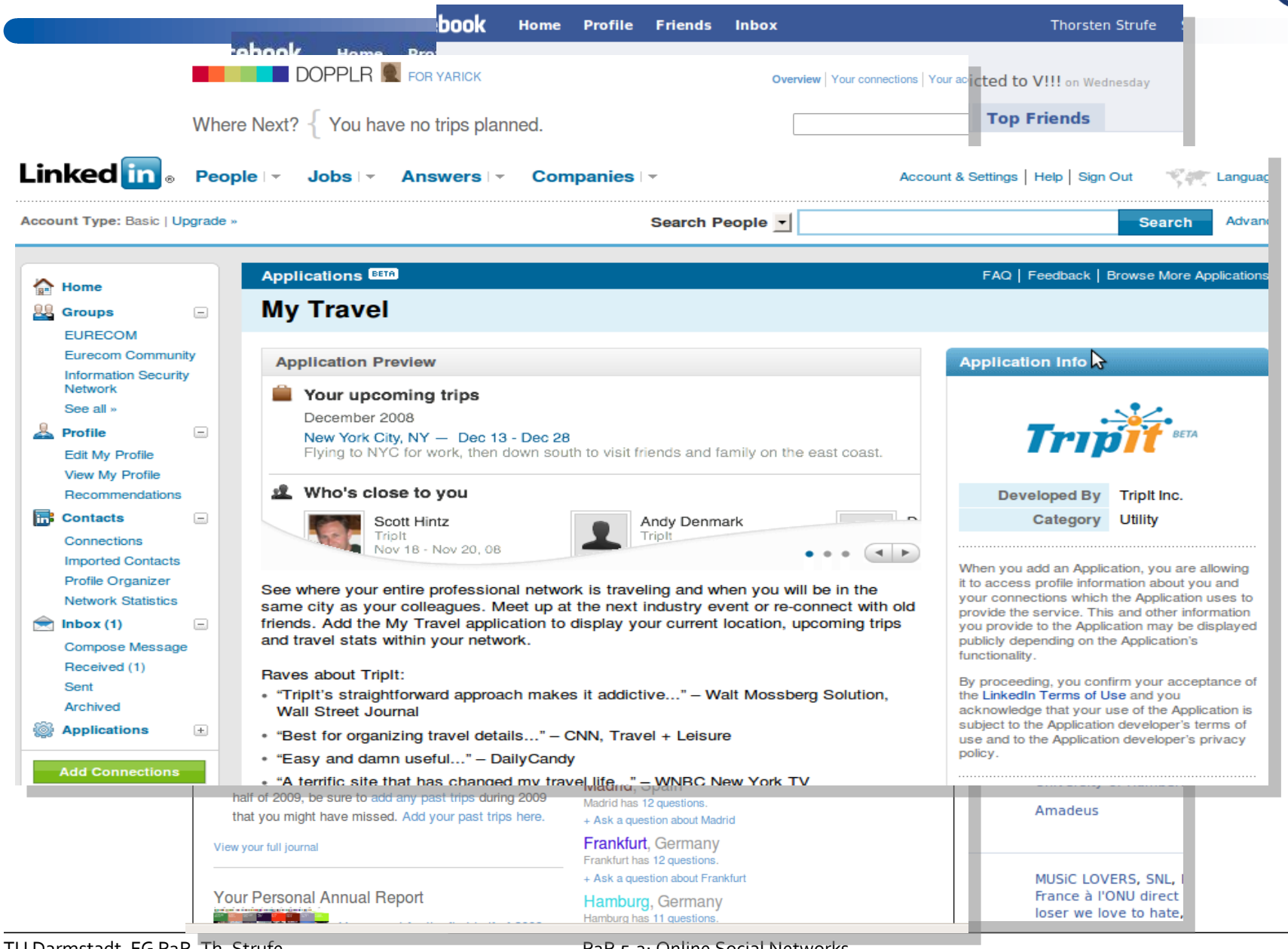
Joachim's 50½ Geburtstag

Otto's 65. Geburtstag

Auf zu den Bildern!

…with high barriers.

So Today *everybody* Shares Some Data…



…with calculated side effects…

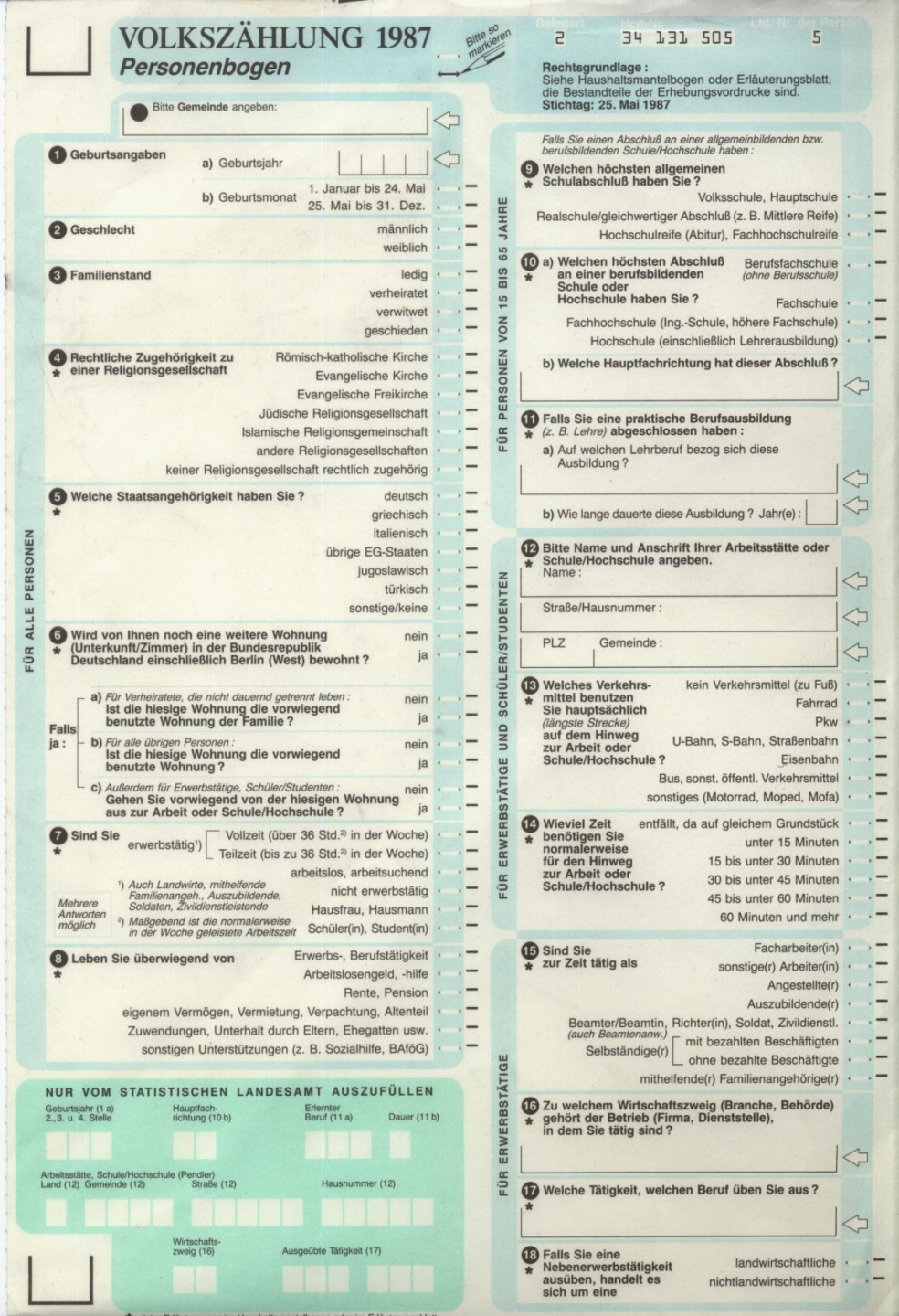
…increasingly immersive to daily life…

loopt

tweetspotting

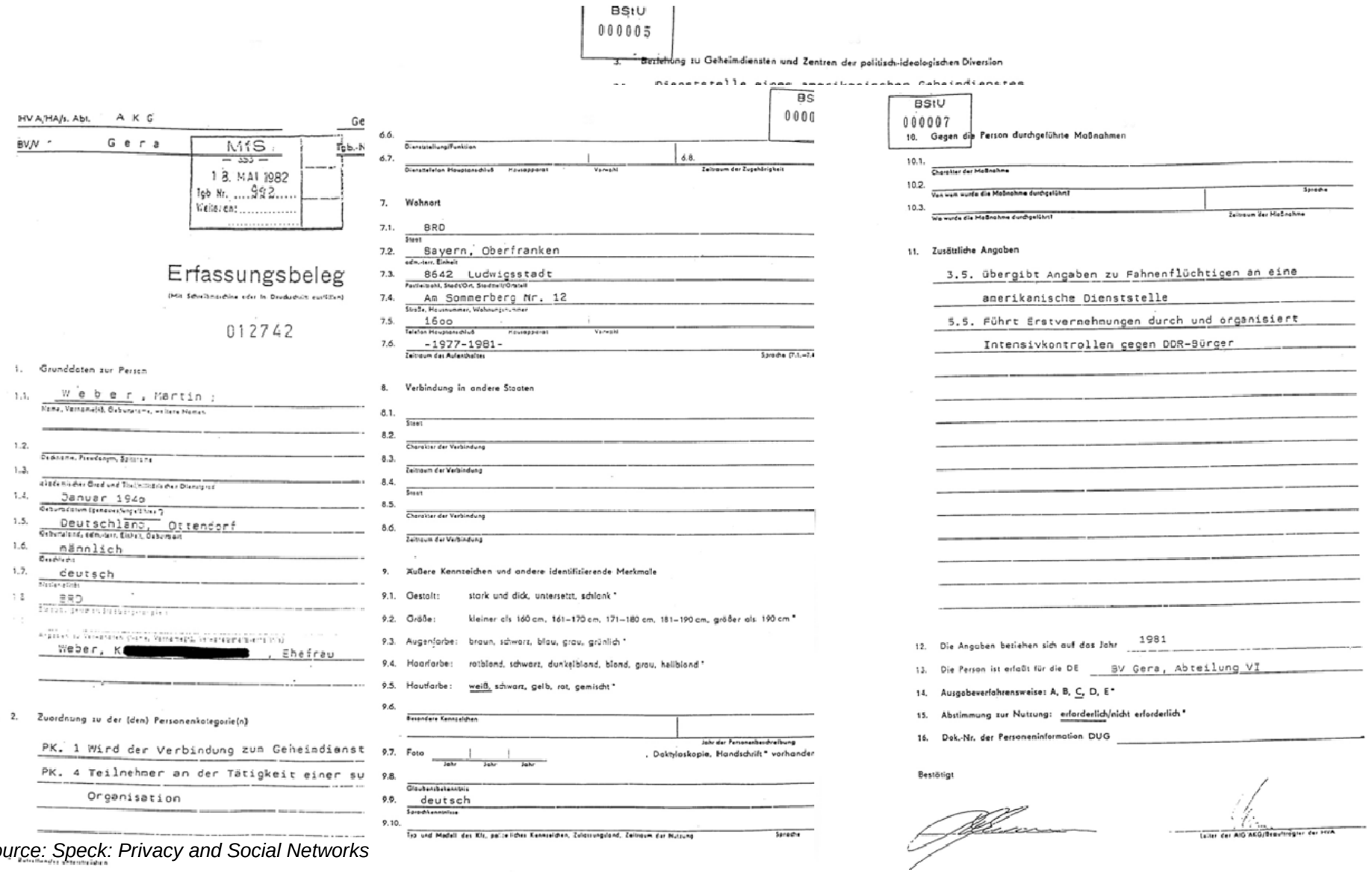
„Volkszählung“ 1987

- Last public census in Germa
- Scheduled for 1981 (delayec
- Significant public opposition
  - Fear of a surveillance society
  - The transparent citizen („gläs
  - Bounty for discovered Germa
  - Appeal for civil disobediance
  - Finally accomplished in 1987
  - Consequence: „25% inherent
  - Significant gap between censu
  - (So let‘s just get a unifying ta



http://upload.wikimedia.org/wikipedia/commons/4/43/Volkszaehlung.JPG

… completely willingly…



BStU 000005

3. Beziehung zu Geheimdiensten und Zentren der politisch-ideologischen Diversion

HVA/HA/s. Abt. A K G

BV/V Gera

MfS
— 355 —
18. MAI 1982
Tgb Nr. 992
Weiterer:

Erfassungsbeleg

(Mit Schreibmaschine oder in Druckschrift ausfüllen)

012742

1. Grunddaten zur Person

1.1. Weber, Martin;
Name, Vorname(n), Geburtsname, weitere Namen

1.2.
Decknamen, Pseudonym, Spitzname

1.3.
Akademischer Grad und Titel/militärischer Dienstgrad

1.4. Januar 1940
Geburtsdatum (genauer/ungefähr)

1.5. Deutschland, Ottendorf
Geburtsland, adm.-terr. Einheit, Geburtsort

1.6. männlich
Geschlecht

1.7. deutsch
Nationalität

1.8. BRD
Staatsbürgerschaft

Angaben zu Verwandten (Name, Vorname(n), Verwandtschaftsverhältnis)

Weber, K[redacted], Ehefrau

2. Zuordnung zu der (den) Personenkategorie(n)

PK. 1 Wird der Verbindung zum Geheimdienst

PK. 4 Teilnehmer an der Tätigkeit einer su

Organisation

Aufnahmeanlaß unterstreichen

6.6. Dienststellung/Funktion

6.7. Dienststellen Hauptanschluß | Hausapparat | Vorwahl | 6.8. Zeitraum der Zugehörigkeit

7. Wohnort

7.1. BRD
Staat

7.2. Bayern, Oberfranken
adm.-terr. Einheit

7.3. 8642 Ludwigsstadt
Postleitzahl, Stadt/Ort, Stadtteil/Ortsteil

7.4. Am Sommerberg Nr. 12
Straße, Hausnummer, Wohnungsnummer

7.5. 1600
Telefon Hauptanschluß | Hausapparat | Vorwahl

7.6. -1977-1981-
Zeitraum des Aufenthaltes | Sprache (7.1.–7.4

8. Verbindung in andere Staaten

8.1. Staat

8.2. Charakter der Verbindung

8.3. Zeitraum der Verbindung

8.4. Staat

8.5. Charakter der Verbindung

8.6. Zeitraum der Verbindung

9. Äußere Kennzeichen und andere identifizierende Merkmale

9.1. Gestalt: stark und dick, untersetzt, schlank *

9.2. Größe: kleiner als 160 cm, 161–170 cm, 171–180 cm, 181–190 cm, größer als 190 cm *

9.3. Augenfarbe: braun, schwarz, blau, grau, grünlich *

9.4. Haarfarbe: rotblond, schwarz, dunkelblond, blond, grau, hellblond *

9.5. Hautfarbe: weiß, schwarz, gelb, rot, gemischt *

9.6. Besondere Kennzeichen | Jahr der Personenbeschreibung

9.7. Foto Jahr Jahr Jahr, Daktyloskopie, Handschrift * vorhanden

9.8. Glaubensbekenntnis

9.9. deutsch
Sprachkenntnisse

9.10. Typ und Modell des Kfz., polizeiliches Kennzeichen, Zulassungsland, Zeitraum der Nutzung | Sprache

BStU 000007

10. Gegen die Person durchgeführte Maßnahmen

10.1. Charakter der Maßnahme

10.2. Von wem wurde die Maßnahme durchgeführt? | Sprache

10.3. Wo wurde die Maßnahme durchgeführt? | Zeitraum der Maßnahme

11. Zusätzliche Angaben

3.5. übergibt Angaben zu Fahnenflüchtigen an eine amerikanische Dienststelle

5.5. Führt Erstvernehmungen durch und organisiert Intensivkontrollen gegen DDR-Bürger

12. Die Angaben beziehen sich auf das Jahr 1981

13. Die Person ist erfaßt für die DE BV Gera, Abteilung VI

14. Ausgabeverfahrensweise: A, B, C, D, E *

15. Abstimmung zur Nutzung: erforderlich/nicht erforderlich *

16. Dok.-Nr. der Personeninformation DUG

Bestätigt

Leiter der AIG AKG/Beauftragter der HVA

Source: Speck: Privacy and Social Networks

Facebook, Twitter & Co. (Online Social Networks)

Online Social Networks (*Social Networking Services)*

"web-based services that allow individuals to

(1) construct a public or semi-public profile within a bounded system,

(2) articulate a list of other users with whom they share a connection, and

(3) view and traverse their list of connections and those made by others within the system."

- Is this really all?

*"the **communication** of participants through **message exchange**, **commenting** on the profiles of others (or previous interactions, e.g. in recommendations), which merely is a message exchange with the aim to annotate the addressed profile, and the wealth of **applications** (starting from simple ``poking'' mechanisms to a variety of ``gift'' and ``likeness'' applications for interactions between users)"*

Source1: boyd et al.: Social Network Sites

Source2: cutillo et al.: Privacy Preserving Social Networking through Decentralization

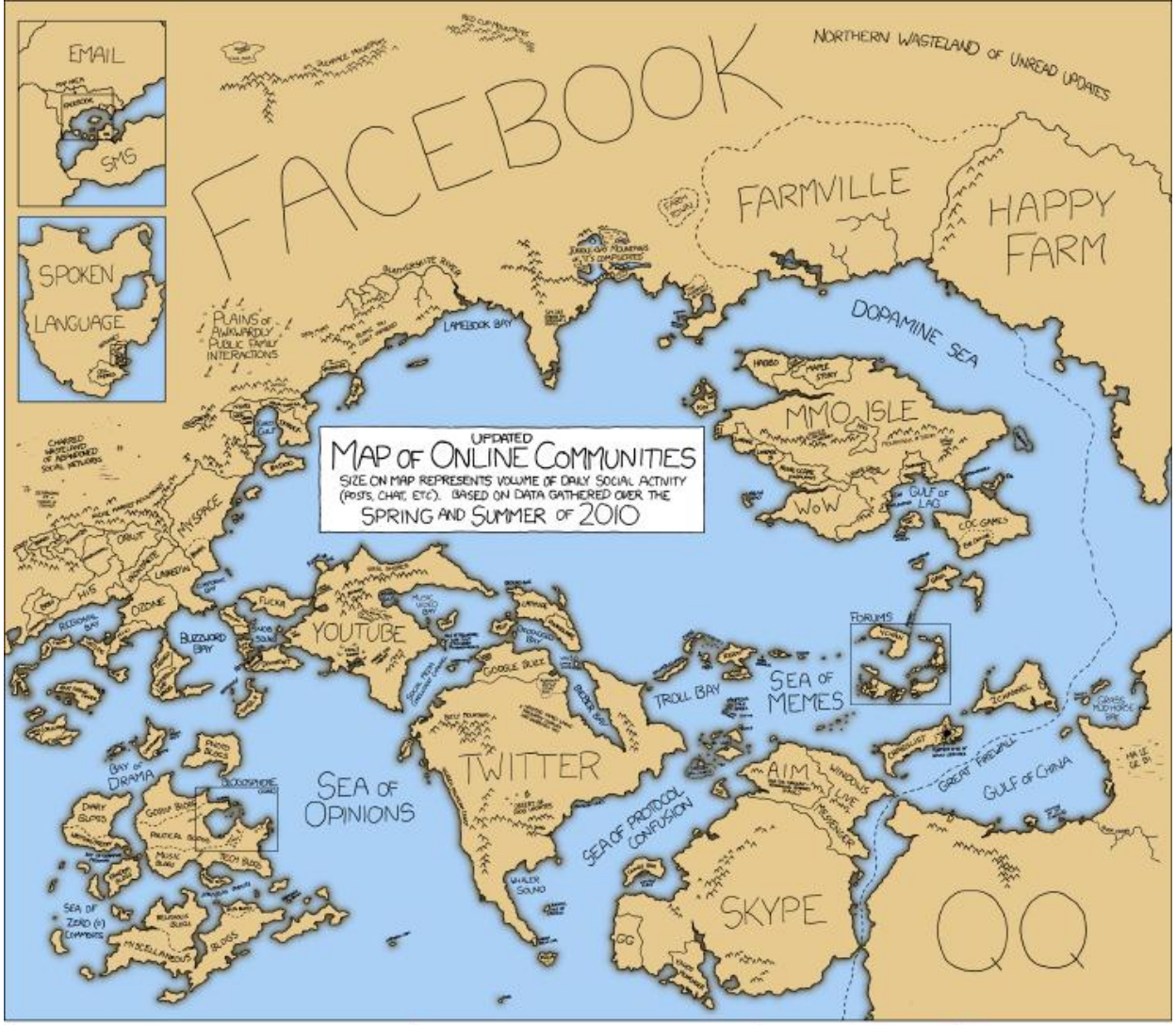
Online Social Networks

- Simplified, walled-garden version of „the Web“:
  - Easy to set-up pages („***profiles***“) of ***individuals*** (… and companies…)
  - Links reflecting ***real-world relations*** between individuals
  - Possibility to share user generated content

- …including messaging
  - “Guest book” / “Wall” (asynchronous broadcast)
  - Email (asynchronous unicast)
  - Chat (~ synchronous unicast)

- Collaborative applications / games

! Different target audience / application domain
  - Private and personal OSN
  - Public and professional OSN (business-oriented)



ABOUT THIS MAP

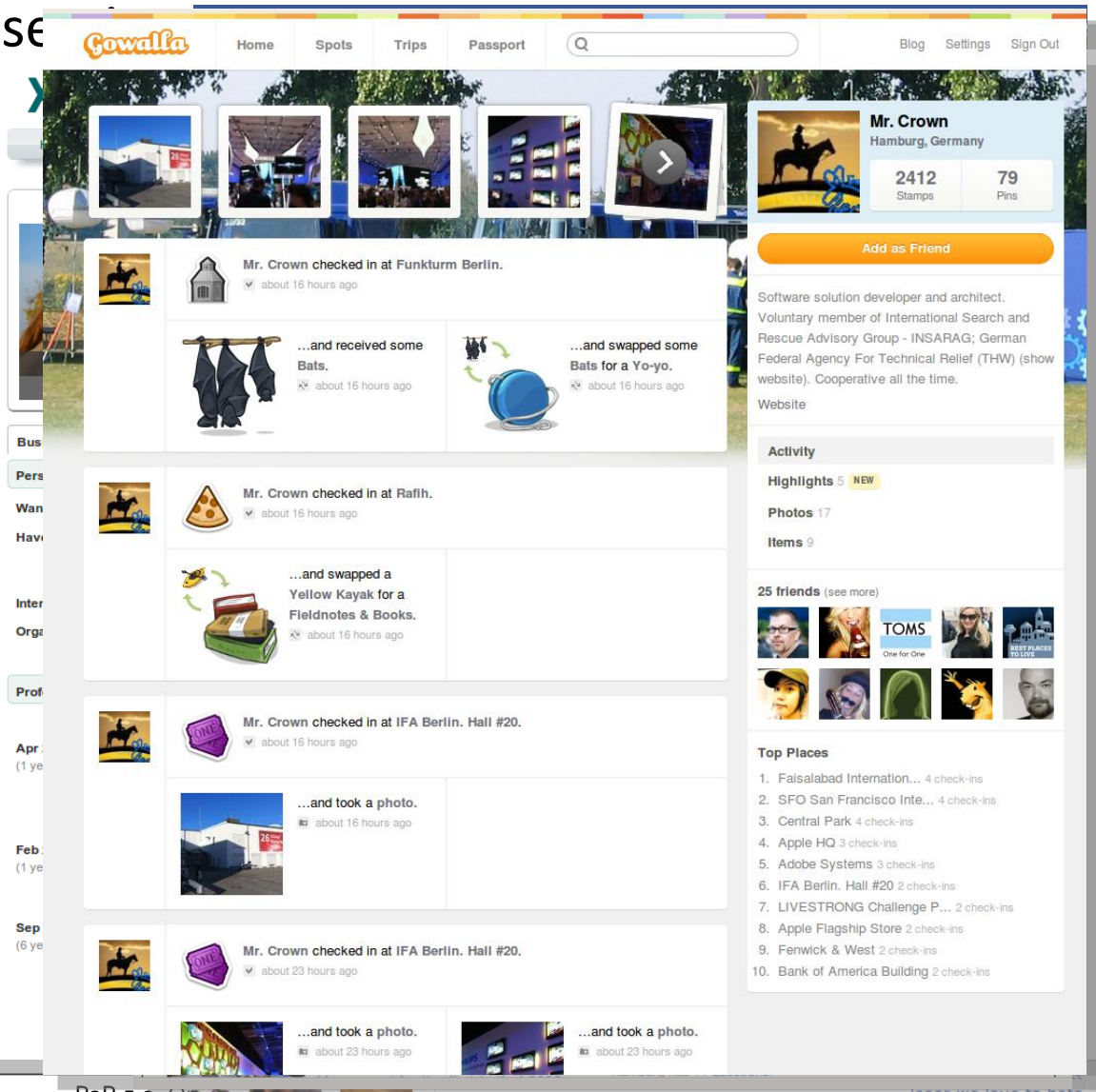
COMMUNITIES RISE AND FALL, AND TOTAL MEMBERSHIP NUMBERS ARE NO LONGER A GOOD MEASURE OF A COMMUNITY'S CURRENT SIZE AND HEALTH. THIS UPDATED MAP USES SIZE TO REPRESENT TOTAL SOCIAL ACTIVITY IN A COMMUNITY — THAT IS, HOW MUCH TALKING, PLAYING, SHARING, OR OTHER SOCIALIZING HAPPENS THERE. THIS MEANT SOME COMPARING OF APPLES AND ORANGES, BUT I DID MY BEST AND TRIED TO BE CONSISTENT.

ESTIMATES ARE BASED ON THE BEST NUMBERS I COULD FIND, BUT INVOLVED A GREAT DEAL OF GUESSWORK, STATISTICAL INFERENCE, RANDOM SAMPLING, NONRANDOM SAMPLING, A 20,000-CELL SPREADSHEET, EMAILING, CATOLING, TEA-LEAF READING, GOAT SACRIFICES, AND GUT INSTINCT (I.E. MAKING THINGS UP).

SOURCES OF DATA INCLUDE GOOGLE AND BING, WIKIPEDIA, ALEXA, BIG-BOARDS.COM, STUMBLEUPON, WORDPRESS, AKISMET, EVERY WEBSITE STATISTICS PAGE I COULD FIND, PRESS RELEASES, NEWS ARTICLES, AND INDIVIDUAL SITE EMPLOYEES. THANKS IN PARTICULAR TO FOLKS AT LAST.FM, LIVEJOURNAL, REDDIT, AND THE NEW YORK TIMES, AS WELL AS SYSADMINS AT A NUMBER OF SITES WHO SHARED STATISTICS ON CONDITION OF ANONYMITY.

Target Audience and Domain

- Professional business se
- Private and personal
- Niche/tailored servic
  - “Micro blogging”: Tw
  - Business trips and m
  - Location-based achie



Why Bother?

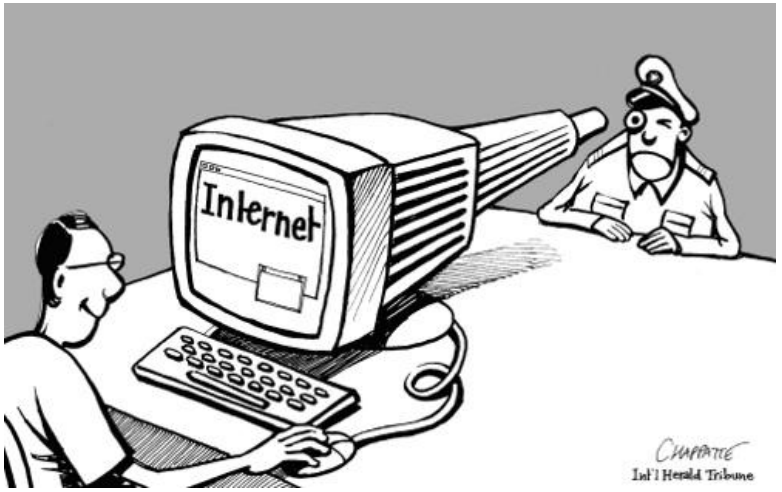
Top 10 Web Brands for January 2010 (U.S., Home and Work)					
RANK	Brand	Unique Audience (000)	Time Per Person (hh:mm:ss)	MOM UA % Change	MOM Time % Change
1	Google	152,708	1:23:54	4.10%	-16.90%
2	Yahoo!	134,561	[illegible]	4.30%	-26.80%
3	Facebook	116,329	7:01:41	5.80%	9.70%
4	MSN/WindowsLive/Bing	109,425	[illegible]	1.20%	-18.10%
5	YouTube	99,525	1:02:27	7.60%	-10.30%
6	AOL Media Network	82,306	1:01:14	-6.80%	-57.80%
7	Wikipedia	64,917	0:15:59	10.70%	-2.70%
8	Fox Interactive Media	62,112	1:23:28	1.00%	-9.10%
9	Amazon	60,772	0:22:34	-8.60%	-32.90%
10	Ask Search Network	57,776	0:12:35	10.70%	-11.40%

Source: The Nielsen Company

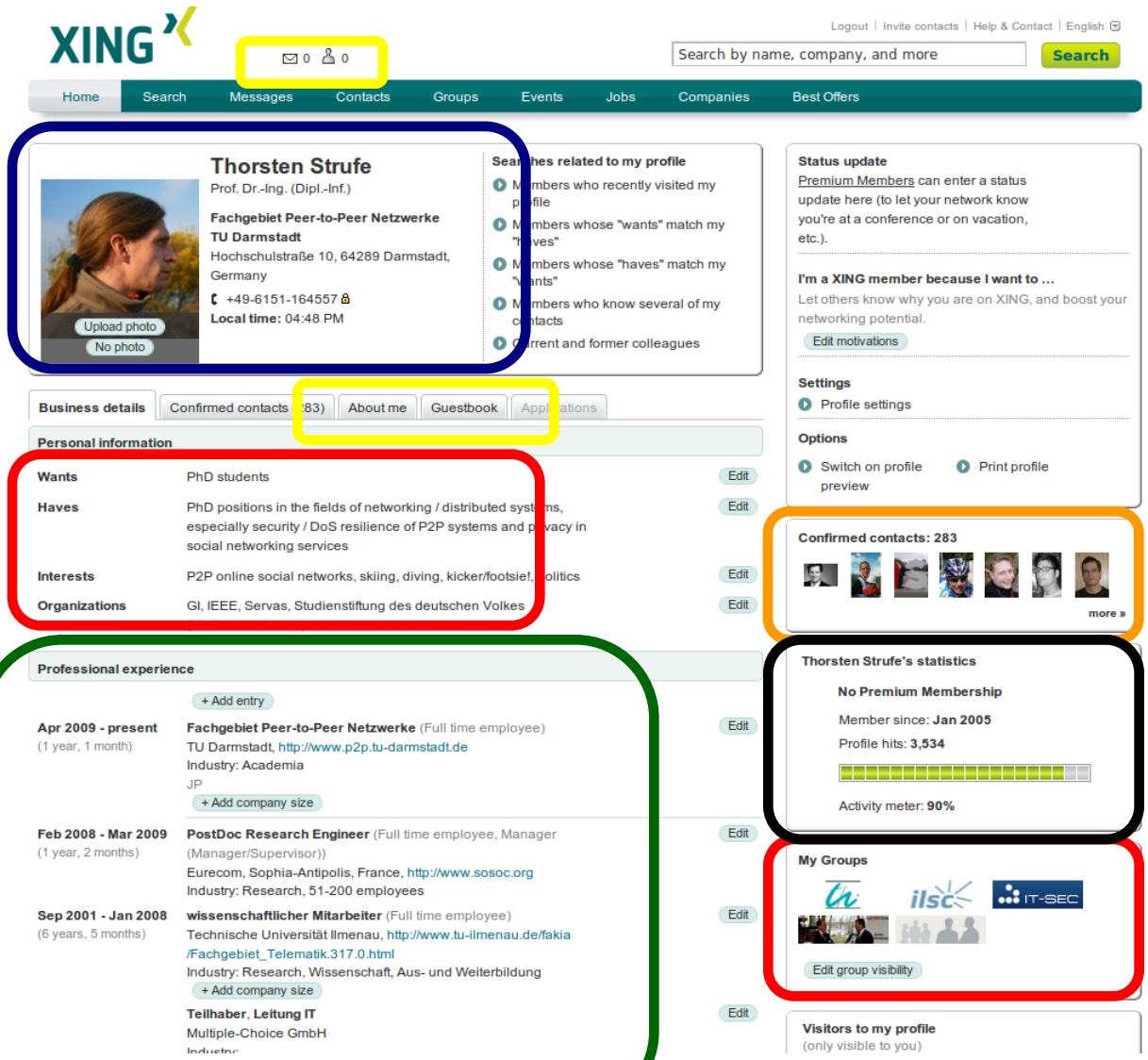
Google-owned sites like Gmail.com.

Characterizing User Behavior in OSN

- Understanding the behavior of users in OSN
- Why?
  - It's interesting! ☺
  - Plus: we need to know to build better (P2P) OSN…
- Questions of interest
  - Sessions (when, how long, - active, - often?)
  - Preferences / services used
  - Popularity of content / pages
  - Scope of access / reciprocity?
- Here: focus on profile popularity…



What's in a "Professional" Profile



- *Identifying info*
  - *Name*
  - *Photo*
  - *Address…*
- *CV*
  - *Current/prev. employments*
  - *Educational track*
- *Interests*
  - *Personal/professional*
  - *Wants/haves*
  - *Interest Groups*
- *Personal contacts*
- *Messaging*
- *Statistics*

Popularity of Profiles – Intuitive Beliefs

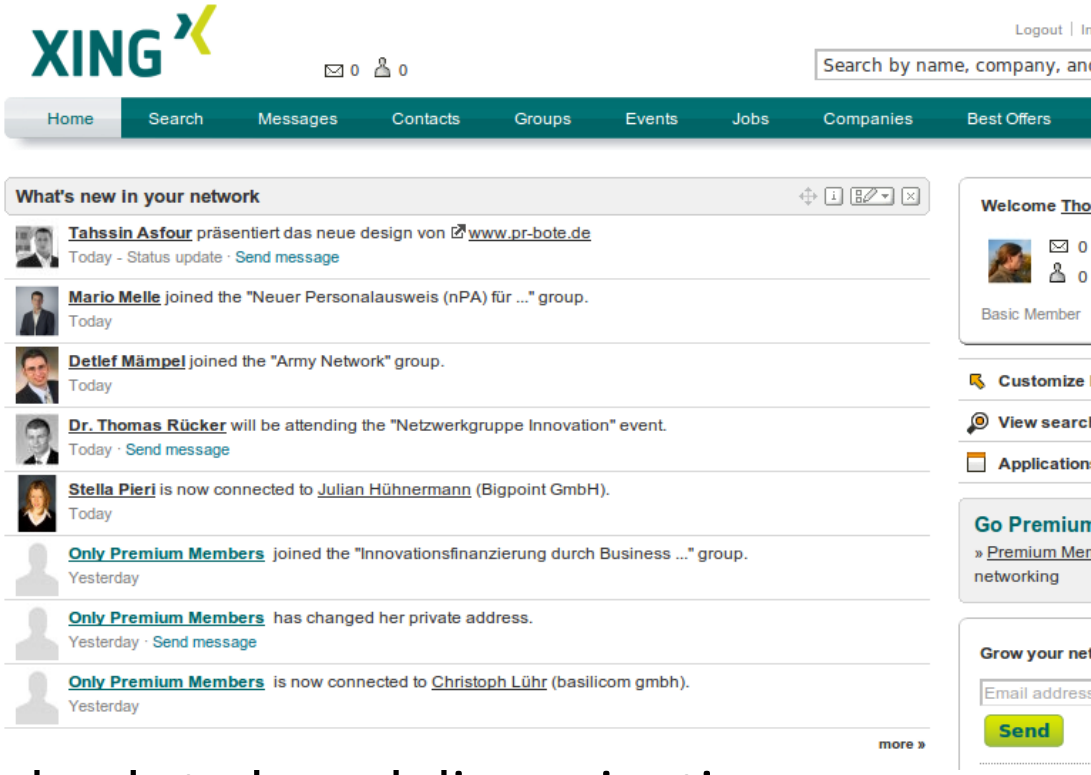
- Which profiles are “popular”?
  - Measured in frequency of requests
  - Possible to correlate to properties of user/profile?
  - Which profiles do we have to keep available (and by which means?) ;-)

- Why?
- Common beliefs…
  - *“Profile of women are much more often visited than profile of men”*
  - *“Profiles with pictures are more interesting than profiles without”*
  - *“Old/experienced profiles attract more views”*
  - *“The profiles of active users are more attractive”*
  - *“Users with many friends are sought and viewed more often”*
  - *“Last name starting with a letter late in the alphabet sucks…”* (c/list pages…)

- Reflecting: how are users lead?
  - Assuming the users generally follow links (rather than searching for content)
  - What do they see as “home” – the front page of the OSN?

The Front Page (and how to get on it)

- "Home page" (after login) usually almost identical
  - Info on profile owner
  - Updates from the provider (and advertisement)
  - Feed of news from "friends"/contacts
- Activity in the news feed:
  - Changes to profile
  - Status updates
  - Birthdays
  - Contact list maintenance (adding friends)



- Note: having many friends leads to broad dissemination…

Data Collection

- How can we gather the data?
  - Access to server logs (***Ha!***)
  - Surveys & Interviews (problems of scale)
  - Traffic logging (problems of scope)
  - ***Crawling***/API access (problems of scale, incompleteness of information, sampling)

- Crawls gather only limited data
  - Does ***not*** sufficiently allow inference on ***sessions***
  - Mainly comprises of plain, static ***profile info*** and ***social graph***
  - Generally does ***not*** include data about ***popularity***

- Regular monitoring:
  - Collect changes to profiles
  - Frequent, regular measurements over long period of time needed

Studying a Prominent OSN

- "xing" selected for the study
  - Business/professional OSN, similar to LinkedIn
  - 8 Mio users, mainly from central Europe
  - xing profiles include
    - Registration date
    - ***activity meter***
    - ***hit counter*** (number of profile impressions for popularity)
    - Weak privacy settings (professional profiles are there to be seen)
  - Visitors to profile visible (to paying users – no stalking, unlike LinkedIN, facebook)

- Crawling / monitoring the ***complete*** OSN is ***infeasible***
  - 8mio profiles, most >15 pages of contact lists (up to 160k contacts!, 10 per page)
  - Access per page takes ~ .5s, complete crawl takes > 275 h (if all goes well)
  - each page > 150 KB, > 17 TB in total
  - Providers don't like this much… (rate control, disabled accounts, blocked IPs)

- Large, ***random*** sample needed for meaningful results

Sampling and Monitoring

- Random sampling
  - Conducted random walks (25k, 5k, ~1k)
  - Wish for the crawls:
    - Diverse graphs without overlap
    - Collect “john does” (no outliers, no abandoned profiles)
  - Covered over 2Mio unique profiles in total
  - Starting at diverse “edges” (AUS,DE,PL,RUS,TR,UK,US)
  - all converged to D.A.CH

- Selected sub graph without overlap
  - 31.643 unique profiles (25k, 5k, 1.6k random walks)
  - Gender automatically derived via website on international first names

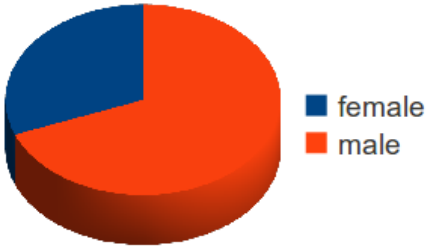
- Subsequently frequently monitored for a long period of time
  - Since Nov 2009
  - At least twice daily
  - Only core data needed (no pictures, friend list not regularly since # on profile)

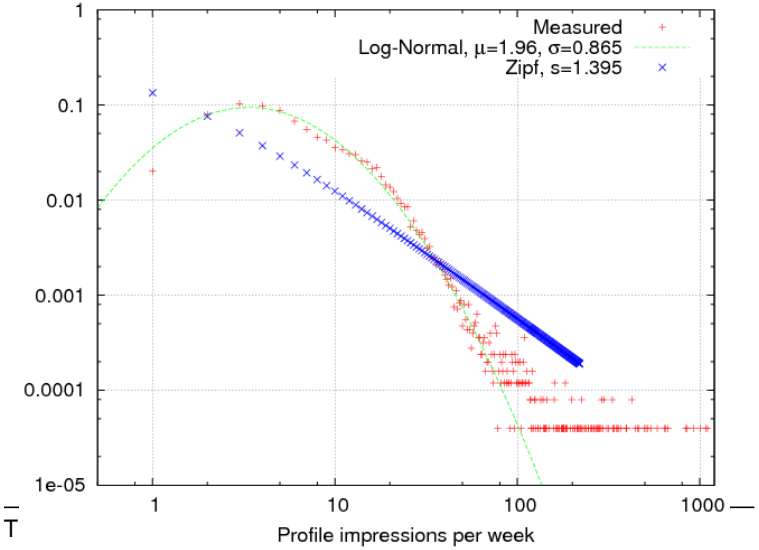
Monitored Data

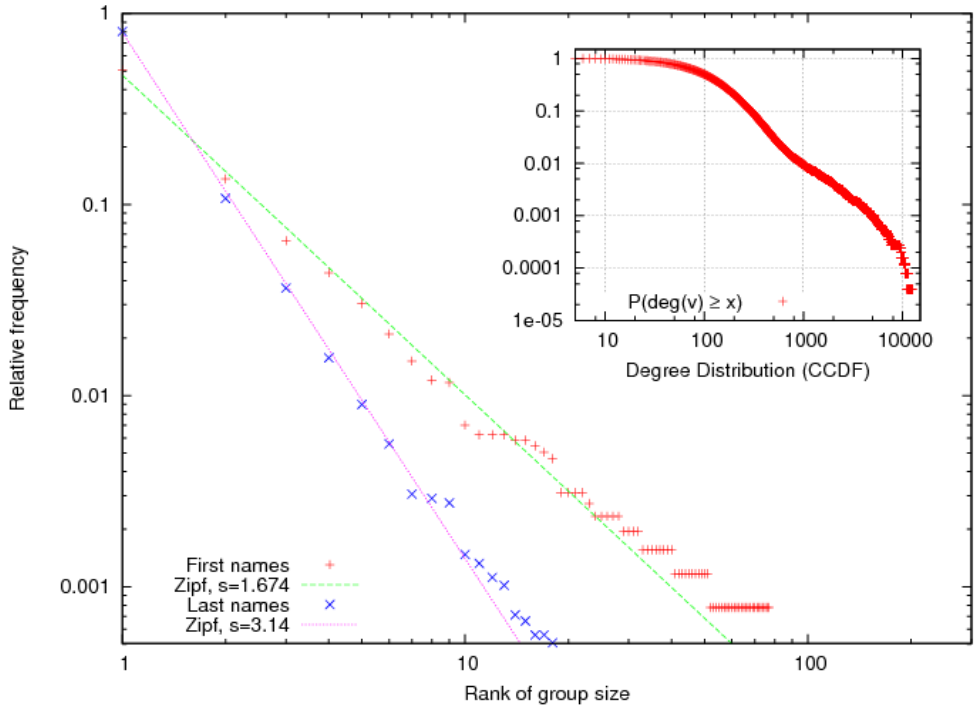
- Identifying Data
  - Name, Image, Gender
- CV
  - Current employment, universities attended, claimed spoken languages
- Interests
  - Interests as stated
  - Number of subscribed groups, subscribed groups, number of members in groups, number of messages in respective groups, languages of group
- Contact list information
  - Number of contacts
  - Complete list of contacts gathered infrequently
- Statistics
  - Registration date, number of profile impressions, activity meter
  - Timestamp of crawl

Analyzing & Validating the Sample

- 3 months section of monitoring data analyzed (Nov ’09 – Jan ’10)
- Some profiles removed (celebrities, abandoned profiles)
- Remaining sample
  - 25.274 (7.824 / 17.450) (31% vs. 34%)
  - Degree dist. (~PL, min 5, max 12.332)
  - Name frequency follows Zipf
    - First names: s= 1.67
    - Last names: s= 3.14
  - Binned popularity dist. log-normal







Activity, Group Membership

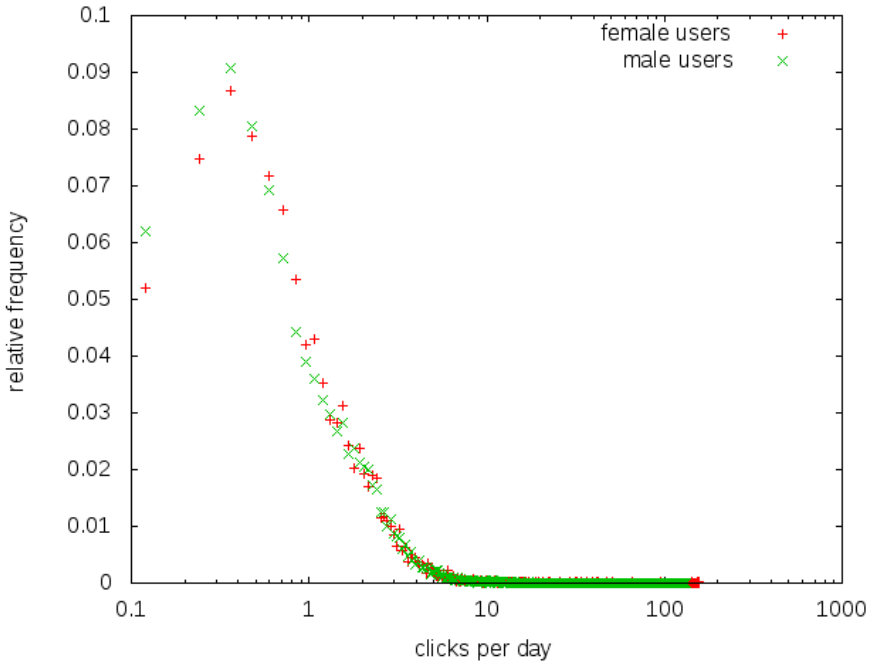
- Activity
  - “Activity meter” in profiles very coarse grained
  - Derived “profile alteration frequency” as alternative
  - Men are slightly more active than women (to both metrics)
  - Profiles without image belong to inactive users

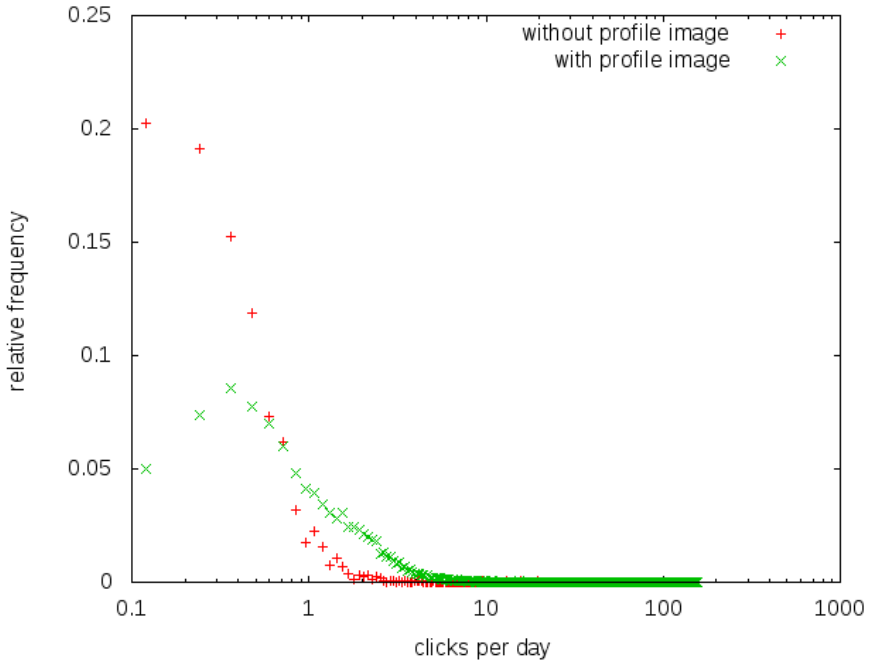
- Membership in Groups
  - Wide range of group membership
  - Max 511, Mdn 3
  - >5k profiles are not registered to any group

Testing Groups of users (non-parametric)

- Testing popularity between groups (non-parametric)

- *"Profile of women are much more often visited than profile of men"*
  - Male vs. female: 0.039 vs. 0.041 (*Mdn*)
  - No significant difference
- *"Profiles with pictures are more interesting than profiles without"*
  - With picture significantly higher popularity (0.5 pi / d)





Correlational Tests (Pearson's *r*)

- *"The profiles of active users are more attractive"*
  - Activity as given on the profiles ("***activity meter***")
  - $r \approx 0.17$, no noteworthy correlation
  - Activity measured in ***group memberships***
  - $r \approx 0.37$ (higher for men, lower for women)
  - *A*ctivity meter is very coarse grained
  - Activity measured in ***profile alterations***
  - $r \approx 0.62$ ($0.61 < r < 0.63$) ***high correlation***

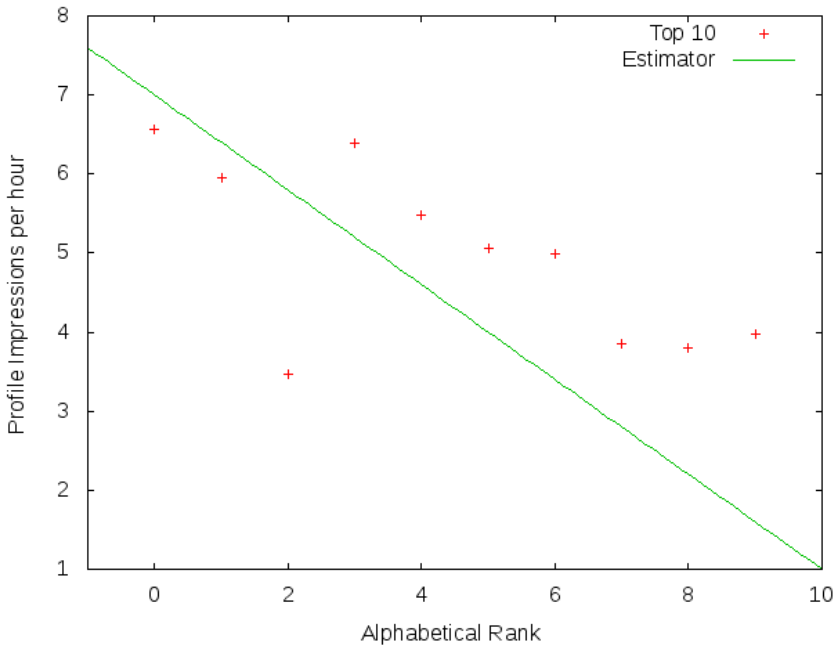
  → Popularity correlates with activity of users (profile alterations/group activity)

- *"Users with many friends are sought and viewed more often"*
  - Correlating popularity to the degree of profiles
  - $r \approx 0.75$, ***high correlation***
  - Stronger for women: $0.81 < r < 0.83$ vs. men: $0.74 < r \leq 0.75$

- ***Combination s/what unsurprising: changes are published at friend's profiles…***

Correlational Tests (odds n' sods ;-)

- *"Old/experienced profiles are viewed more frequently"*
  - Preferential attachment / experience could lead to higher popularity
  - Ho rejected, but ***r*** ≈ 0.11, ***no noteworthy correlation***
- *"Last name starting with a letter late in the alphabet sucks…"*
  - Ho not rejected, there is no correlation.
  - *Taking the "rich-club", however…*
  - Top 5% profiles: *r ≈ - 0.09*
  - Top 2‰ profiles: *r ≈ - 0.22*
  - Top 1 ‰ profiles: *r ≈ - 0.29*
  - Top 10 profiles: *r ≈* ***- 0.9 !!***



Summarizing the User Model

- Selected large sample of profiles in *business oriented* OSN
- Monitored profile properties and popularity (in pi/h)

- Profile Popularity *can* be predicted. Relates to
  - Providing image
  - Activity (diligence of maintaining profile)
  - Number of friends and contacts

- What we take away (P2P OSN)
  - Nice correlation with activity/friends (P2P & replicating at friends…)

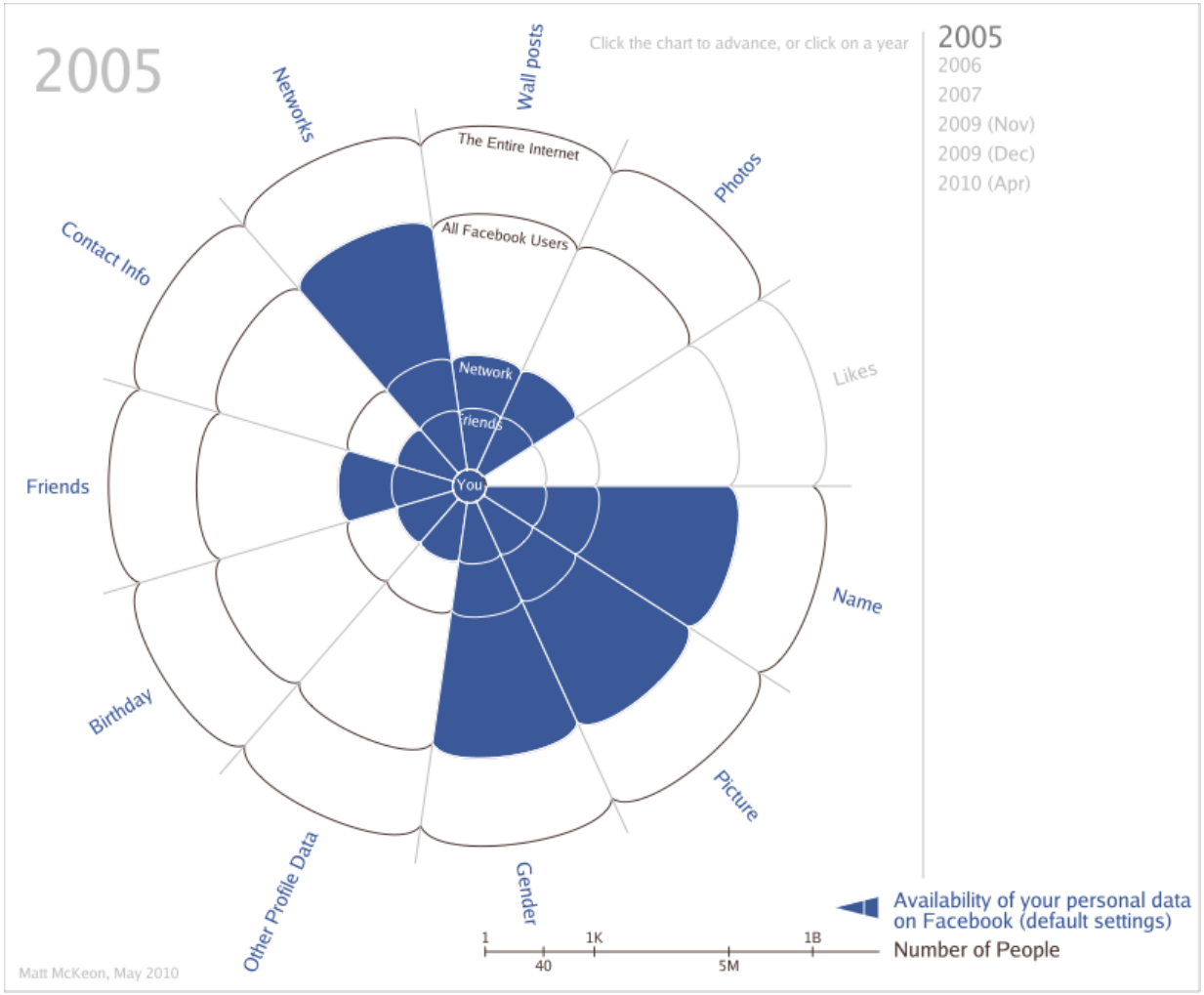
- Future Work
  - Kept monitoring, but results quite stable…
  - Analyze data from DB and server access logs “spi” (fb-like personal osn)
    - Are profile requests “local” (viewing friends…)?
    - Is interest mutual?
    - Can we learn more on the sessions?
  - Struggle to get/analyze more data!

Facebook and Privacy Settings

- Facebook largest and most successful OSN:
  - Over 800 Mio active users
  - 50 Mio user after 3 years (2004-2007)
  - 50 mio users per quarter since 2007
  - ½ Mio registrations ***daily***
  - De-facto reference

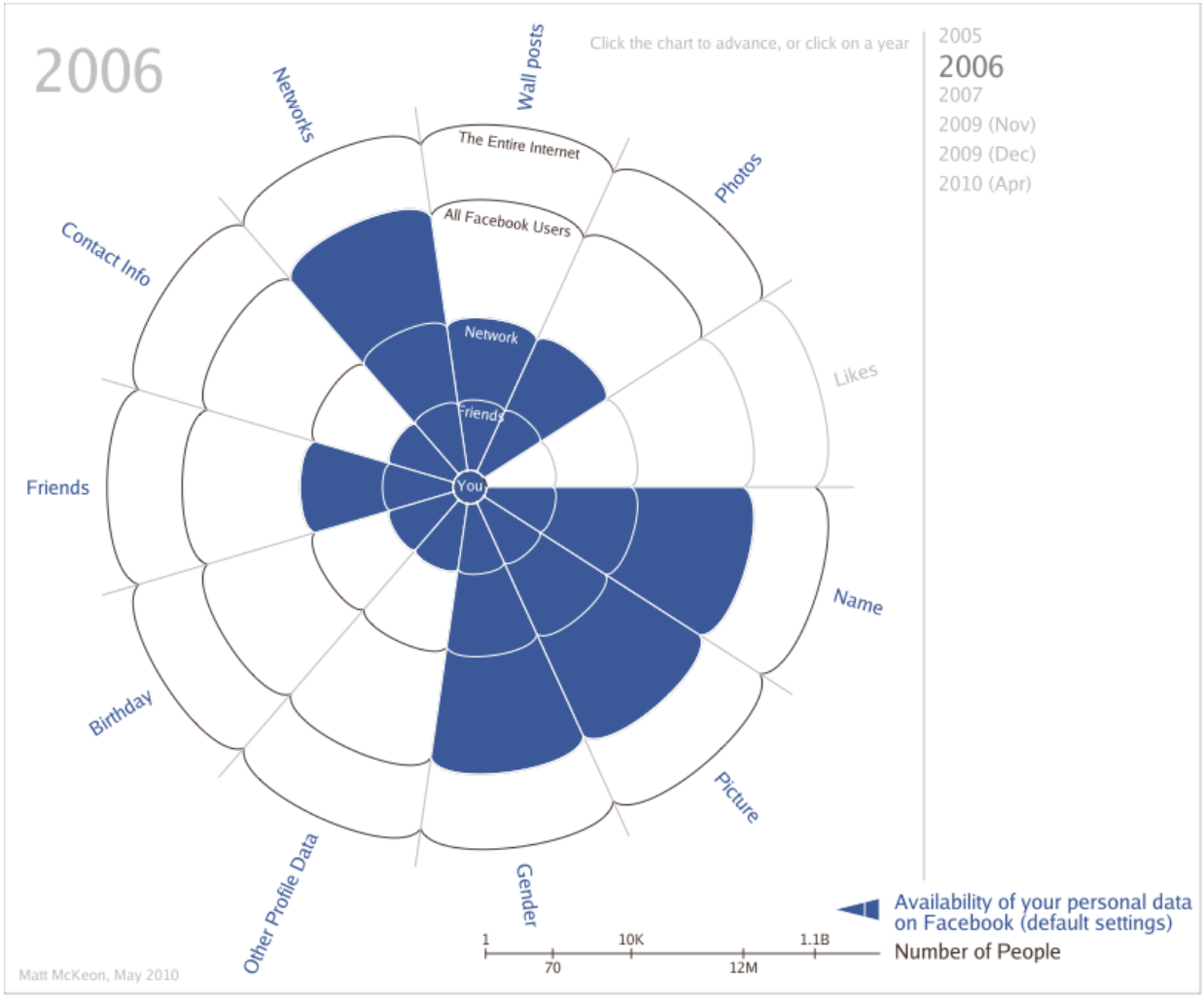
- Potential reasons for the dominance
  - Quick following of new hypes
    - "Apps" (Extensions to service, integration of 3rd party services)
    - "Friend feed": construct virtual village (following twitter's success)
    - Facebook places (following foursquare and gowalla's success)
  - Stalking service (you don't know who watched your profile => if this means success, users need to make their private data public!)

Facebooks Privacy-Evolution: 2005

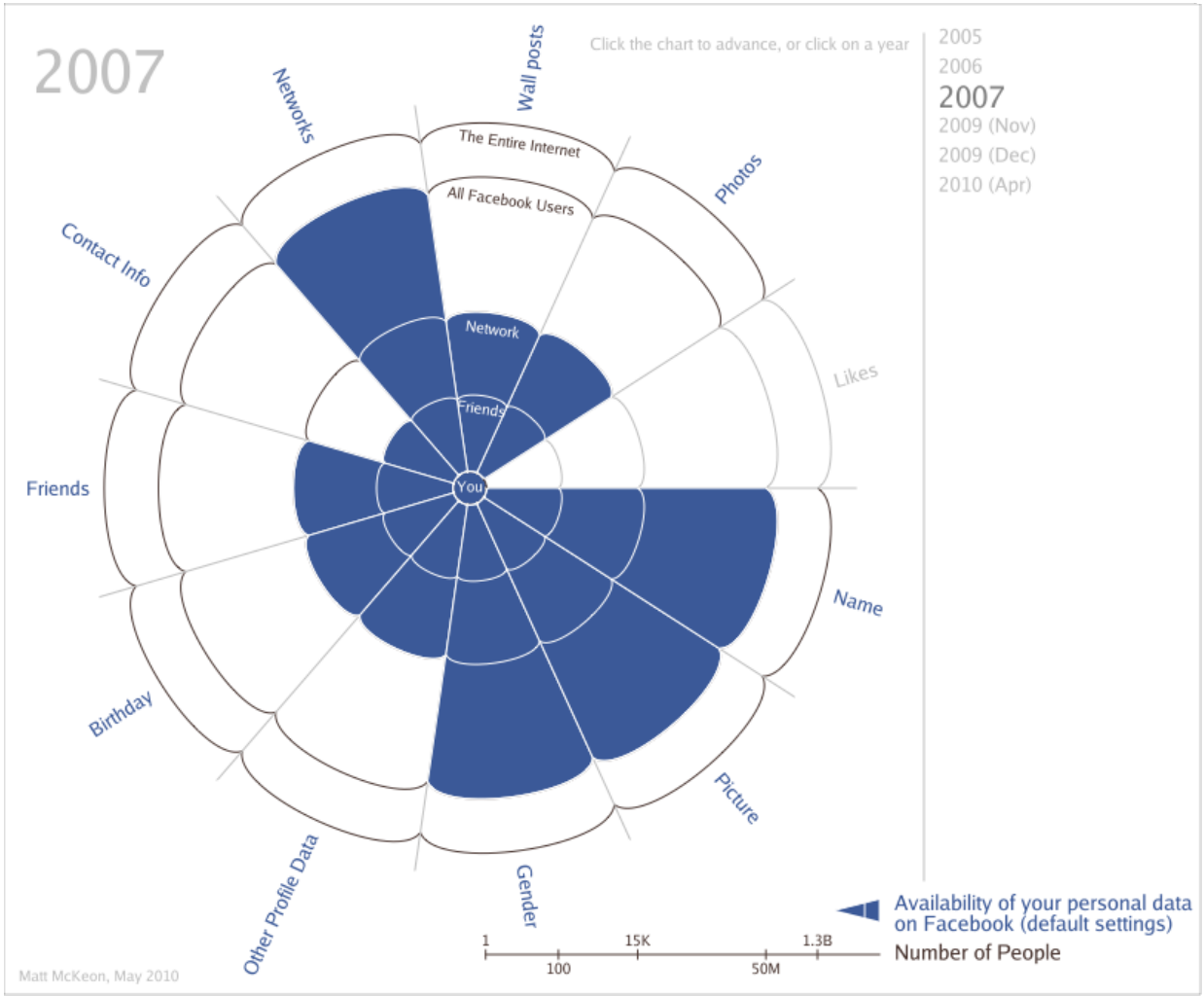


Source: McKeon

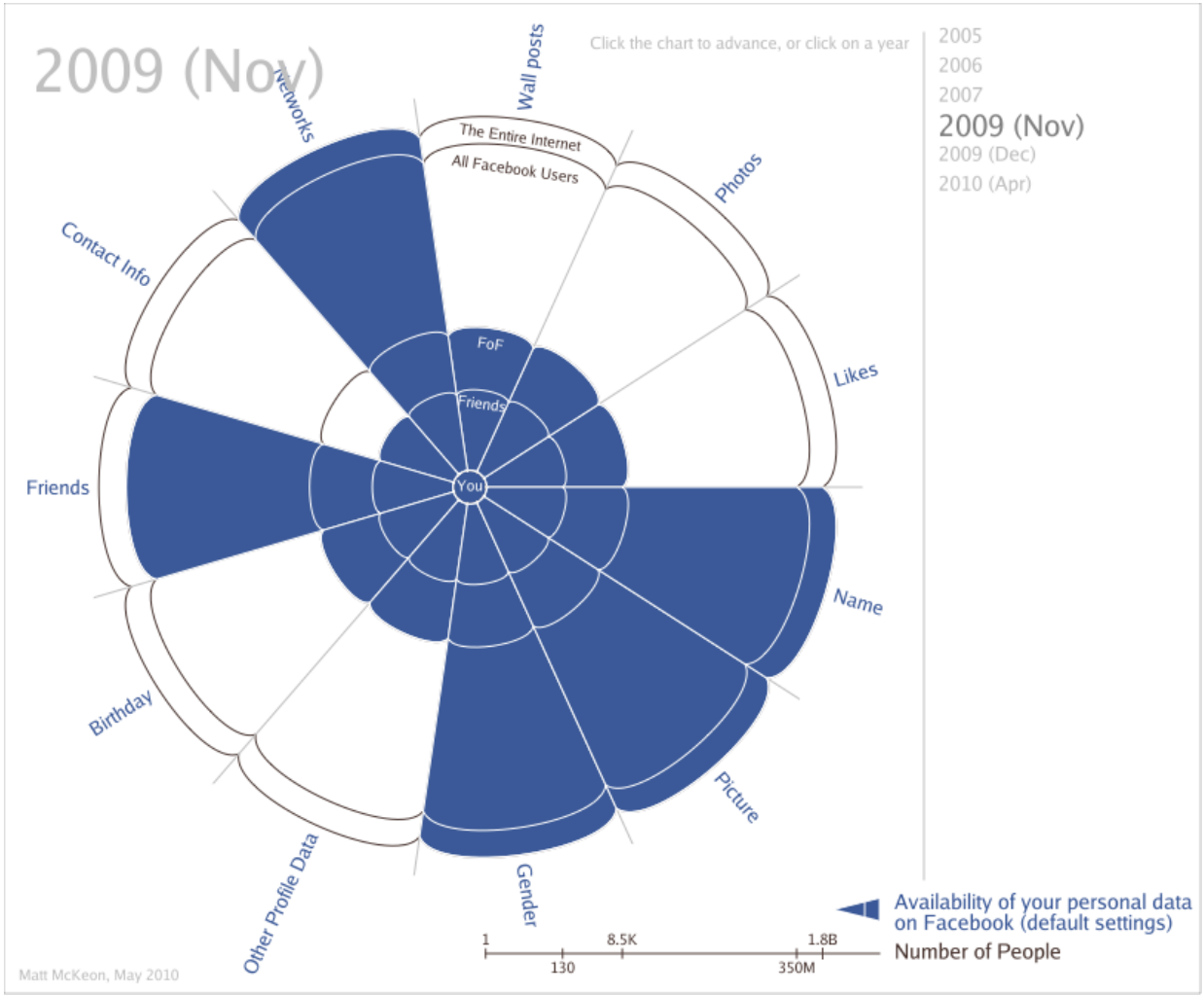
Facebooks Privacy-Evolution: 2006



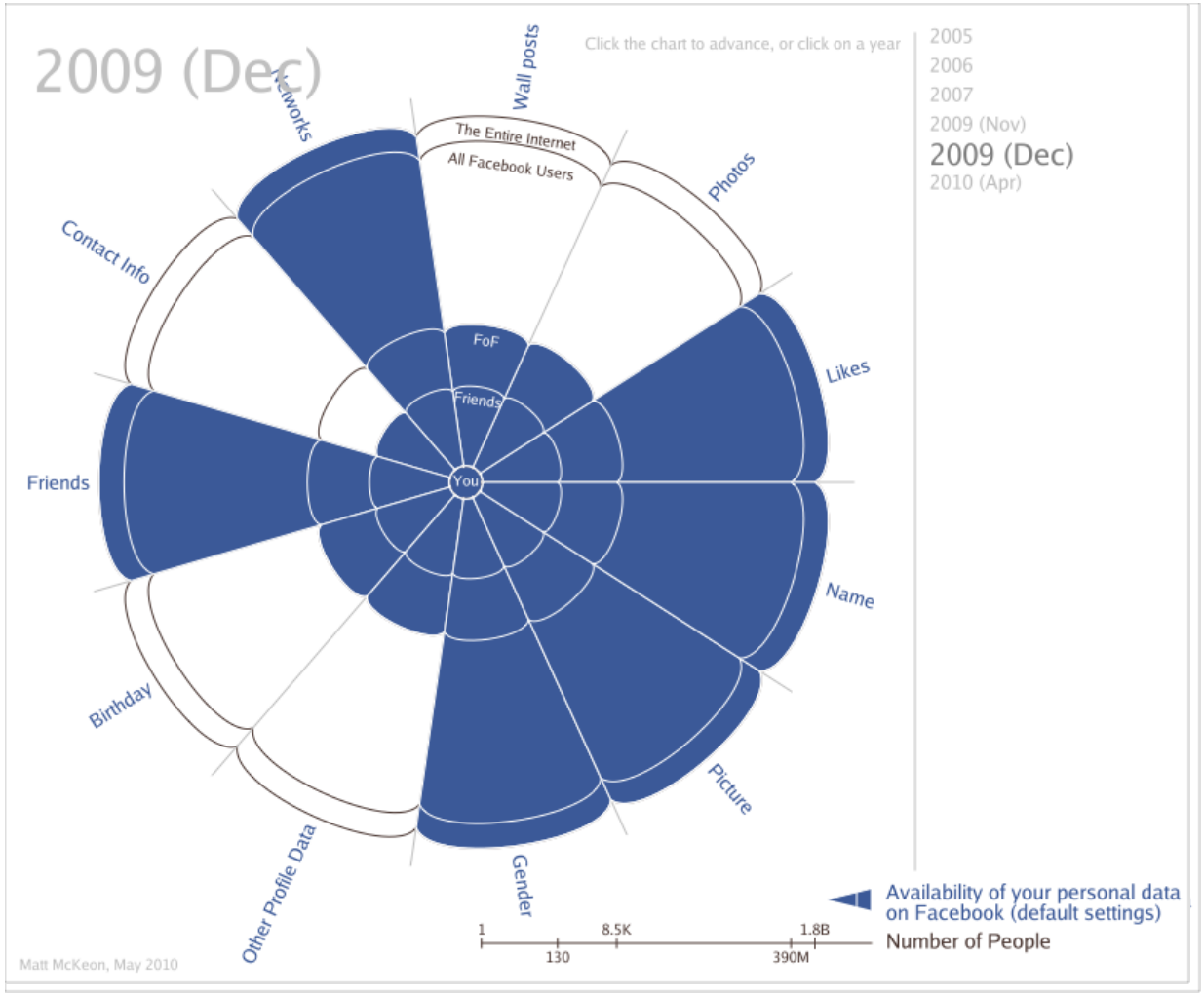
Facebooks Privacy-Evolution: 2007



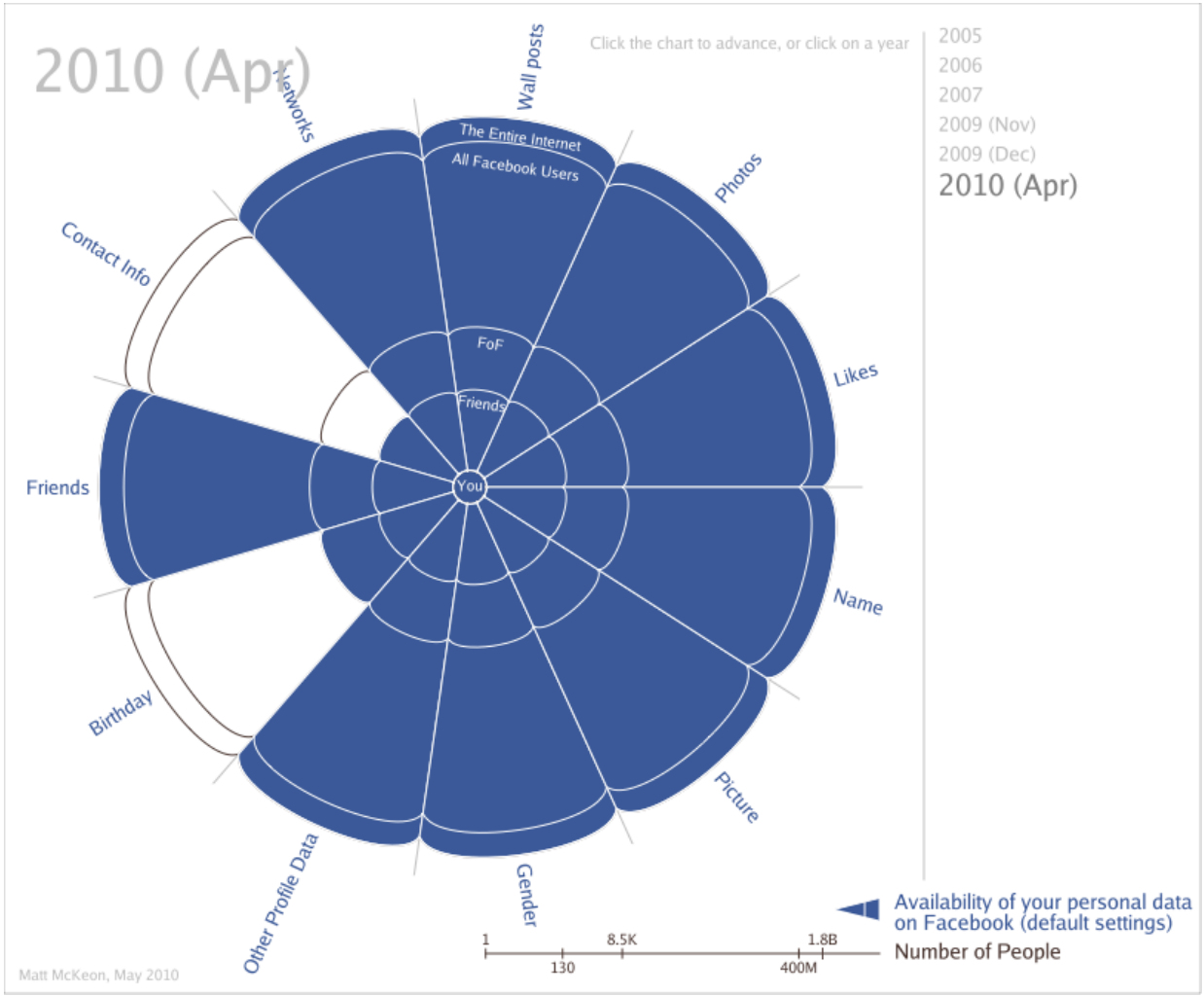
Facebooks Privacy-Evolution: 2009



Facebooks Privacy-Evolution: 2009



Facebooks Privacy-Evolution: 2010



Social Networking Sites as Targets

- Information on social networks is sensitive and attractive by nature:
  - E-mail addresses
  - Postal addresses
  - Educational background
  - Gender/Relationship status/partner
  - Spoken languages

- This information is valuable for attackers
  - Spear phishers
  - Spammers (implicit trust between users)
  - Malware authors (effective infection medium)

Cloning Attacks on Social Networks

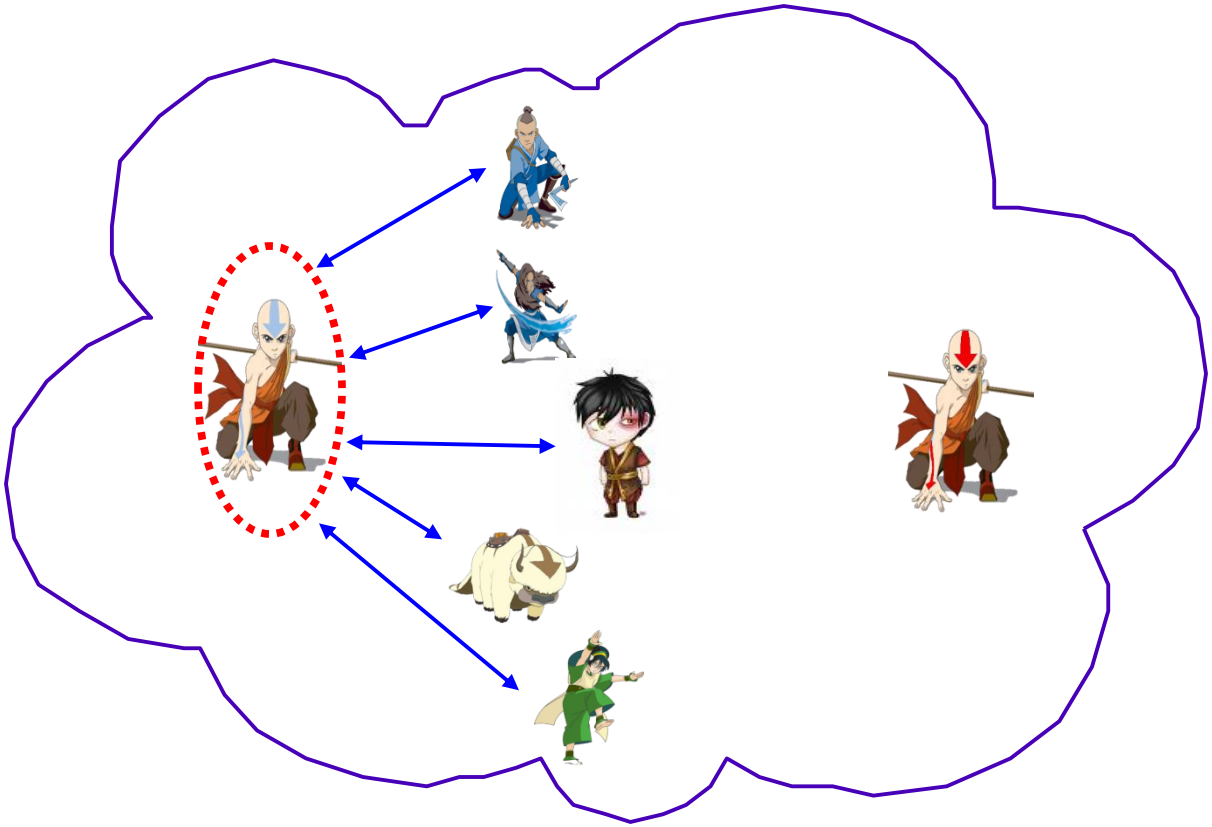
- Target: get on the friend list of real users to get access to their personal information and their circle of trust

- Two Cloning Attacks
  - Clone the account of an existing user inside the same network and send friend requests to her contacts
  - Clone the victim profile into a different social network where she is not registered and contact her friends

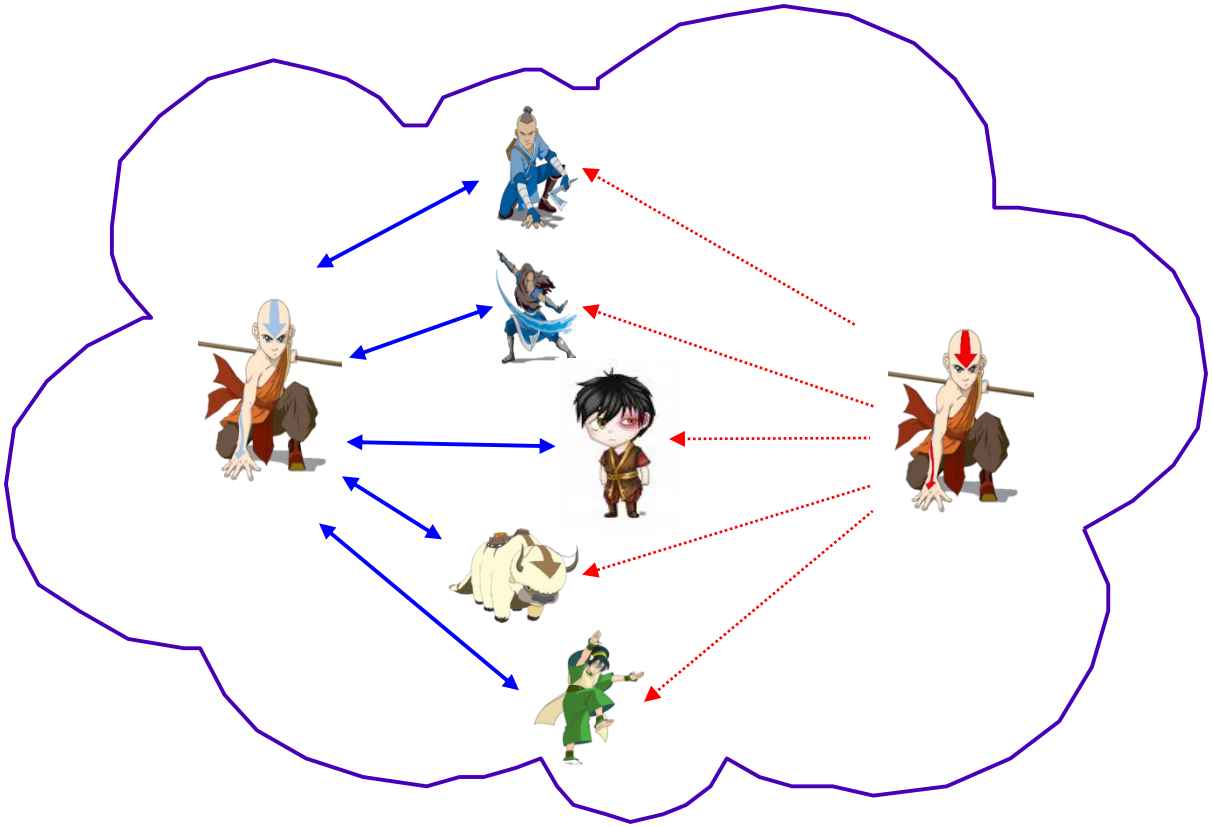
Aim of Attacks and Experiments

- Is it possible for an attacker to launch impersonation attacks on a large scale against a number of popular social networking sites?
  - Facebook (international)
  - XING (international)
  - LinkedIn (international)
  - MeinVZ (popular in Germany, Austria, Switzerland)
  - StudiVZ (popular in Germany, Austria, Switzerland)

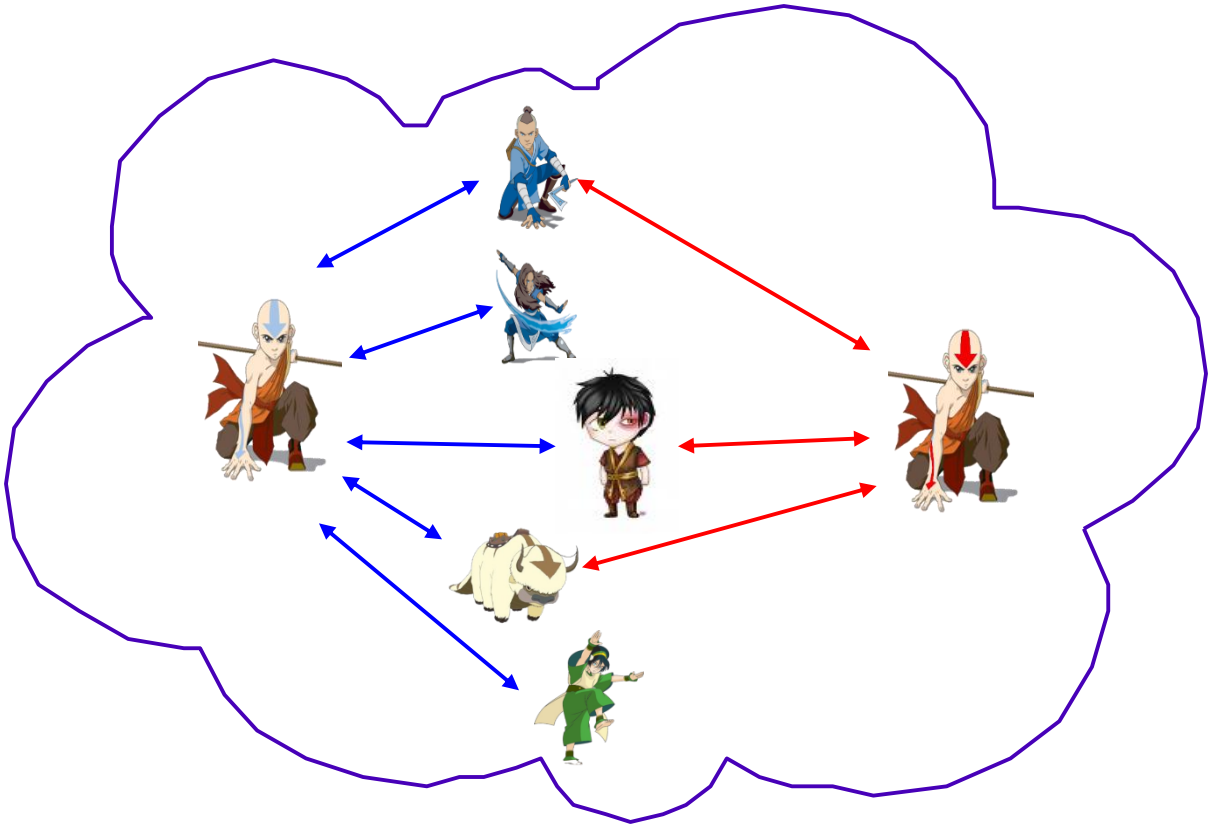
Automated Profile Cloning



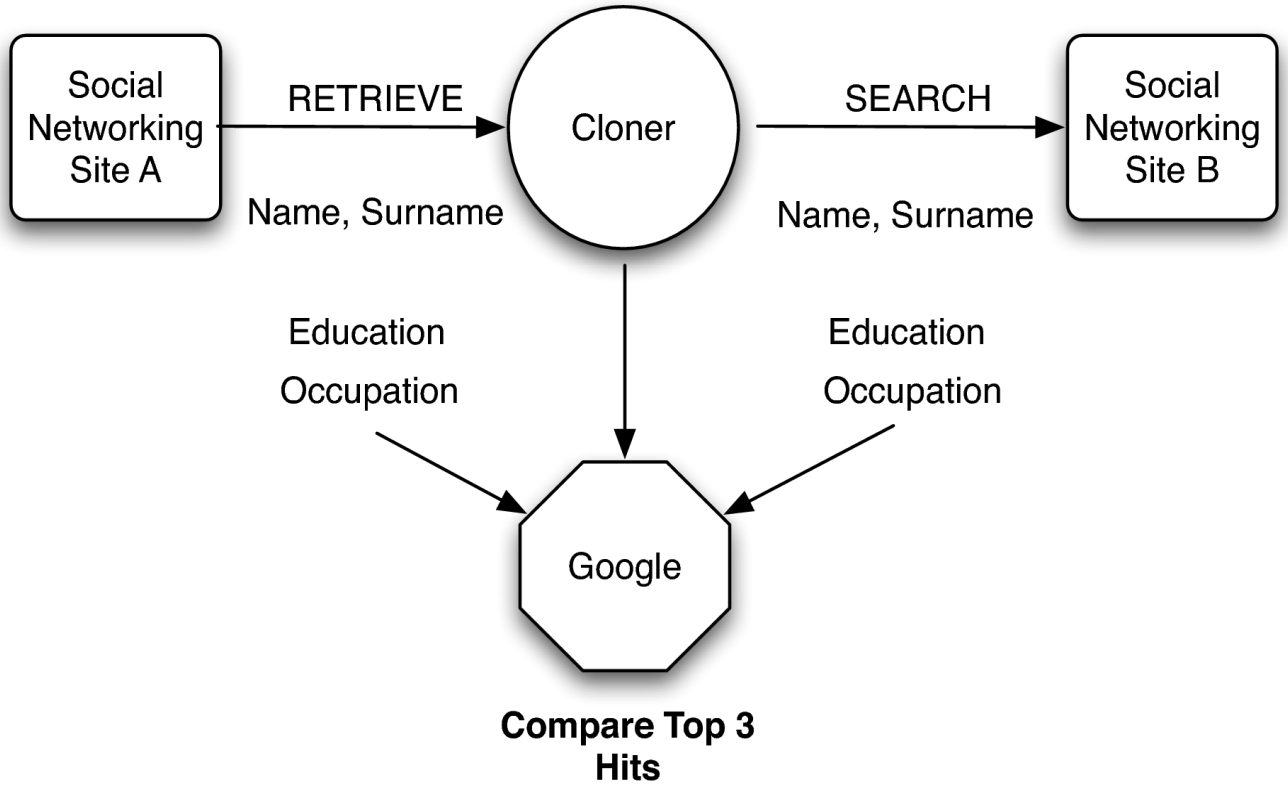
Automated Profile Cloning



Automated Profile Cloning



Cross-Site Profile Cloning



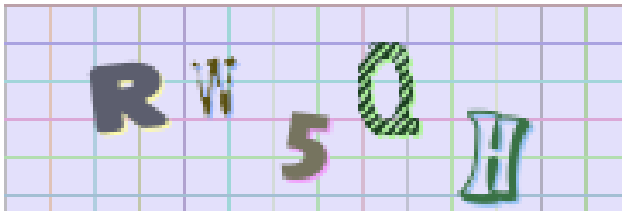
Obstacle: CAPTCHAs

- CAPTCHA: Completely Automated Public Turing test to tell Computers and Humans Apart

- CAPTCHAs are employed to prevent automated programs from accessing and abusing the services

- In order to automate the attacks, a number of CAPTCHA breaking techniques were developed
  - “Quick and dirty”, techniques are not perfect
  - The aim is to break the CAPTCHAs efficiently enough to make automated attacks against several social networking sites possible

MeinVZ and StudiVZ CAPTCHAs



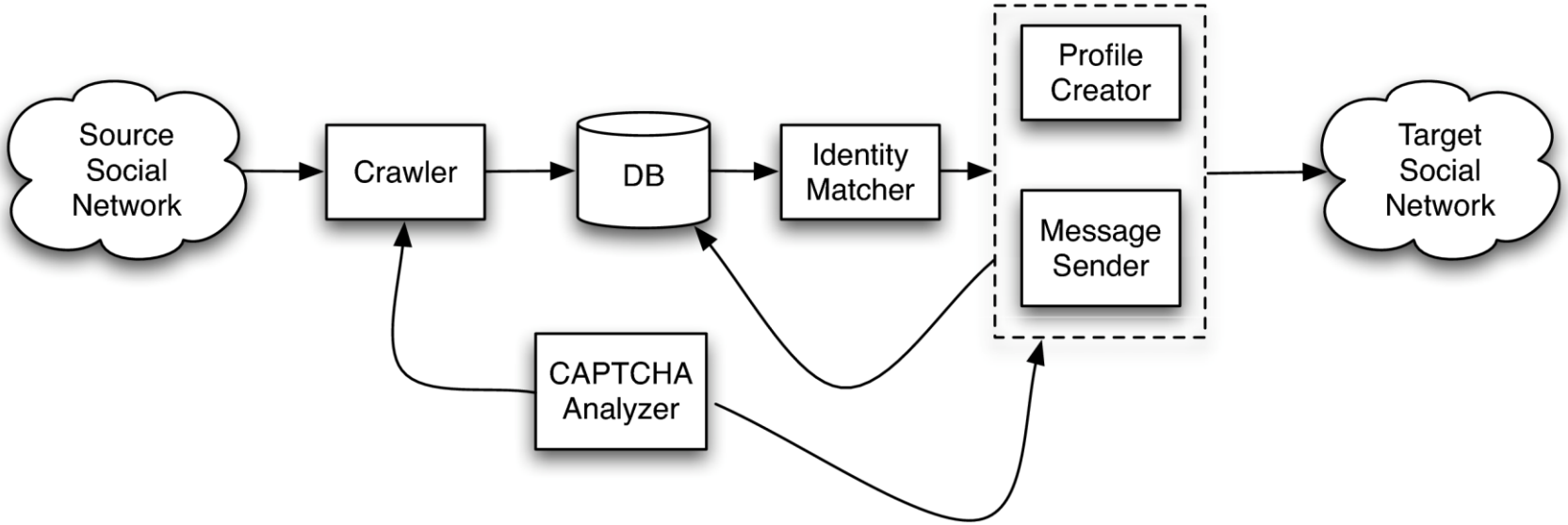
- GD Library (PHP) CAPTCHAs
- CAPTCHAs always contain 5 letters
- Each letter is written in
  - Different font
  - Different background and foreground color
- Often tilted, scaled or blurred
- A simple grid-base noise is added to the image
- Quick script* with success rate of 88.7%

**Cracking the CAPTCHAs was done with serious amounts of help from Michael Roßberg/TU-Ilmenau*

Facebook CAPTCHAs

- Adopts ReCAPTCHA technique
  - Asks words that are encountered while digitizing books that cannot be correctly recognized by the OCR program
  - By solving the CAPTCHAs, the user contributes to the effort to increase the accuracy of the text of the digitized book
- ReCAPTCHA asks meaningful words. Therefore, after solution is found, the word is sought in a dictionary
  - Result additionally submitted to Google as check
- Script with success rate of 7%
- Might seem small, but…
  - If every bot is capable of solving 7 CAPTCHAs per day, a botnet that consists of 10.000 can send 70.000 friend requests per day
  - Attack against Microsoft Live Hotmail had similar success rate

Prototype Implementation: iCloner



Real-World Experiments

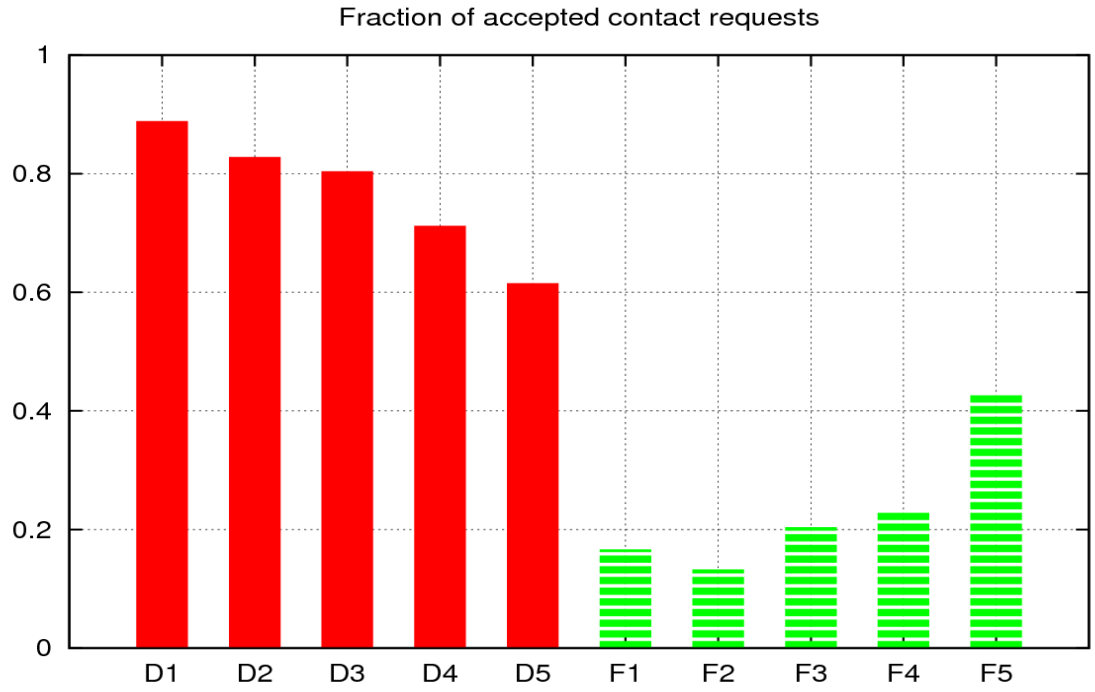
- Is it feasible to perform cloning attacks in the real-world?
- Questions:
  - Can an attacker launch large-scale attacks?
  - How willing are users to accept friendship requests from forged profiles of people who are already in their friendship lists?
  - Is it possible to efficiently find two identical accounts in two different social networks?

Profile Crawling Experiments

- StudiVZ and MeinVZ
  - Displays CAPTCHA if large number of requests come from one account
  - To collect as much information as possible, without being noticed, 16 accounts were created, and separately used for crawling
  - Collected 5M profiles with contact information, and 1.2M complete user profiles
- XING
  - Does not display CAPTCHA, but disables the account if the account requests around 2000 pages consecutively
  - 118,000 accounts were crawled

Profile Cloning Experiments

- Attack: duplicate the profiles of five users (D1,…,D5) and create fictitious profiles (F1,…,F5 as control group)



Profile Cloning Experiments

- Do the users really trust their friends in their friend list?
- Would they click the link seen in the message below?

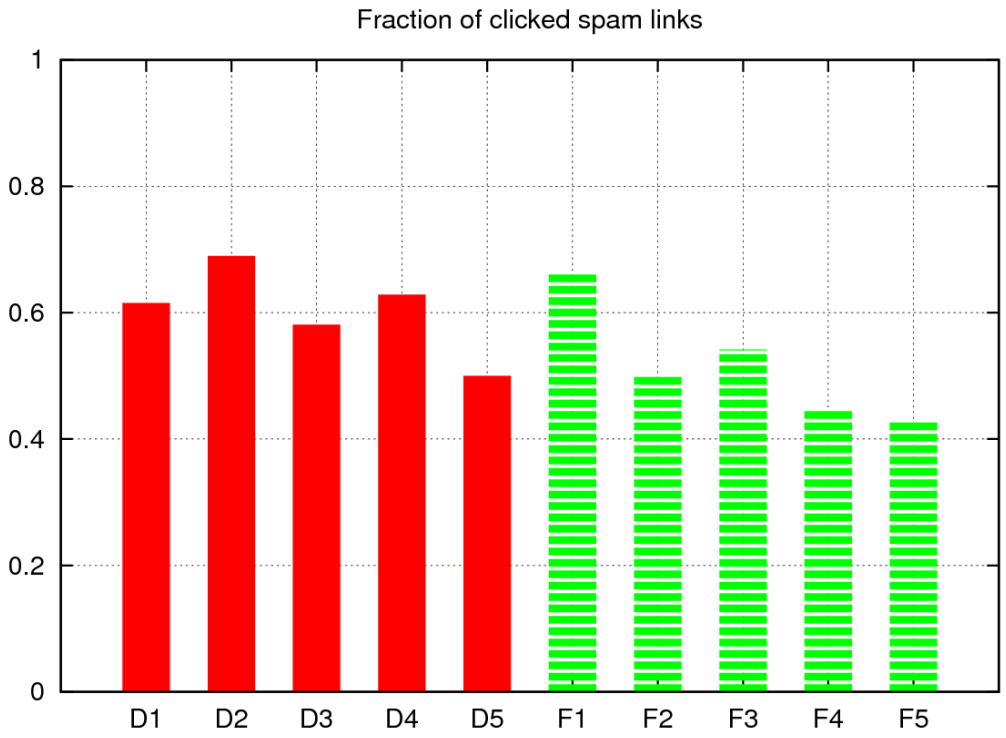
```
Hey, I put some more pictures online. Check them
here!:

http://193.55.112.123/userspace/pix?user=<account>
&guest=<contact>&cred=3252kj5kj25kjk325hk}

Ciao, <account first-name>
```

Profile Cloning Experiments

- Click through rate for messages from duplicate / fictitious profiles

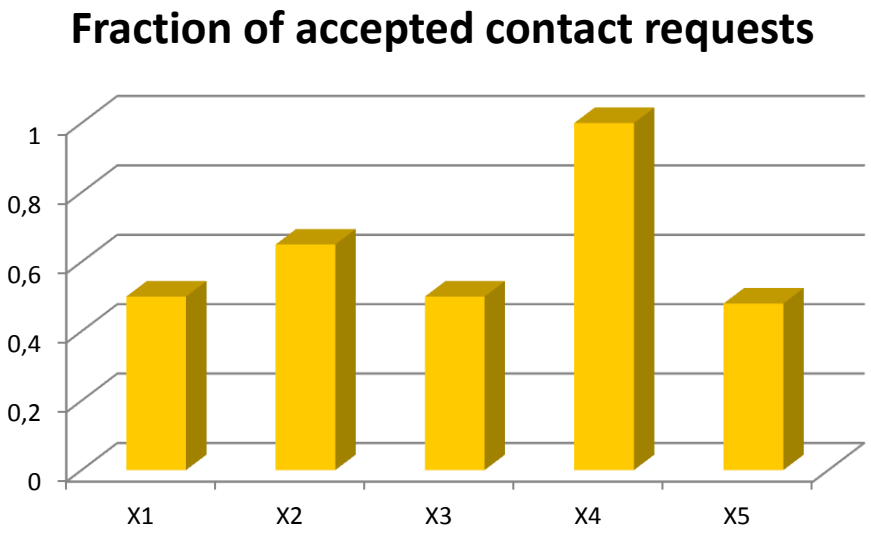


Cross-Site Profile Cloning Experiments

- Cloning profiles that exist on XING, but not on LinkedIn
- The success of the cross-site profile cloning depends on the number of users that have a profile in both of the networks
- From around 30.000 crawled profiles in XING, 3.700 were also registered in LinkedIn

Cross-Site Profile Cloning Experiments

- Clone 5 users from XING to LinkedIn
- iCloner identified 78 out of 443 XING friend contacts that were also registered in LinkedIn
- Fraction that has actually accepted the contact requests:



How Can Social Network Security be Improved?

- Make the users aware of the risk
- Make privacy controls easier to use (and show consequences!)
- Provide a stronger way to authenticate users

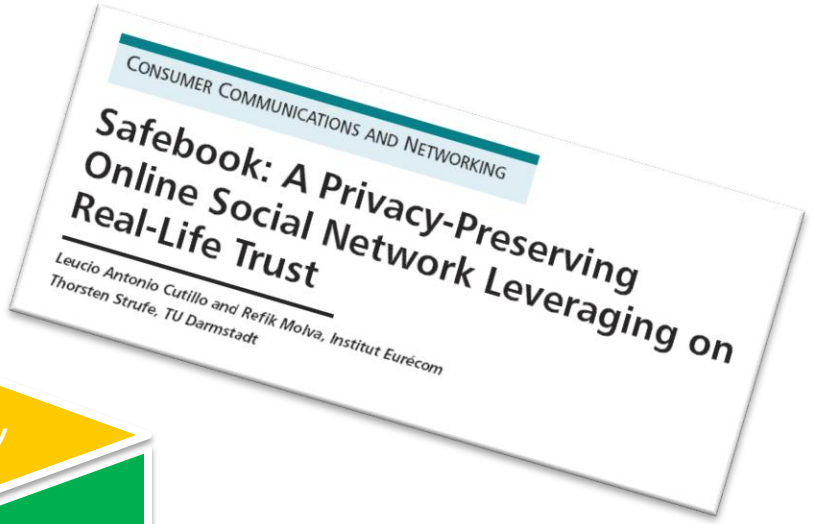
- Increase defense (rate ctrl against massive automatic harvesting):
  - Make CAPTCHAs more difficult to break
  - Reduce response rates
  - Limit number of accessible profiles

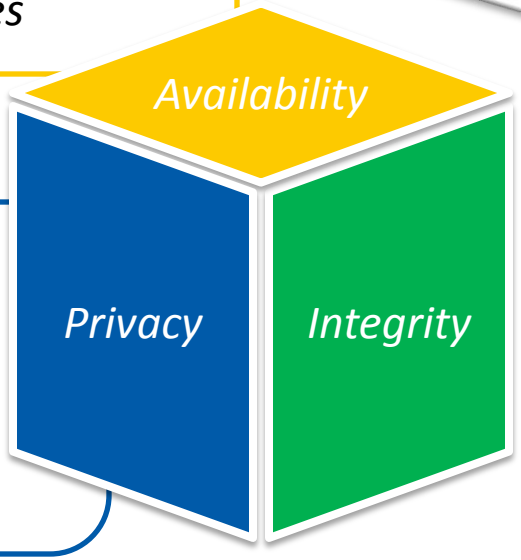
- Apply anomaly detection techniques to detect:
  - Crawling
  - Massive amount of friend requests or messages
  - Profiles with identical/similar information
  - Attempts to automatically solve CAPTCHAs

- …decentralize, but that comes a bit later ;-)

Security Objectives in OSNs

1. *Of data and services*
2. *Robustness against censorship*
3. *Prevent from seizure or hijacking of identities*

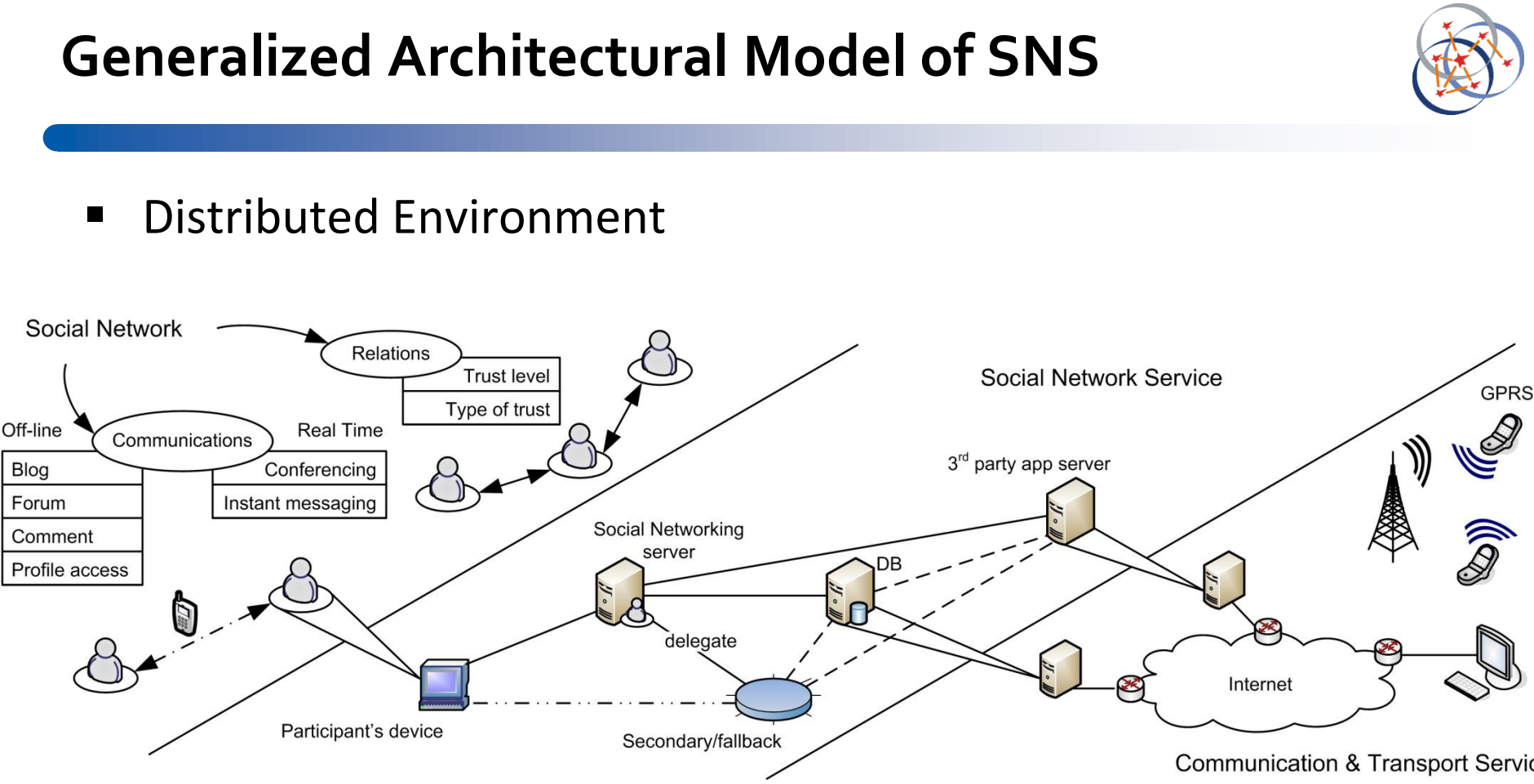




1. *Only trusted parties can access:*
   - *Personal Information*
   - *Communication*
2. *Possibility to hide any data about any user*

1. *Protect user data from unauthorized modification and tampering*
2. *Ensure the link between OSN accounts and people*

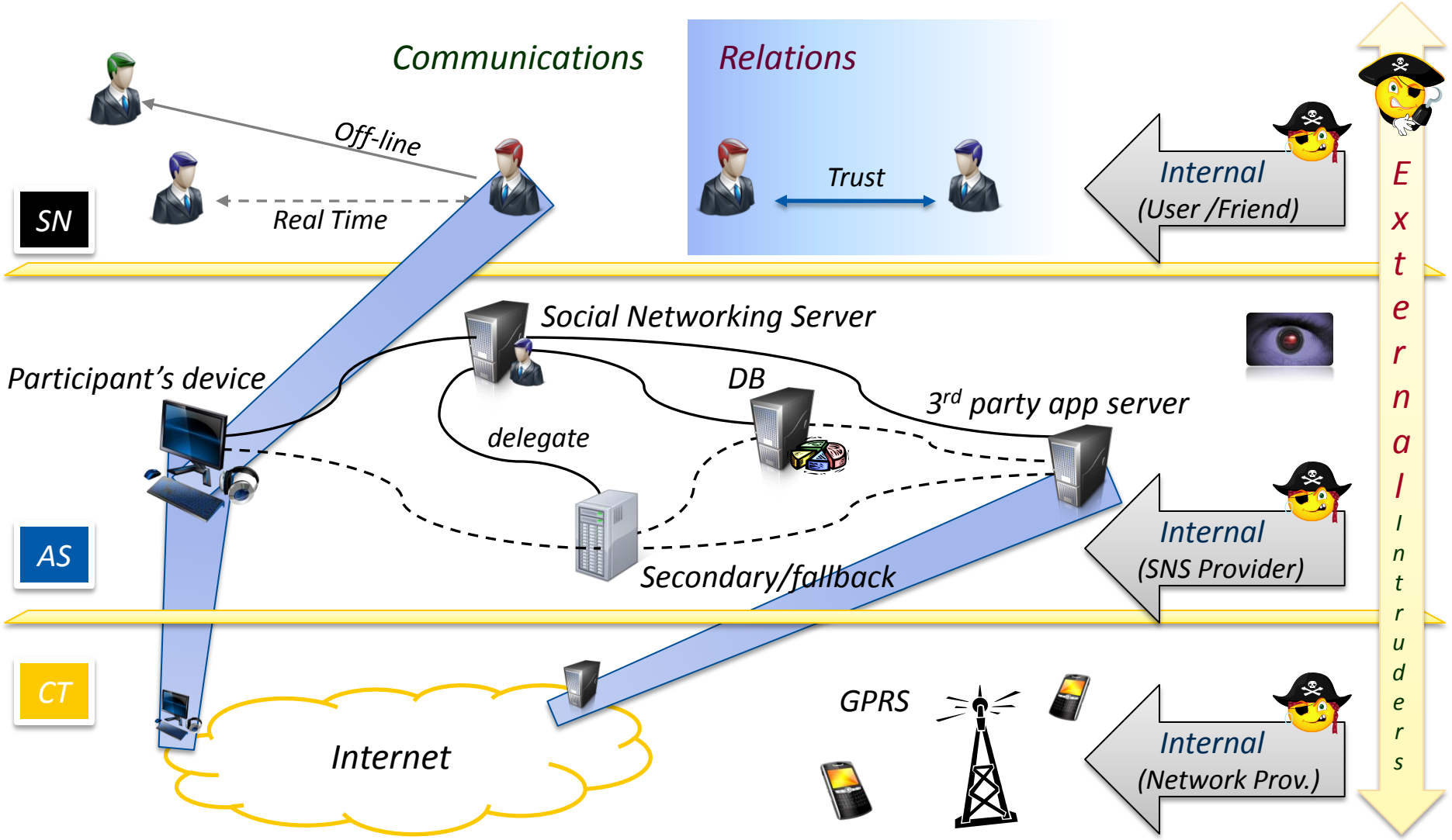
Generalized Architectural Model of SNS



- Distributed Environment

- Main SNS centralized
- Additional 3rd party application servers

Security Analysis of OSNs

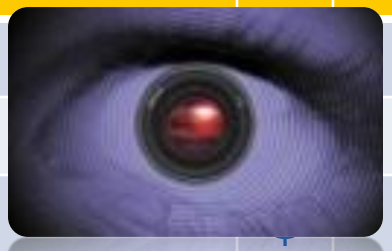


Current Status of OSNs

- *Ease of data leakage*
- *Ease of impersonation*
- *Limited privacy support*
- *Lack of flexibility in privacy*

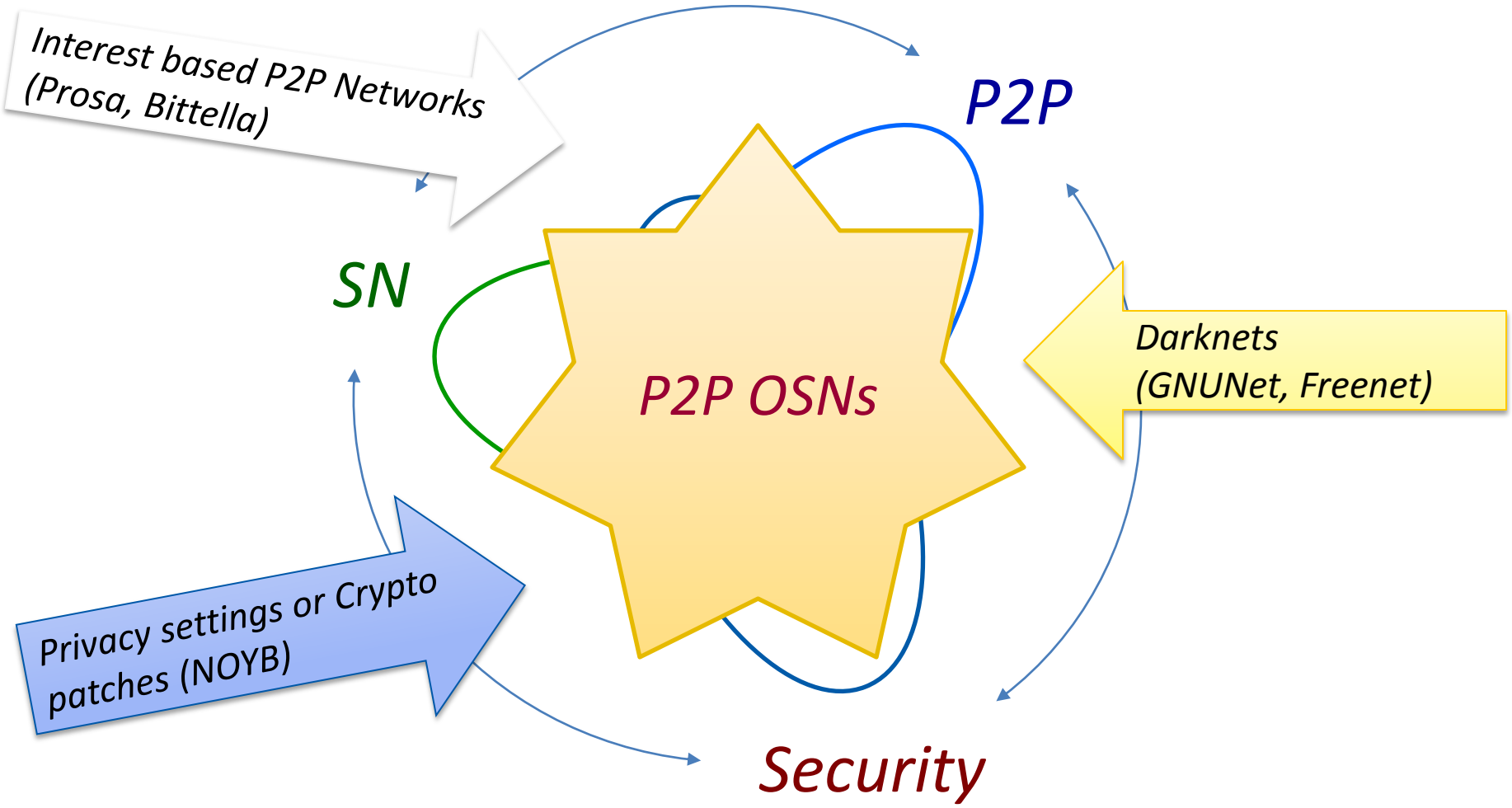
Attacks	P	I	A
Id Theft	◎	◎+	◎
Profile Cloning	◎	◎+	
Profile Porting	◎	◎+	
Fake profiles			
Profiling			
Communication tracking	◎		
Secondary Data collection	◎		

Attaks	P	I	A
Image retrieval			
Fa…on			
		◎	
Defamation		◎	
Censorship			◎
Collusion	◎	◎	◎

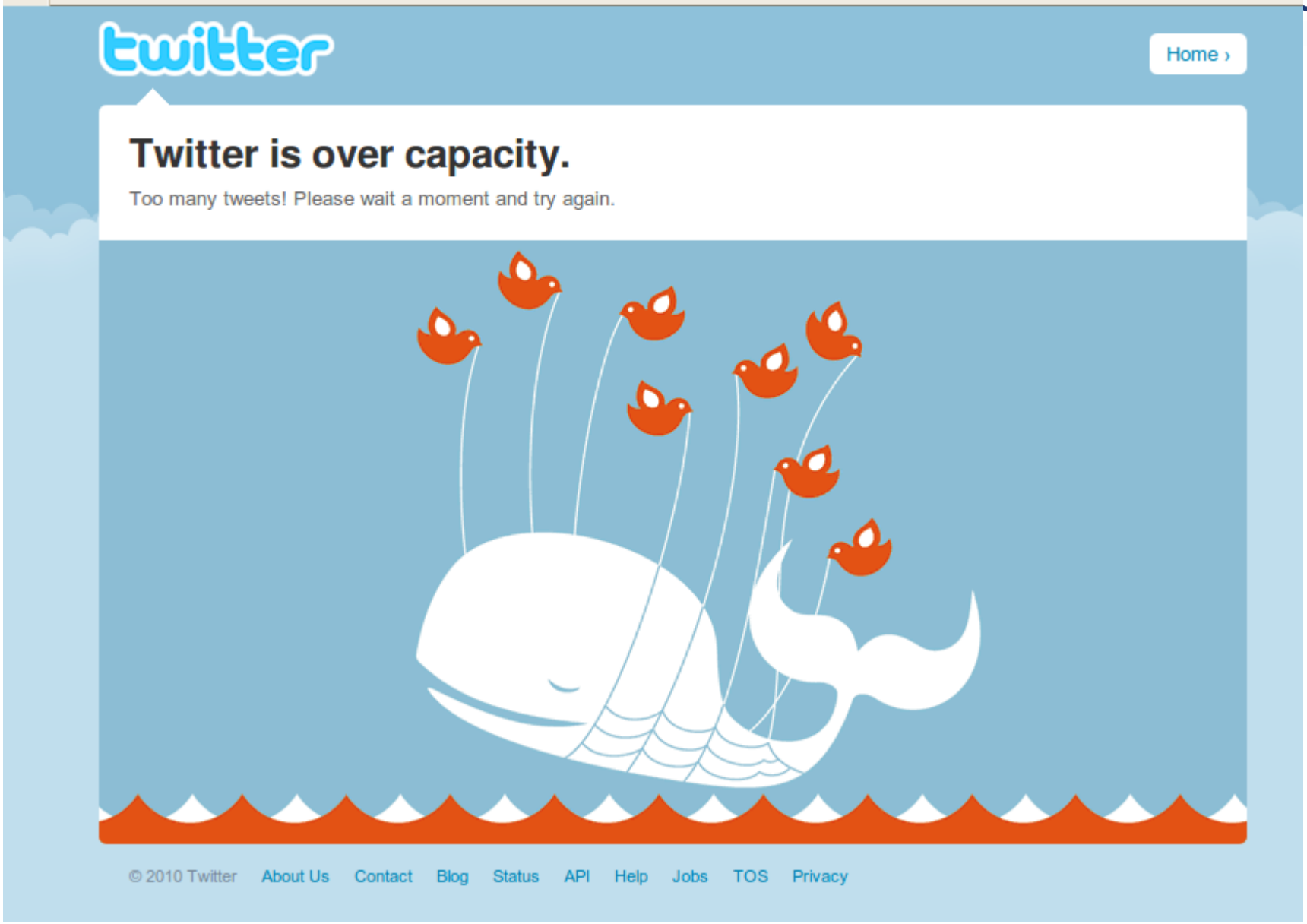


SNS provider as a "BIG BROTHER"

Current solutions



Quest to Decentralize



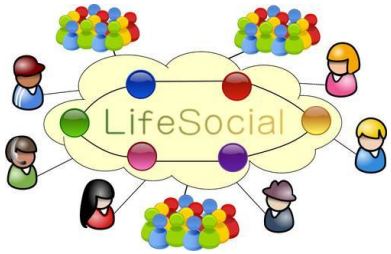
Plethora P2P / Decentralized Social Services

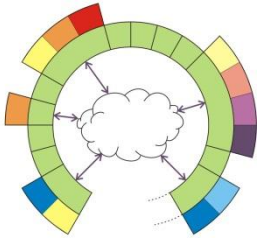
- FOSS
  - BuddyPress, CrabGrass, Cobs, DaisyChain, Diki, Elgg, FETHR, GNUNet, Gossple, Jappix, Lorea, Mycella, Movim, PeerScape, Pinax, StatusNet
- Commercial Approaches
  - diaspora, wuala, LifeSocial
- Academia
  - Friend-of-a-Friend, FriendStore, HelloWorld, LifeSocial, LotusNet (Likir), PeerSon, Safebook, SocialCircle, Tribler, Vis-a-Vis
- Focus on systems that
  - implement social networking and publication functions
  - Provide running software or a comprehensive protocol / system description
  - Actually are decentralized

Selected Systems and Proposals

- diaspora
- Friend-of-a-Friend
- LifeSocial
- LotusNet / Likir
- PeerSon
- Safebook
- Vis-A-Vis

diaspora*





PeerSoN

Classifying Decentralized OSN

- Type of storage / service provision
  - Infrastructure-based
    - Dedicated Web-Servers
    - Deployed in the cloud
  - Peer-to-Peer-based
  - Hybrid
- Granularity of service provision
  - Replicating whole service (profile and provision)
  - Distributed storage of attributes
- Level of integration
  - Stand alone system
  - Extension of existing systems
- Resource sharing incentives
  - None
  - Social cooperation
  - Payed premium services

Approach	*Type of Service*	*Service Granulariy*	*Integration*	*Incentives*
diaspora	web-based	complete	external	Premium
FoaF	web-based	complete	external	
vis-à-vis	Cloud	complete	external	
LotusNet	P2P	split	stand alone	
PeerSon	P2P	split	external	
Safebook	P2P	complete	stand alone	Social Coop
LifeSocial	hybrid	split	stand alone	Premium

Challenges for P2P OSN

- Let‘s use P2P to increase privacy in OSN!
- ***How could we proceed?***
- ***Which challenges are we about to encounter?***

- Performance
  - Properties: P2P → filesharing! OSN: small, short lived objects
  - Extreme heterogeneity (friends, uploaded content, resources)
  - User behaviour / sessions (diurnal patterns, session- and inter session times)
  - Incentives
  - Mobile users vs. p2p / recursive routing, etc…

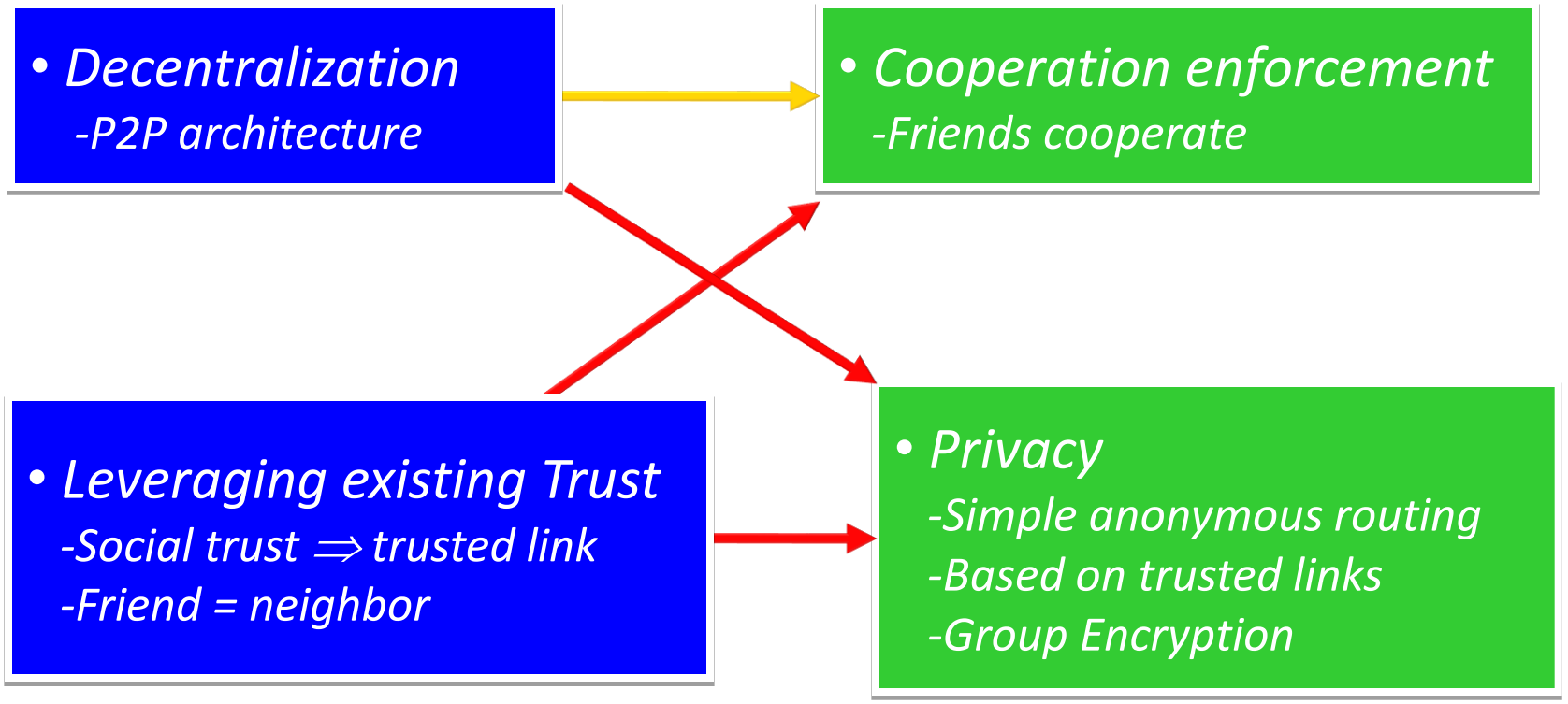
- Security
  - How can we „guarantee“ availability?
  - How do you search for somebody?
  - Key management!?
  - Identification services (we don‘t want fake accounts!)

Safebook : *Privacy-Preserving Online Social Networking*

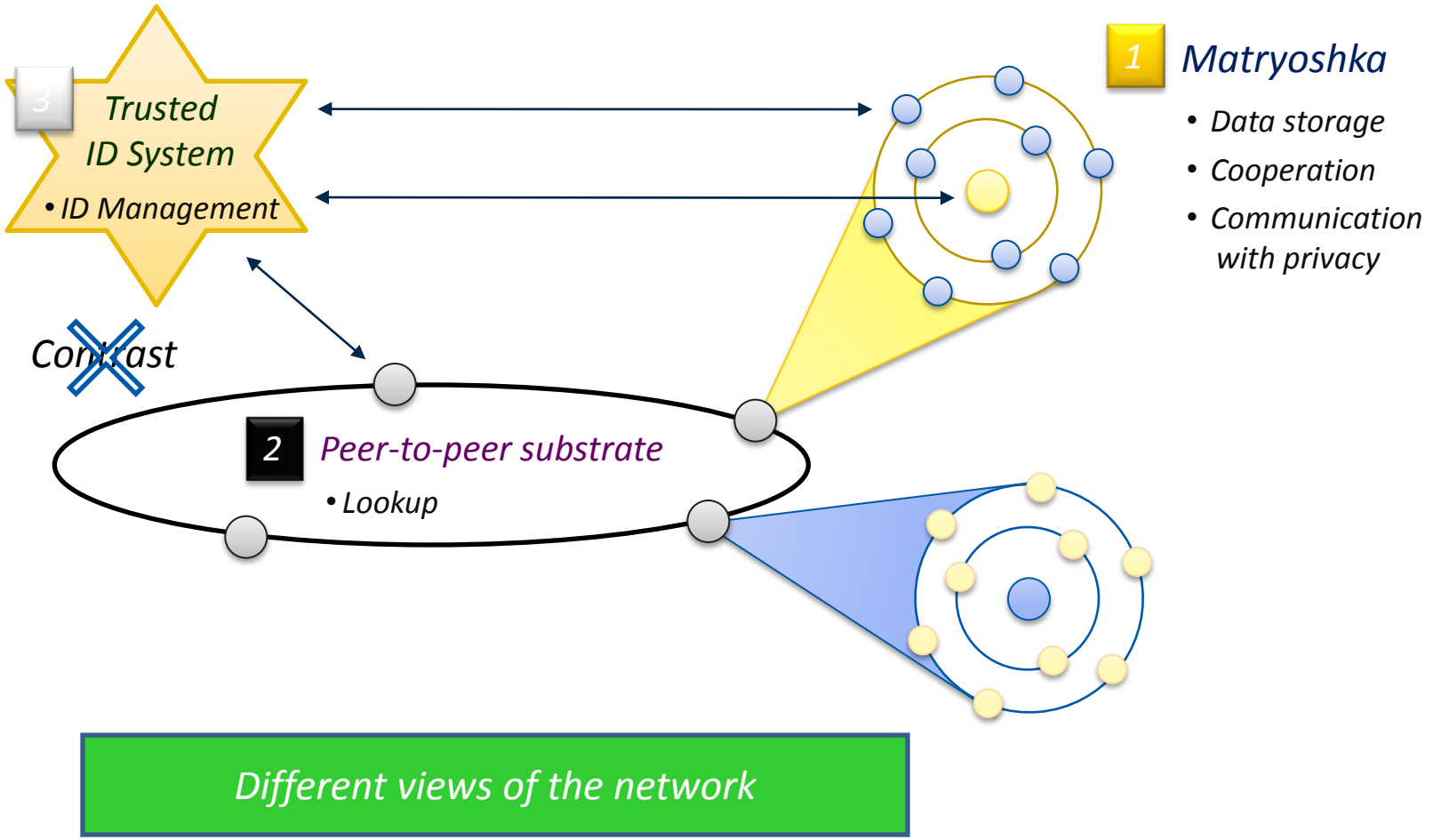
http://www.safebook.us

L. A. Cutillo, R. Molva, M. Önen, T. Strufe

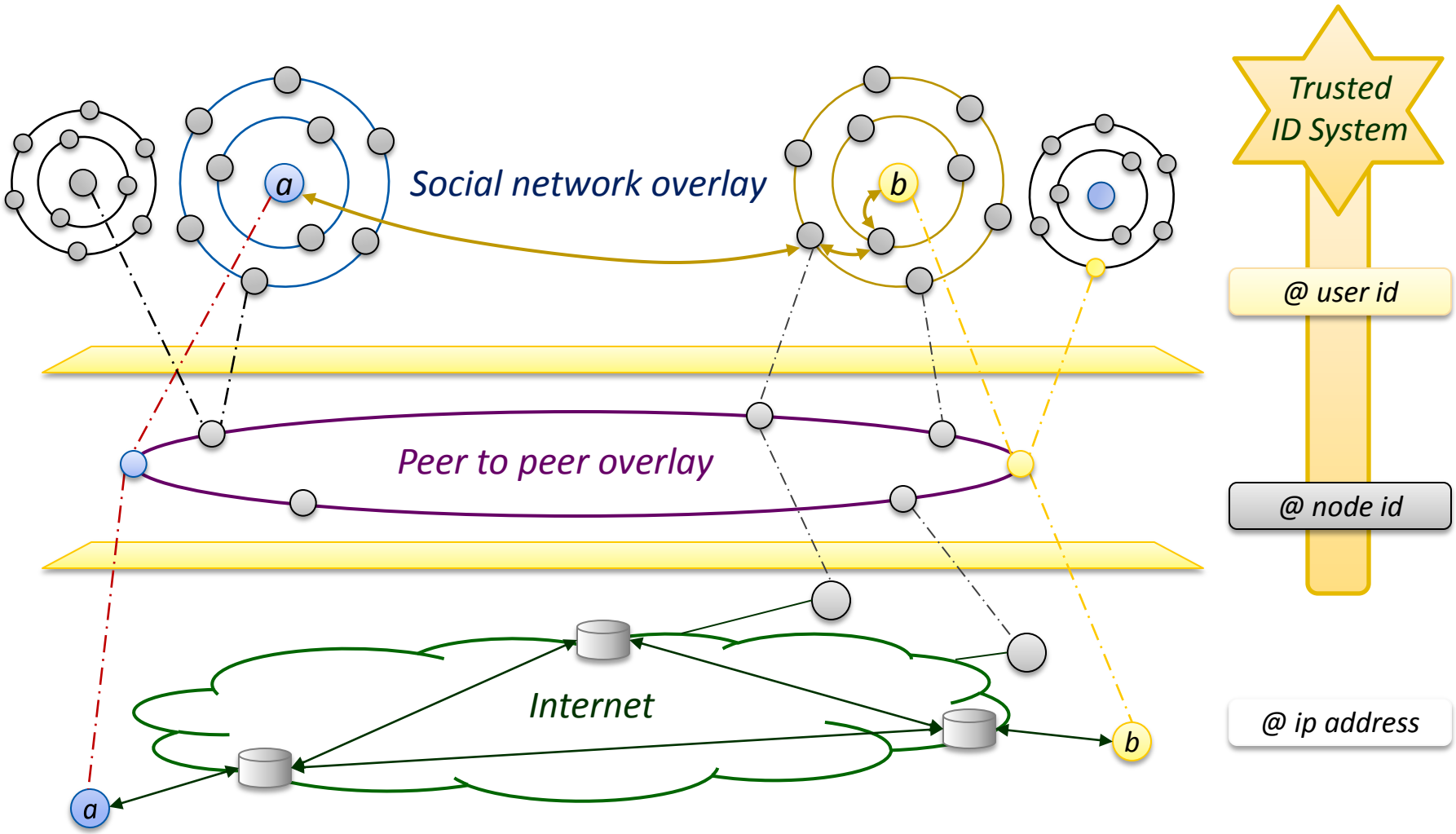
Design Principles



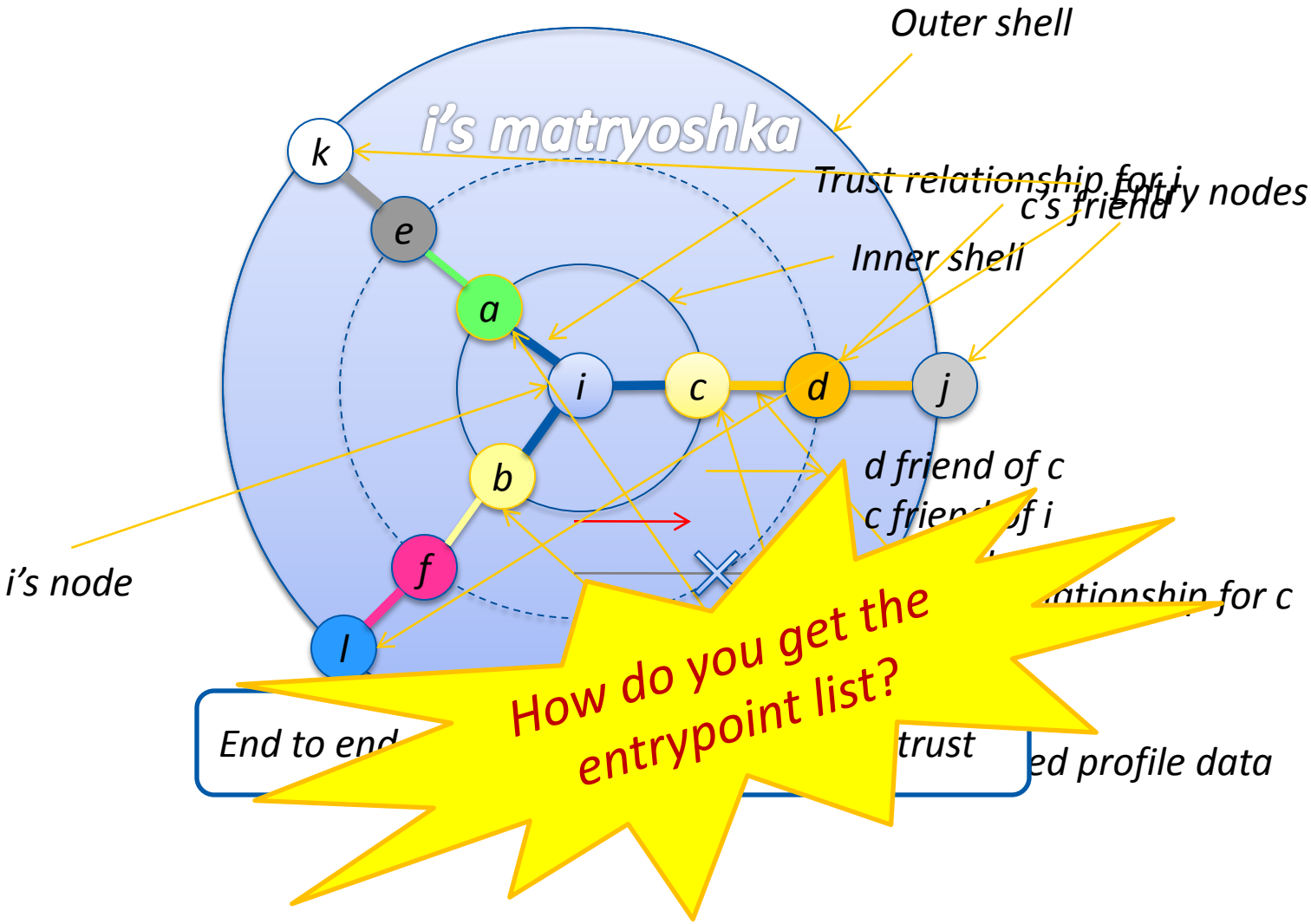
Architecture



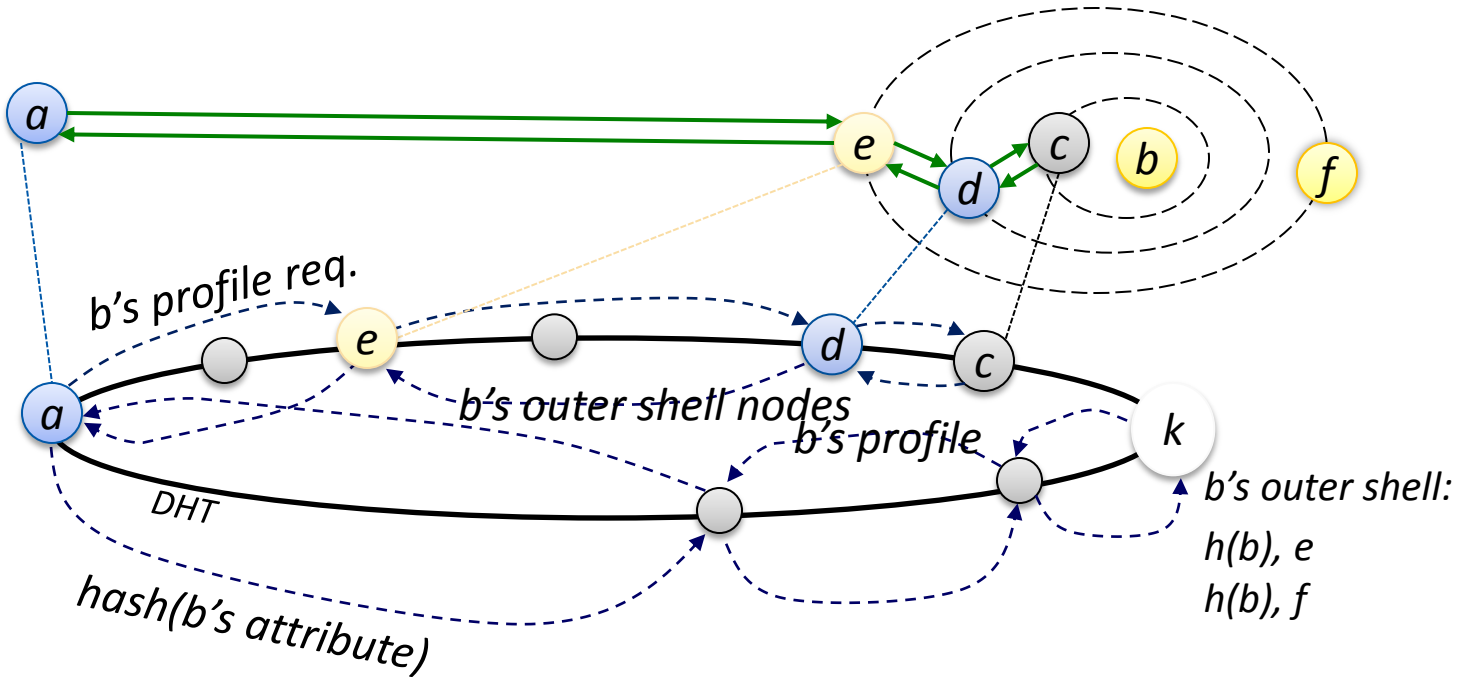
Network view



User i's Matryoshka



Finding it, using P2P: *a* looks for *b*

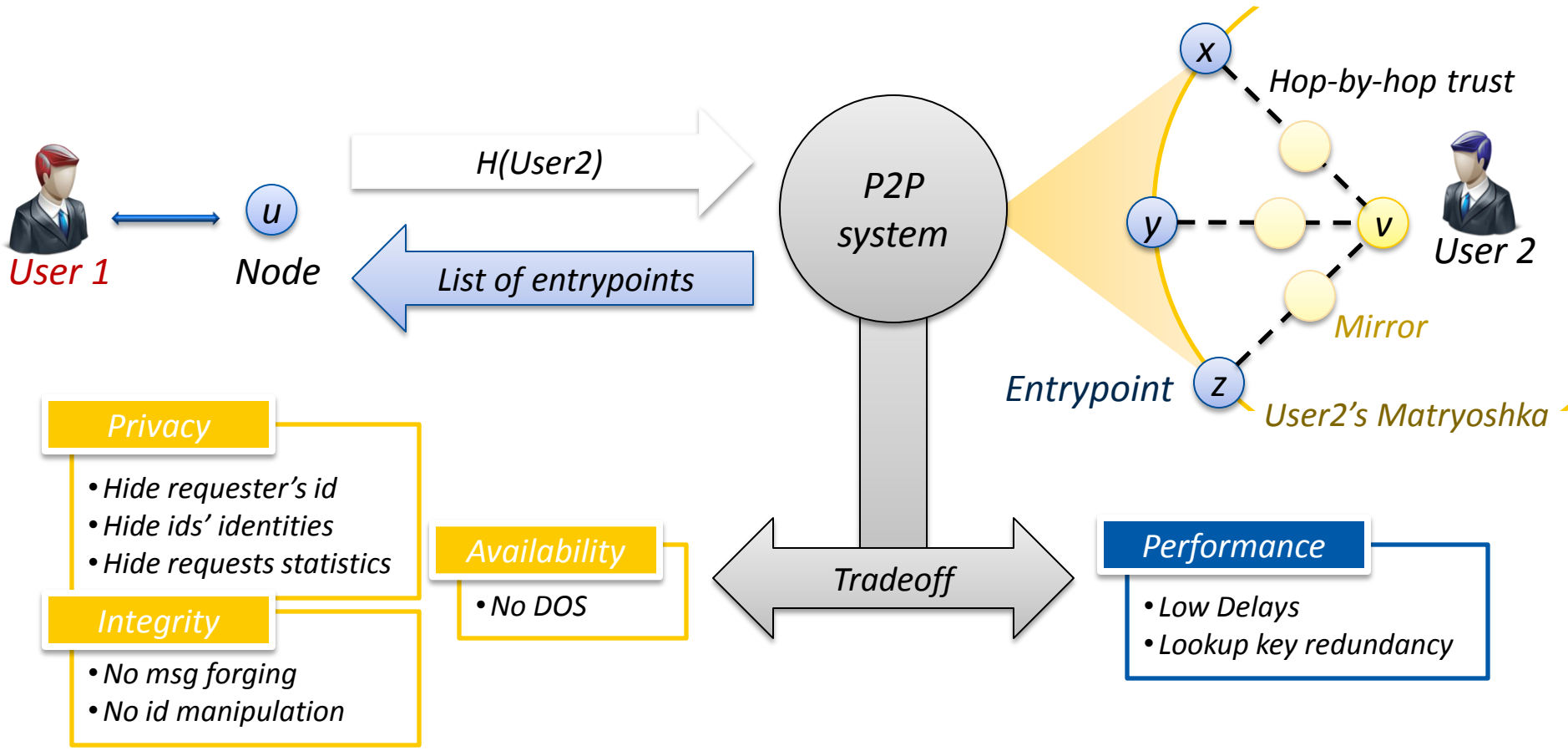


lookup
- *a looks for b's entry nodes*
- *k provides b's outer shell nodes*

data request
- *a sends profile data request to a b's entry node*

Data reply
- *One of b's inner shell nodes answers*

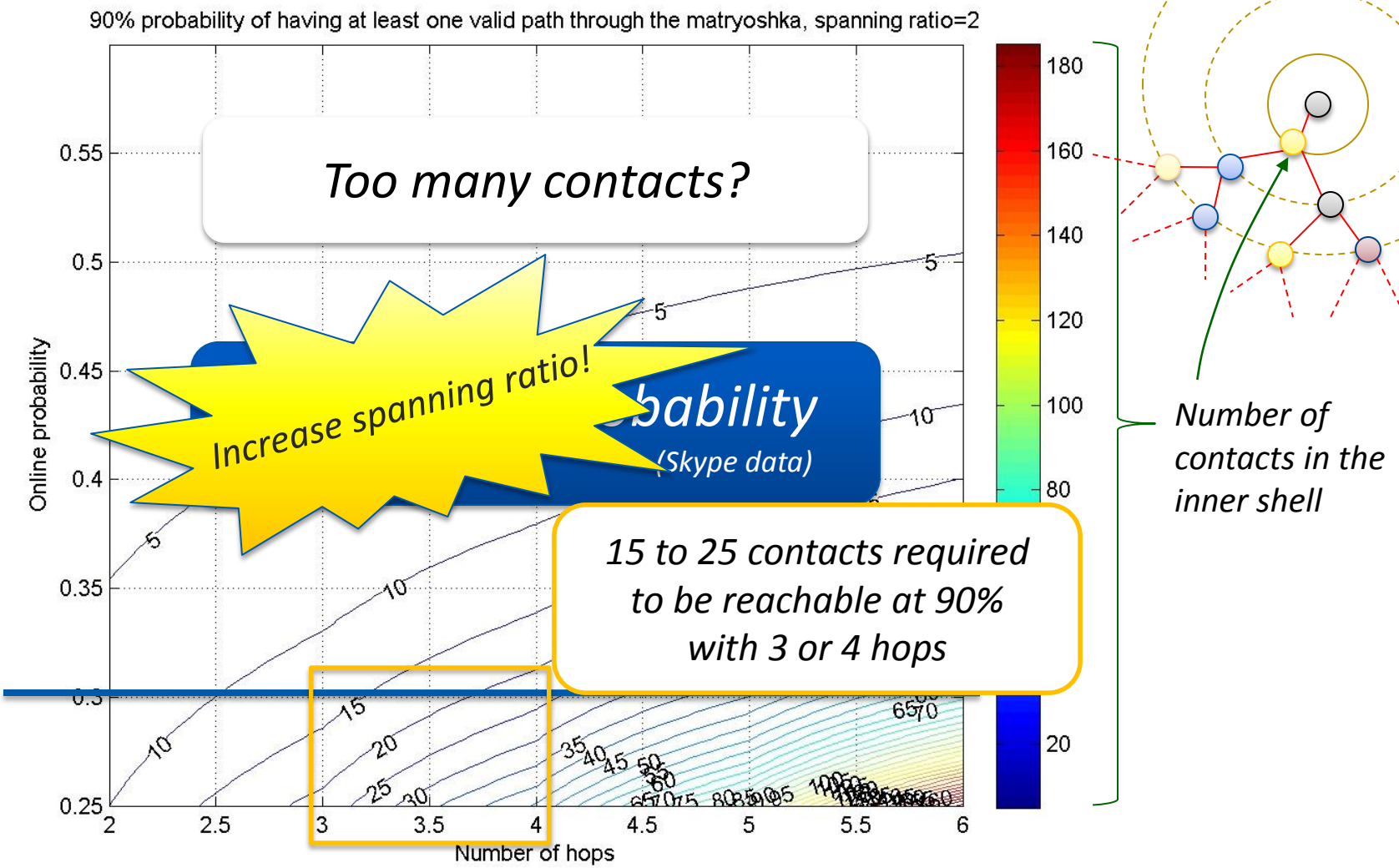
Role of P2P system



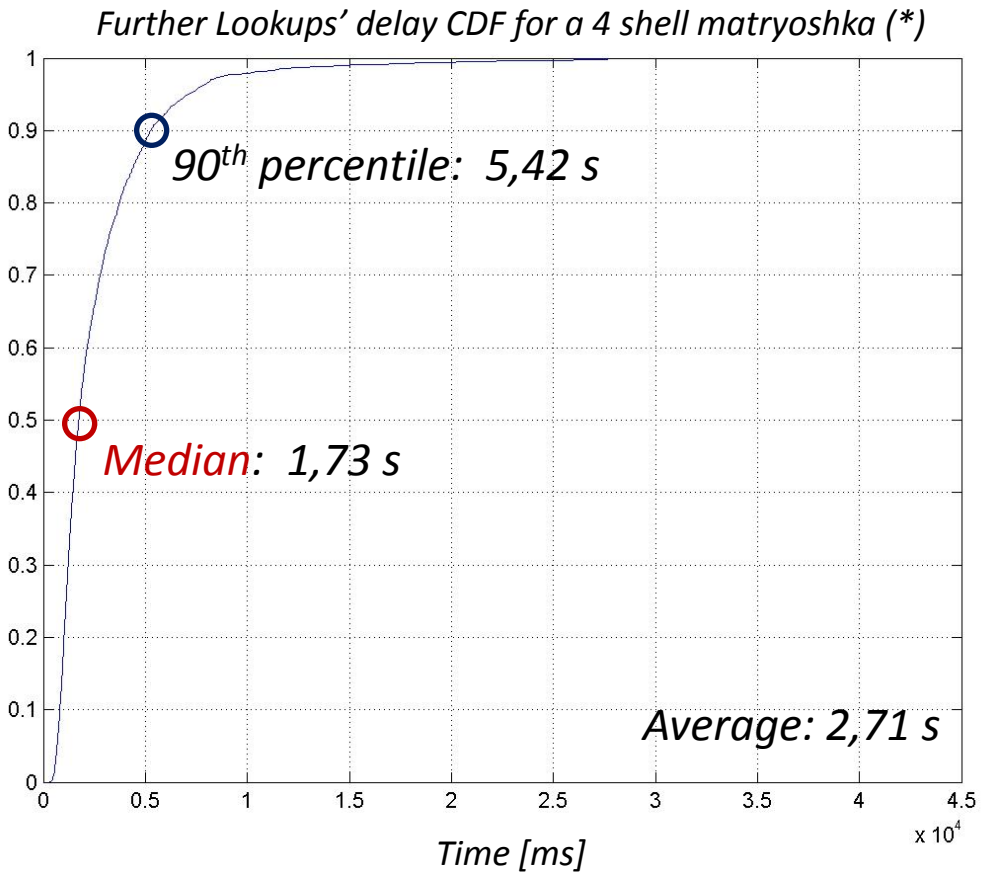
Evaluation of the scheme (1)

Privacy	*Friendship relations hidden through Matryoshkas*
	Untraceability through pseudonymity and anonymous routing
Cloning prevention	*ID management*
Dos prevention	
Access control	*Key management*
Availability	*Data replication at friends' nodes*

Performance - Reachability



Performance - Delay



Total data lookup time:

$T_{dl} = T_{DHT} + T_{Mat}$

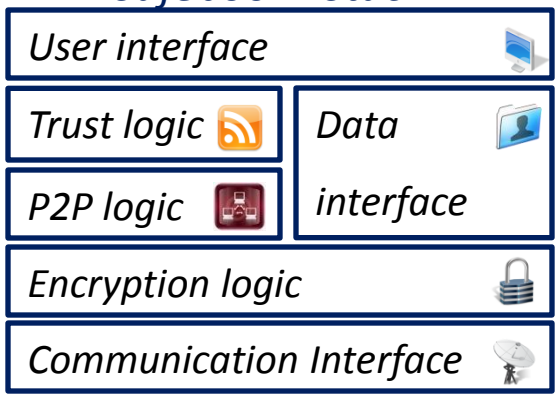
- *Further lookups: T_{DHT}=0 thanks to caching*

(*) Data computed by applying the Monte Carlo sampling technique on single hop delay measurements and on delay measurement for a successful DHT key lookup in KAD (Biersack..)

Current work on Safebook

Antonio, Thorsten, Melek, Refik PLUS Students' team: Nadja Hille, Djawad Mamedov, Jens Trinh, Paolo Viotti, Luca Boasso, Etienne Peron, Jean-Baptiste Barrau, Daniel Puscher, Felix Günther,…

Safebook "Stack"



- *Incentive Mechanisms* → *Enhance cooperation*
- *Prototyping Safebook* → *Mainly at EURECOM*
- *Simulating Safebook* → *Partly in Darmstadt*
- *Modelling Safebook* → *Protocol optimizations*
- *Modelling OSN and Users* → *Understand the Users first*

Start … *End*

Summary

- Online Social Networks very popular
- Important to know user behavior
- Privacy issues a serious problem

- Decentralization one possibility
- Large number of projects that leverage the “social” term

- Good (even though challenging) example for an application that gains from the properties of P2P

- A few real approaches to decentralize
  - Privacy, availability, cost (load balancing)

- Safebook as a case study