



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



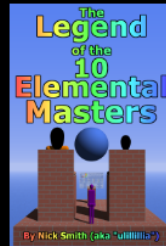
- What are Online Social Networks and how have they evolved?
 - A short overview of the history of networked services
- Why are they interesting for research?
 - A peek at their prevalence
- How do people actually use them?
 - Characterizing user behavior in OSN
- The privacy problem
- Movements to decentralize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



```
Running On: Wildcat v5/winserver          Connected To node:2
Your Sysop: Frog Man                    Today's Date Is: 3/26/202
Telnet: bbs.frogland.net                The Time Is: 4:47 PM
WWW: http://www.frogland.net
>Note: Check Out http://www.bbsdot.com slashdot for BBS's's!!!!!!!!!!!!
      [ Only One Account Per Person ]

What is your user name? █
```

...around came the World Wide Web...



Major announcement

My book is now on sale! The page detailing about my book has been updated as well, now including the full size version of the front cover image.

Announcement

It's now official - my book is going into book stores. When you'll see it, that I don't know. Lulu states 6 to 8 weeks so this implies a time from April 13 to April circumstances - I couldn't order my second proof copy. **Updated Mar 3, 2010**

Ulimilab City Software logo - a space city drifts freely among the background stars with colorful 3D text in front.

1 Welcome!

Welcome to Ulimilab City! In Ulimilab City, you'll find tips and tricks that'll help with math, computer operation, and a few others. I have free games you can make materials like dice or marbles. My extensive dream journal with over 750 entries is probably one of the largest online dream journals. From school and my game, I have some stories available. Along with that, I have two computer games I'm working on, "The Supernatural Olympics" and a 2D RPG game. Best website. The only ads, my own creations from 2002, are in one isolated area. My site is best viewed at 800x600 resolution at true color (24 or 32-bit color) 1024x768 is optimal.

Lost? Don't know how to navigate? I've explained [the basic design](#) of my site (needs to be updated) and how it's organized from page to page. The category

Note: My website is currently undergoing a redesign so some pages, especially those that haven't been updated in at least one year, will have a somewhat ending with ".html" are the very old documents not having been updated since July of 2005 at the latest.

2 News and latest updates

Ulimilab contains the latest news and updates in more detail than described in this page and events going back when I began it in late 2001. This page only



Lou's page

1995 Annual Report

- [Letter to Investors/Financials](#)
- [The Future of Computing](#)
- [1994 Annual Report Review](#)

Financial information

- [IBM stock quotation](#)
- [Quarterly results](#)
- [Stockholder services](#)

Employment

IBM Planetwide

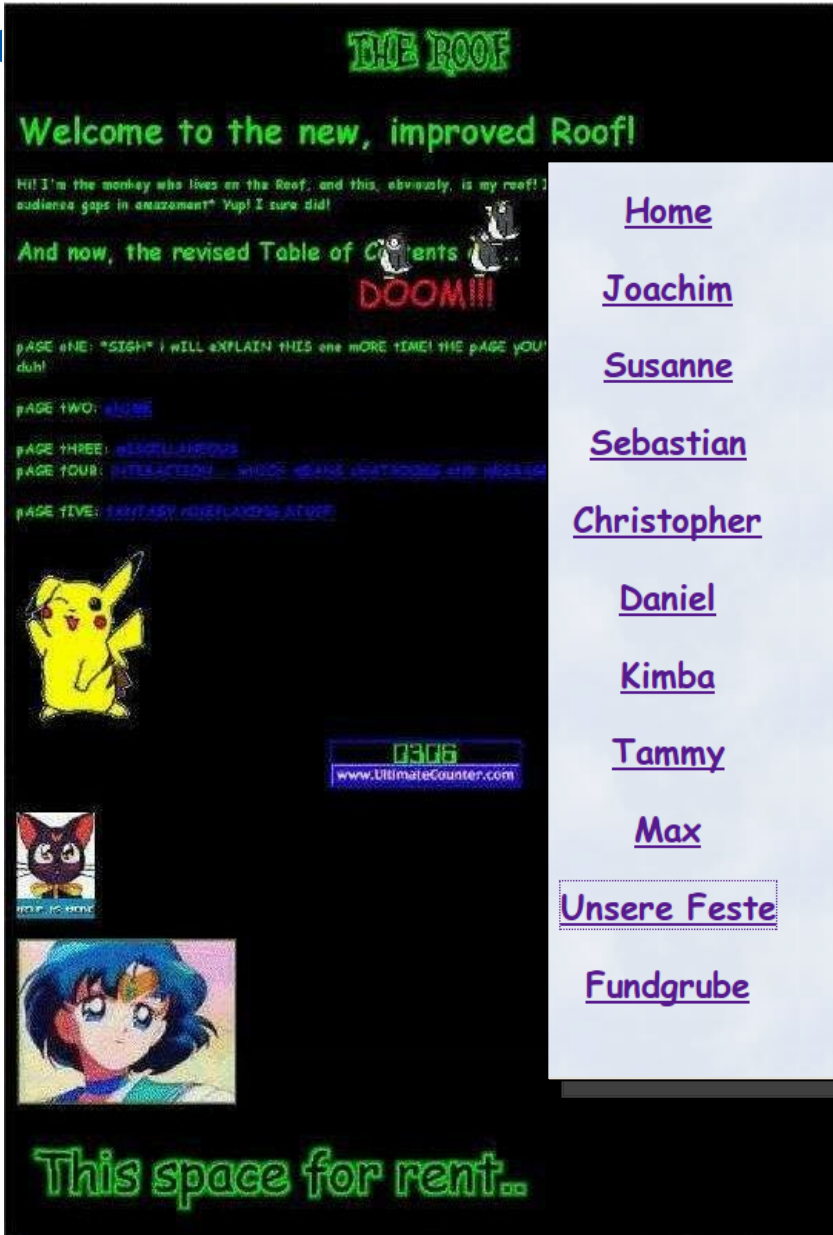
Other corporate activities

- [IBM and the environment](#)
- [IBM and the Global Information Infrastructure](#)
- [Philanthropy](#)

[[IBM home page](#) | [Order](#) | [Search](#) | [Contact IBM](#) | [Help](#) | [\(C\)](#) | [\(TM\)](#)]

<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



...with high barriers.



World Wide Web

The WorldWideWeb (W3) is a universe of documents.

Everything there is online about the project, [Mailing lists](#), [Policy](#)

[What's out there?](#)

Pointers to the world's

[Help](#)

on the browser you are

[Software Products](#)

A list of W3 project co

[Mail robot](#), [Library](#))

[Technical](#)

Details of protocols, fo

[Bibliography](#)

Paper documentation o

[People](#)

A list of some people involved

[History](#)

A summary of the history of the

[How can I help?](#)

If you would like to support the

[Getting code](#)

Getting the code by [anonymous](#)

Welcome to

FIAT

Benvenuti sul Web-server FIAT !



Il server è attualmente in sviluppo: la
[home-page](#) definitiva,
per ora riservata ai gestori del server, sarà disponibile prossimamente.
Per ulteriori informazioni Vi invitiamo a visitare il server di
[ITS s.r.l.](#)
oppure potete scrivere a webmaster@its.it

FIAT Copyright © 1995.

Take a look at a demo of one of our latest technologies, ProVision.


BBN ON THE WORLD WIDE WEB

We get the Internetworking for your business

Press Releases | Investor Relations | Jobs

Customer Connection | Business Partners | BBN Features

BBN PROVISION



Bandwidth On Demand

ALSO INSIDE

[BBN Planet Customer Support Online!](#)

[BBN Corporation 1996 Annual Meeting](#)

[BBN Planet Network Map](#)

<http://www.fiat.it>



people would line up to look at old day at the museum-- course you ne world at the end (those'll be ne 'Net.) So the museum heads artwork, etc. Some of the old ad lines are priceless-- like "Drink Coca-Cola," an idea from '86 (that's 1886) that left no room for misinterpretation, and 1906's "The Great National Temperance Drink." What?

<http://www.cocacola.com>

So Today *everybody* Shares Some Data...



Facebook Home Profile Friends Inbox Thorsten Strufe

DOPPLR FOR YARICK Overview | Your connections | Your activity | invited to V!!! on Wednesday

Where Next? { You have no trips planned. Top Friends

LinkedIn People | Jobs | Answers | Companies Account & Settings | Help | Sign Out Language

Account Type: Basic | Upgrade » Search People Search Advanced

Applications BETA FAQ | Feedback | Browse More Applications

My Travel

Application Preview

Your upcoming trips
December 2008
New York City, NY — Dec 13 - Dec 28
Flying to NYC for work, then down south to visit friends and family on the east coast.

Who's close to you

Scott Hintz
Triplt
Nov 18 - Nov 20, 08

Andy Denmark
Triplt

See where your entire professional network is traveling and when you will be in the same city as your colleagues. Meet up at the next industry event or re-connect with old friends. Add the My Travel application to display your current location, upcoming trips and travel stats within your network.

Raves about Triplt:

- "Triplt's straightforward approach makes it addictive..." – Walt Mossberg Solution, Wall Street Journal
- "Best for organizing travel details..." – CNN, Travel + Leisure
- "Easy and damn useful..." – DailyCandy
- "A terrific site that has changed my travel life..." – WNBC New York TV

half of 2009, be sure to add any past trips during 2009 that you might have missed. [Add your past trips here.](#)

[View your full journal](#)

Your Personal Annual Report

Madrid, Spain
Madrid has 12 questions.
[+ Ask a question about Madrid](#)

Frankfurt, Germany
Frankfurt has 12 questions.
[+ Ask a question about Frankfurt](#)

Hamburg, Germany
Hamburg has 11 questions.

Application Info

Developed By Triplt Inc.

Category Utility

When you add an Application, you are allowing it to access profile information about you and your connections which the Application uses to provide the service. This and other information you provide to the Application may be displayed publicly depending on the Application's functionality.

By proceeding, you confirm your acceptance of the [LinkedIn Terms of Use](#) and you acknowledge that your use of the Application is subject to the Application developer's terms of use and to the Application developer's privacy policy.

Amadeus

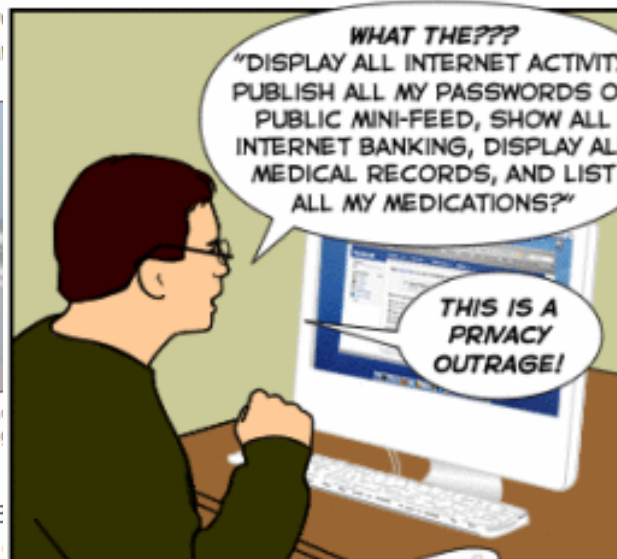
MUSIC LOVERS, SNL, France à l'ONU direct loser we love to hate,

...with calculated side effects...



The Joy of Tech™

by Nitrozac & Snaggy



joyoftech.com

3/26/facebook-e-robbery/



11/

robbery

ow you choose y
reveal.

cent-seeming stat
atment."

and called Fire De
g was caught on c
oked suspiciously

Friday's American

...increasingly immersive to daily life...



go



Loop



tweetspotting

„Volkszählung“ 1987

- Last public census in Germany
- Scheduled for 1981 (delayed)
- Significant public opposition
 - Fear of a surveillance society
 - The transparent citizen („gläserner Bürger“)
 - Bounty for discovered German citizens
 - Appeal for civil disobedience
 - Finally accomplished in 1987
 - Consequence: „25% inherent tax“
 - Significant gap between census and reality
 - (So let's just get a unifying tax)

VOLKSZÄHLUNG 1987
Personenbogen

Bitte Gemeinde angeben: _____

1 Geburtsangaben

a) Geburtsjahr: _____
b) Geburtsmonat: 1. Januar bis 24. Mai / 25. Mai bis 31. Dez.

2 Geschlecht
männlich / weiblich

3 Familienstand
ledig / verheiratet / verwitwet / geschieden

4 Rechtliche Zugehörigkeit zu einer Religionsgesellschaft
Römisch-katholische Kirche / Evangelische Kirche / Evangelische Freikirche / Jüdische Religionsgesellschaft / Islamische Religionsgemeinschaft / andere Religionsgesellschaften / keiner Religionsgesellschaft rechtlich zugehörig

5 Welche Staatsangehörigkeit haben Sie?
deutsch / griechisch / italienisch / übrige EG-Staaten / jugoslawisch / türkisch / sonstige/keine

6 Wird von Ihnen noch eine weitere Wohnung (Unterkunft/Zimmer) in der Bundesrepublik Deutschland einschließlich Berlin (West) bewohnt?
ja / nein

7 Sind Sie erwerbstätig?
Vollzeit (über 36 Std. in der Woche) / Teilzeit (bis zu 36 Std. in der Woche)
arbeitslos, arbeitssuchend / nicht erwerbstätig
Hausfrau, Hausmann / Schüler(in), Student(in)

8 Leben Sie überwiegend von
Erwerbs-, Berufstätigkeit / Arbeitslosengeld, -hilfe / Rente, Pension / eigenem Vermögen, Vermietung, Verpachtung, Altenteil / Zuwendungen, Unterhalt durch Eltern, Ehegatten usw. / sonstigen Unterstützungen (z. B. Sozialhilfe, BAföG)

9 Welchen höchsten allgemeinen Schulabschluß haben Sie?
Volksschule, Hauptschule / Realschule/gleichwertiger Abschluß (z. B. Mittlere Reife) / Hochschulreife (Abitur), Fachhochschulreife

10 a) Welchen höchsten Abschluß an einer berufsbildenden Schule oder Hochschule haben Sie?
Berufsfachschule (ohne Berufsschule) / Fachschule / Fachhochschule (Ing.-Schule, höhere Fachschule) / Hochschule (einschließlich Lehrerausbildung)

b) Welche Hauptfachrichtung hat dieser Abschluß?

11 Falls Sie eine praktische Berufsausbildung (z. B. Lehre) abgeschlossen haben:
a) Auf welchen Lehrberuf bezog sich diese Ausbildung?
b) Wie lange dauerte diese Ausbildung? Jahr(e):

12 Bitte Name und Anschrift Ihrer Arbeitsstätte oder Schule/Hochschule angeben.
Name: _____
Straße/Hausnummer: _____
PLZ: _____ Gemeinde: _____

13 Welches Verkehrsmittel benutzen Sie hauptsächlich (längste Strecke) auf dem Hinweg zur Arbeit oder Schule/Hochschule?
kein Verkehrsmittel (zu Fuß) / Fahrrad / Pkw / U-Bahn, S-Bahn, Straßenbahn / Eisenbahn / Bus, sonst. öffentl. Verkehrsmittel / sonstiges (Motorrad, Moped, Mofa)

14 Wieviel Zeit benötigen Sie normalerweise für den Hinweg zur Arbeit oder Schule/Hochschule?
entfällt, da auf gleichem Grundstück / unter 15 Minuten / 15 bis unter 30 Minuten / 30 bis unter 45 Minuten / 45 bis unter 60 Minuten / 60 Minuten und mehr

15 Sind Sie zur Zeit tätig als
Facharbeiter(in) / sonstige(r) Arbeiter(in) / Angestellte(r) / Auszubildende(r) / Beamter/Beamtin, Richter(in), Soldat, Zivildienstl. (auch Beamtenanw.) / mit bezahlten Beschäftigten / Selbständige(r) / ohne bezahlte Beschäftigte / mithelfende(r) Familienangehörige(r)

16 Zu welchem Wirtschaftszweig (Branche, Behörde) gehört der Betrieb (Firma, Dienststelle), in dem Sie tätig sind?

17 Welche Tätigkeit, welchen Beruf üben Sie aus?

18 Falls Sie eine Nebenerwerbstätigkeit ausüben, handelt es sich um eine
landwirtschaftliche / nichtlandwirtschaftliche

NUR VOM STATISTISCHEN LANDESAMT AUSZUFÜLLEN

Geburtsjahr (1 a) 2., 3. u. 4. Stelle: _____ Hauptfachrichtung (10 b): _____ Erlernter Beruf (11 a): _____ Dauer (11 b): _____

Arbeitsstätte, Schule/Hochschule (Pendler) Land (12) Gemeinde (12) Straße (12) Hausnummer (12)

Wirtschaftszweig (16): _____ Ausgeübte Tätigkeit (17): _____

... completely willingly...



HVA/MA/s. Abl. A K G Ge

BV/N G e r a M1S Tgb.-N

13. MAI 1982

Tgb. Nr. 0002

Vielbezug

Erfassungsbeleg
(Mit Schreibmaschine oder in Deutschdruck ausfüllen)

012742

1. Grunddaten zur Person

1.1. Weber, Martin
Name, Vorname, Geburtsname, weitere Namen

1.2. Deutsche, Freidenker, Splitter
Organisation, Freidenker, Splitter

1.3. 18. Januar 1940
Geburtsdatum (Geburtsort, Geburtsort)

1.4. Deutschland, Ottendorf
Geburtsort, Geburtsort, Eltern, Geburtsort

1.5. männlich
Geschlecht

1.6. deutsch
Nationalität

1.7. BRD
Geburtsort, Geburtsort, Geburtsort

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

PK. 1 Wird der Verbindung zum Geheimdienst

PK. 4 Teilnehmer an der Tätigkeit einer zu

Organisation

BSIU 000005

1. Zuordnung zu Geheimdiensten und Zentren der politisch-ideologischen Divergenz

2. Identifizierung eines spezifischen Geheimdienstes

BS 0000

BSIU 000007

10. Gegen die Person durchgeführte Maßnahmen

10.1. Charakter der Maßnahme

10.2. Von wann wurde die Maßnahme durchgeführt

10.3. Wie wurde die Maßnahme durchgeführt

11. Zusätzliche Angaben

3.5. Übergibt Angaben zu Fahnenflüchtigen an eine amerikanische Dienststelle

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

12. Die Angaben beziehen sich auf das Jahr 1981

13. Die Person ist erlaubt 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 AGI AGI/Beauftragter der AGI

BSIU 000005

1. Zuordnung zu Geheimdiensten und Zentren der politisch-ideologischen Divergenz

2. Identifizierung eines spezifischen Geheimdienstes

BS 0000

BSIU 000007

10. Gegen die Person durchgeführte Maßnahmen

10.1. Charakter der Maßnahme

10.2. Von wann wurde die Maßnahme durchgeführt

10.3. Wie wurde die Maßnahme durchgeführt

11. Zusätzliche Angaben

3.5. Übergibt Angaben zu Fahnenflüchtigen an eine amerikanische Dienststelle

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

12. Die Angaben beziehen sich auf das Jahr 1981

13. Die Person ist erlaubt 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 AGI AGI/Beauftragter der AGI



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



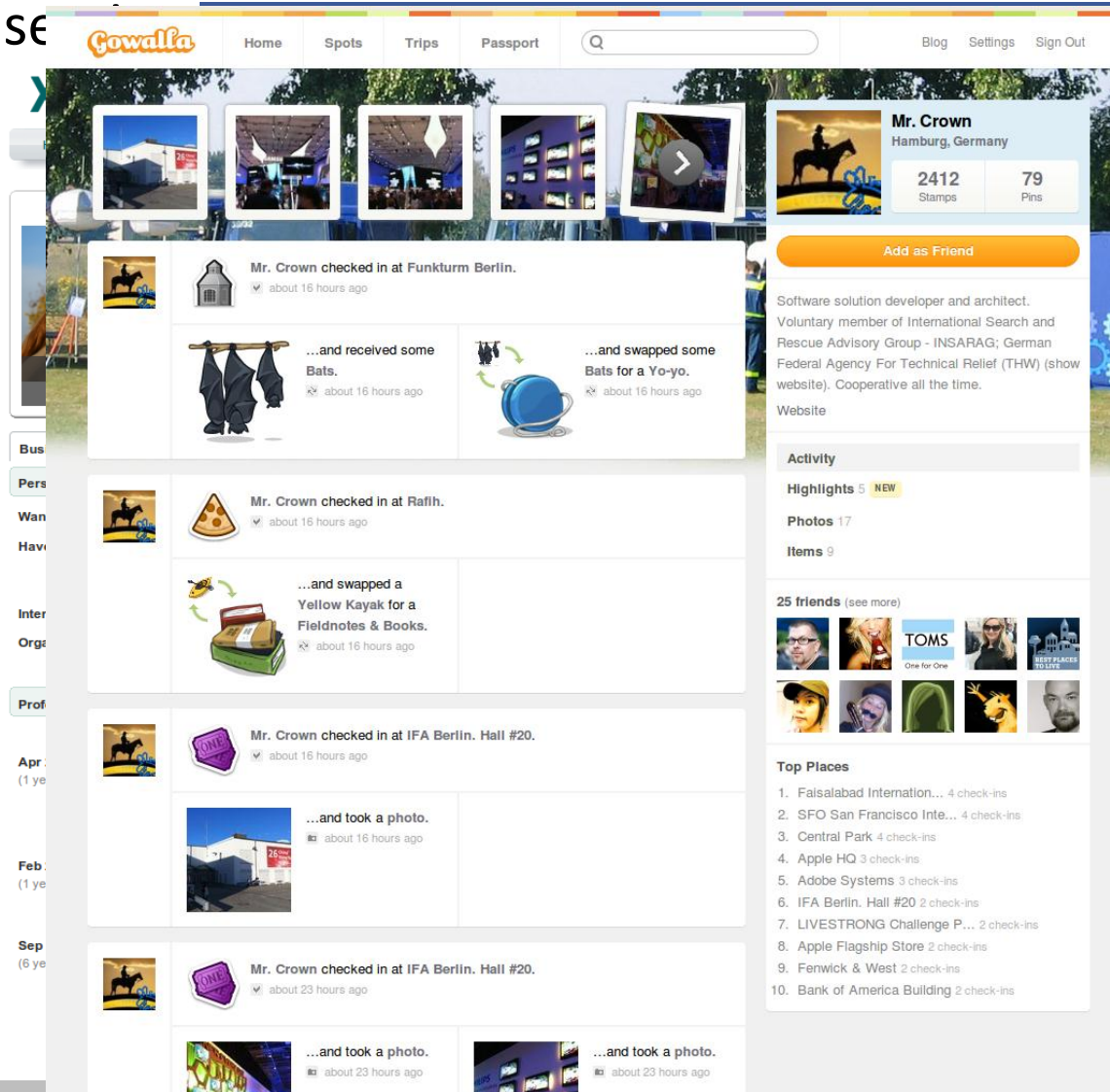
- Simplified, walled-garden version of „the Web“:
 - Easy to set-up pages („**profiles**“) of **individuals** (... and companies...)
 - Links reflecting **real-world relations** between individuals
 - Possibility to share user generated content
 - ...including messaging
 - “Guest book” / “Wall” (asynchronous 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)



Target Audience and Domain



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



Why Bother?



The

CNNMoney.com™
A Service of CNN, Fortune & Money

Symbol

Get Quo

March 2006
Work begins

March 2010
10 billionth

Top 10 Web Brands for January 2010 (U.S., Home and Work)

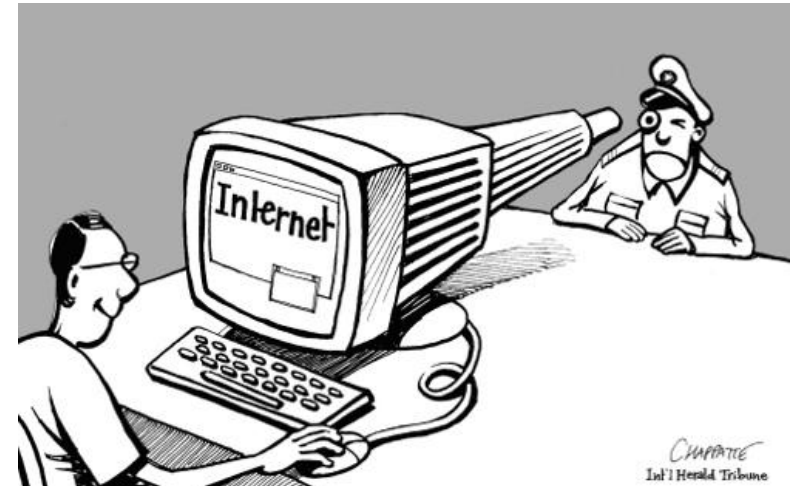
RANK	Brand	Unique Audience (000)	Time Per Person (hh:mm:ss)	MOM UA % Change	MOM Time % Change
1	Google	152,708	1:23:54	4.10%	-16.90%
2	Yahoo!	134,561	2:28:11	4.30%	-26.80%
3	Facebook	116,329	7:01:41	5.80%	9.70%
4	MSN/WindowsLive/Bing	109,425	1:25:22	1.20%	-18.10%
5	YouTube	99,525	1:02:27	7.60%	-10.30%
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

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



XING Logout | Invite contacts | Help & Contact | English

Search by name, company, and more **Search**

Home Search Messages **Contacts** Groups Events Jobs Companies Best Offers

Thorsten Strufe
Prof. Dr.-Ing. (Dipl.-Inf.)
Fachgebiet Peer-to-Peer Netzwerke
TU Darmstadt
Hochschulstraße 10, 64289 Darmstadt, Germany
+49-6151-164557
Local time: 04:48 PM

Upload photo
No photo

Searches related to my profile

- Members who recently visited my profile
- Members whose "wants" match my "haves"
- Members whose "haves" match my "wants"
- Members who know several of my contacts
- Current and former colleagues

Status update
Premium Members can enter a status update here (to let your network know you're at a conference or on vacation, etc.).

I'm a XING member because I want to ...
Let others know why you are on XING, and boost your networking potential.
Edit motivations

Settings
1 Profile settings

Options
1 Switch on profile preview 1 Print profile

Confirmed contacts: 283

Thorsten Strufe's statistics
No Premium Membership
Member since: Jan 2005
Profile hits: 3,534
Activity meter: 90%

My Groups
Edit group visibility

Visitors to my profile
(only visible to you)

Business details Confirmed contacts (83) About me Guestbook Applications

Personal information

Wants	PhD students	Edit
Haves	PhD positions in the fields of networking / distributed systems, especially security / DoS resilience of P2P systems and privacy in social networking services	Edit
Interests	P2P online social networks, skiing, diving, kicker/footsiel, politics	Edit
Organizations	GI, IEEE, Servas, Studienstiftung des deutschen Volkes	Edit

Professional experience

+ Add entry

Apr 2009 - present (1 year, 1 month)	Fachgebiet Peer-to-Peer Netzwerke (Full time employee) TU Darmstadt, http://www.p2p.tu-darmstadt.de Industry: Academia JP + Add company size	Edit
Feb 2008 - Mar 2009 (1 year, 2 months)	PostDoc Research Engineer (Full time employee, Manager (Manager/Supervisor)) Eurecom, Sophia-Antipolis, France, http://www.sosoc.org Industry: Research, 51-200 employees	Edit
Sep 2001 - Jan 2008 (6 years, 5 months)	wissenschaftlicher Mitarbeiter (Full time employee) Technische Universität Ilmenau, http://www.tu-ilmenau.de/fakia/Fachgebiet_Telematik.317.0.html Industry: Research, Wissenschaft, Aus- und Weiterbildung + Add company size Teilhaber, Leitung IT Multiple-Choice GmbH Industry:	Edit

Identifying info

- Name
- Photo
- Address...

CV

- Current/prev. employments
- Educational track

Interests

- Personal/professional
- Wants/haves
- Interest Groups

Personal contacts

Messaging

Statistics



- 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)

The screenshot displays the XING homepage. At the top, the XING logo is visible next to a notification icon showing 0 messages and 0 profile updates. A search bar on the right prompts the user to 'Search by name, company, and...'. Below the header is a navigation menu with links: Home, Search, Messages, Contacts, Groups, Events, Jobs, Companies, and Best Offers. The main content area is titled 'What's new in your network' and contains a list of updates. These include: Tahssin Asfour's status update about a new design for www.pr-bote.de; Mario Melle joining the 'Neuer Personalausweis (nPA) für ...' group; Detlef Mämpel joining the 'Army Network' group; Dr. Thomas Rücker attending the 'Netzwerkgruppe Innovation' event; Stella Pieri connecting to Julian Hühnemann; and several updates from 'Only Premium Members' regarding group joins, address changes, and new connections. A 'more »' link is at the bottom right of the feed. On the right sidebar, there are sections for 'Welcome Th...', 'Customize', 'View search', 'Application', 'Go Premium', and 'Grow your net...', each with additional options or links.

- Note: having many friends leads to broad dissemination...



- 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



- 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)

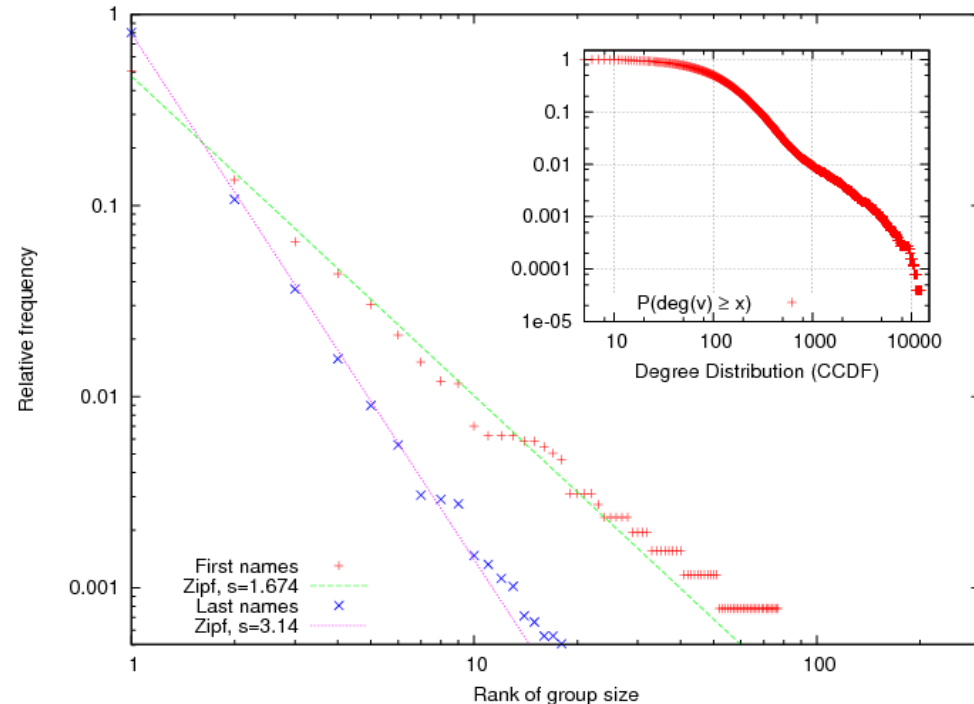
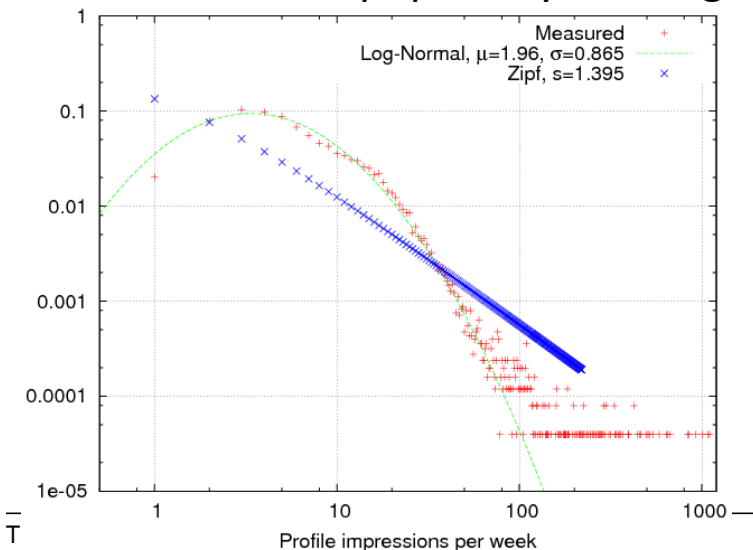
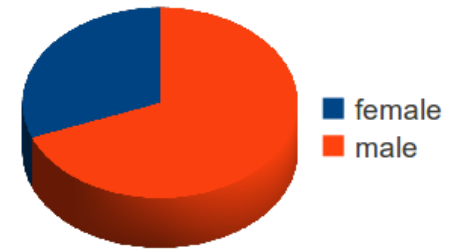


- Identifying Data
 - Name, Image, Gender
- CV
 - Current employment, universities attended, claimed spoken languages
- Interests
 - Interests as stated
 - Number of subscribed groups, subscribed groups, number of members in groups, number of messages in respective groups, languages of group
- Contact 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. (\sim 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





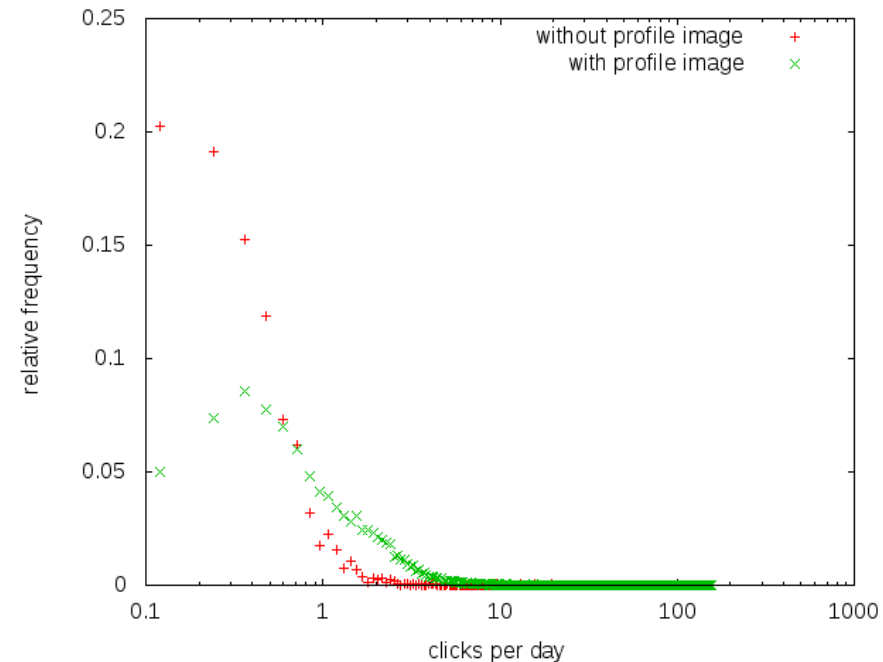
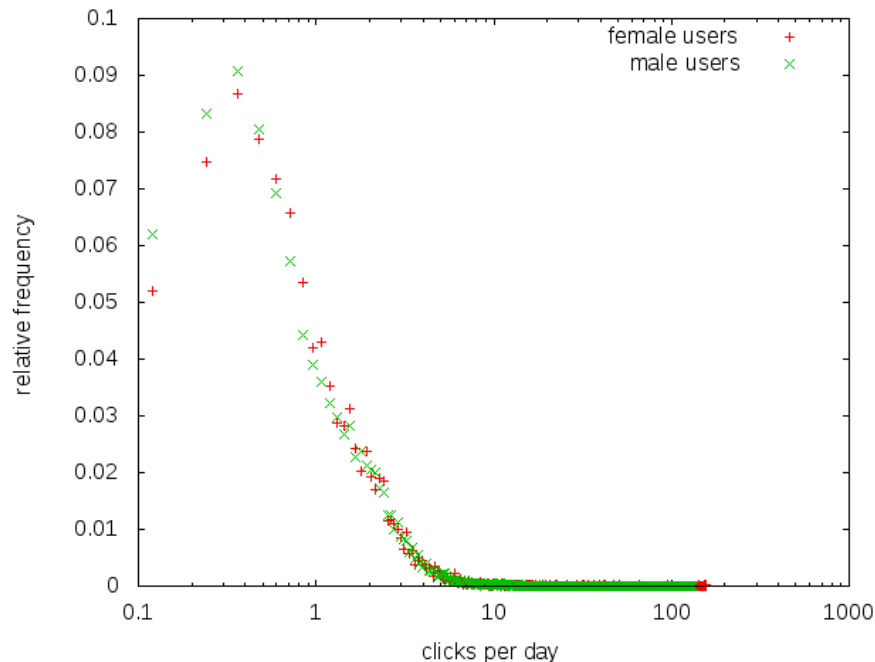
- Activity
 - “Activity meter” in profiles very coarse grained
 - Derived “profile alteration frequency” as alternative
 - Men are slightly more active than women (to both metrics)
 - Profiles without image belong to inactive users

- 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)



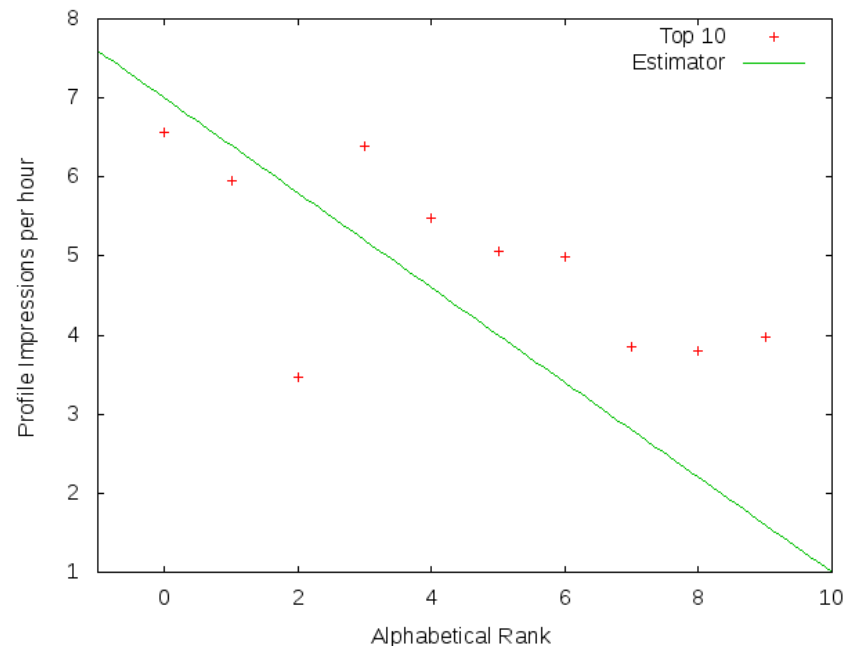
- *“The profiles of active users are more attractive”*
 - Activity as given on the profiles (**“activity meter”**)
 - $r \approx 0.17$, no noteworthy correlation
 - Activity measured in **group memberships**
 - $r \approx 0.37$ (higher for men, lower for women)
 - Activity meter is very coarse grained
 - 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
 - H_0 rejected, but $r \approx 0.11$, **no noteworthy correlation**
- *“Last name starting with a letter late in the alphabet sucks...”*
 - H_0 not rejected, there is no correlation.
 - Taking the “rich-club”, however...
 - Top 5% profiles: $r \approx -0.09$
 - Top 2‰ profiles: $r \approx -0.22$
 - Top 1 ‰ profiles: $r \approx -0.29$
 - Top 10 profiles: $r \approx -0.9 !!$



Summarizing the User Model



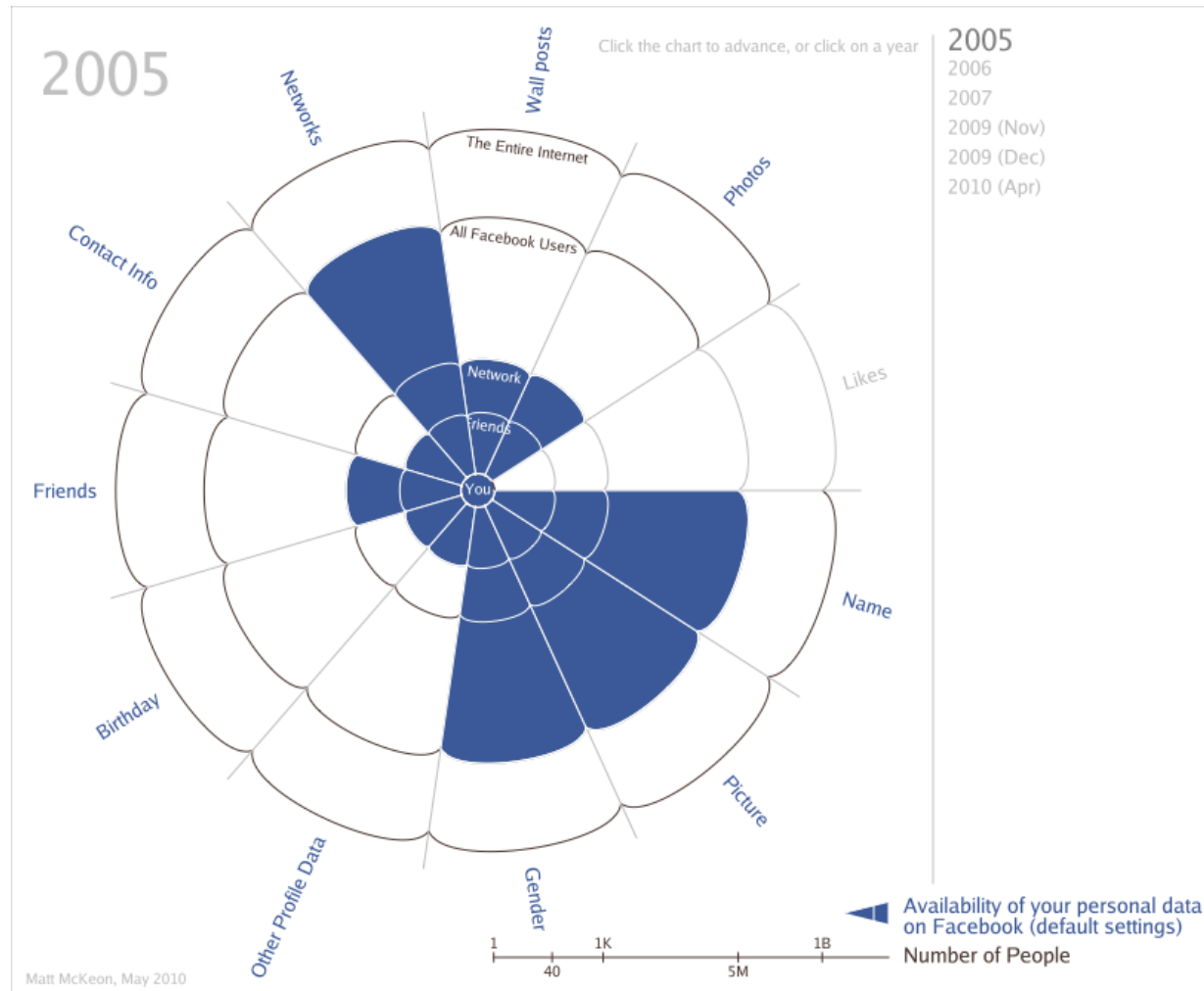
- Selected large sample of profiles in *business oriented* OSN
- Monitored profile properties and popularity (in pi/h)
- Profile Popularity *can* be predicted. Relates to
 - Providing image
 - Activity (diligence of maintaining profile)
 - Number of friends and contacts
- What we take away (P2P OSN)
 - 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 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

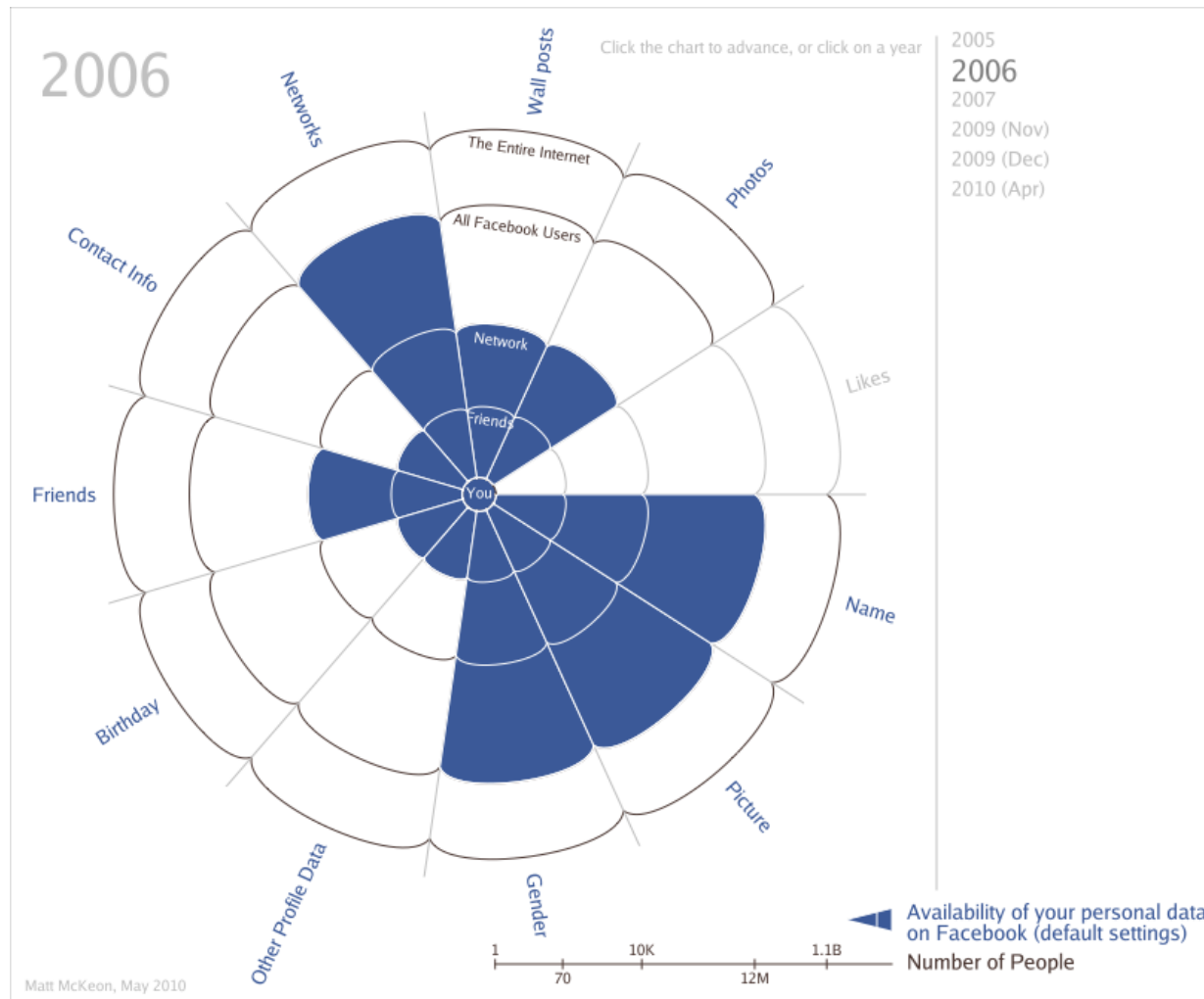
- Potential reasons for the dominance
 - Quick following of new hypes
 - “Apps” (Extensions to service, integration of 3rd party services)
 - “Friend feed”: construct virtual village (following twitter’s success)
 - Facebook places (following foursquare and gowalla’s success)
 - Stalking service (you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

Facebooks Privacy-Evolution: 2005

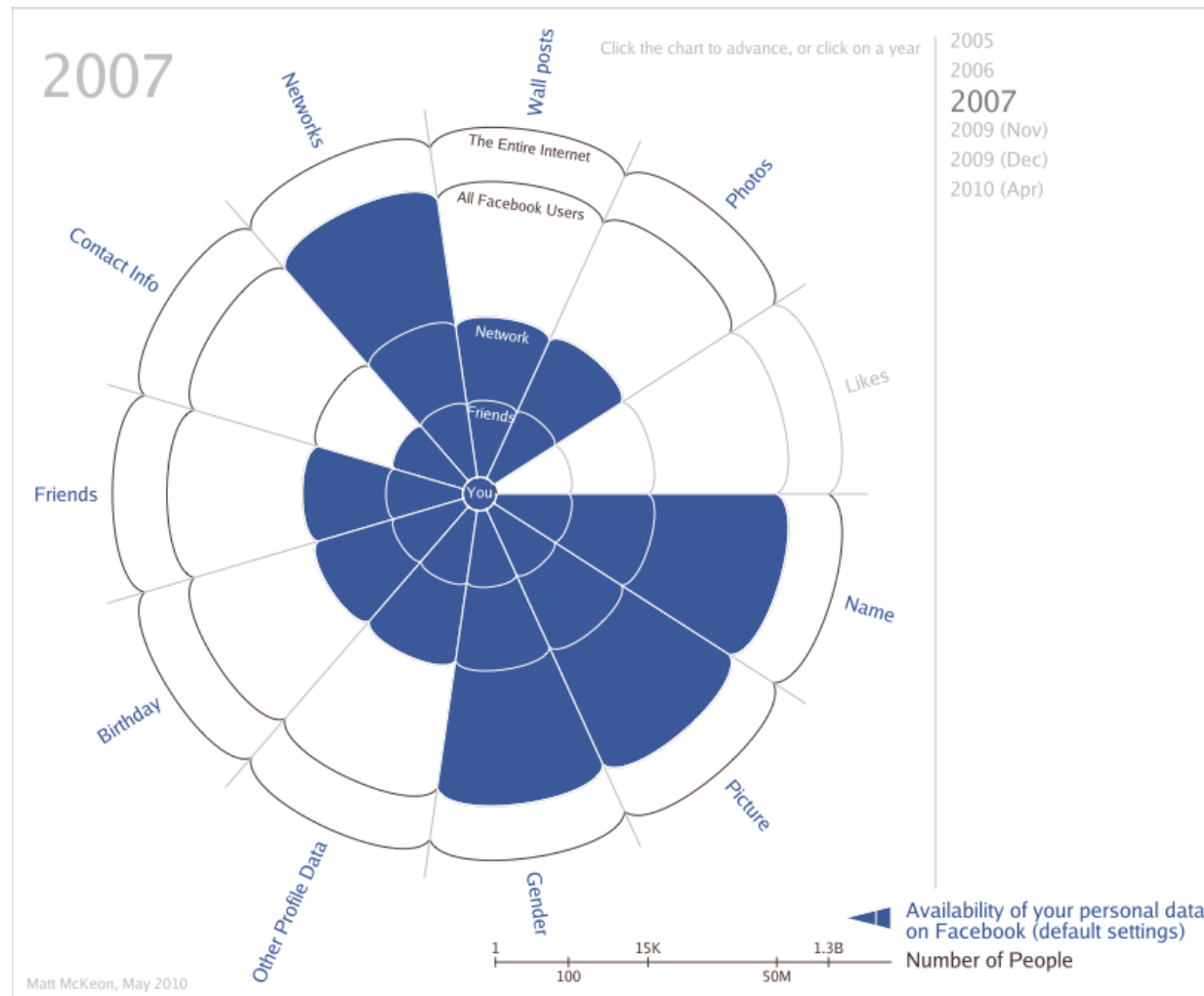


Source: McKeon

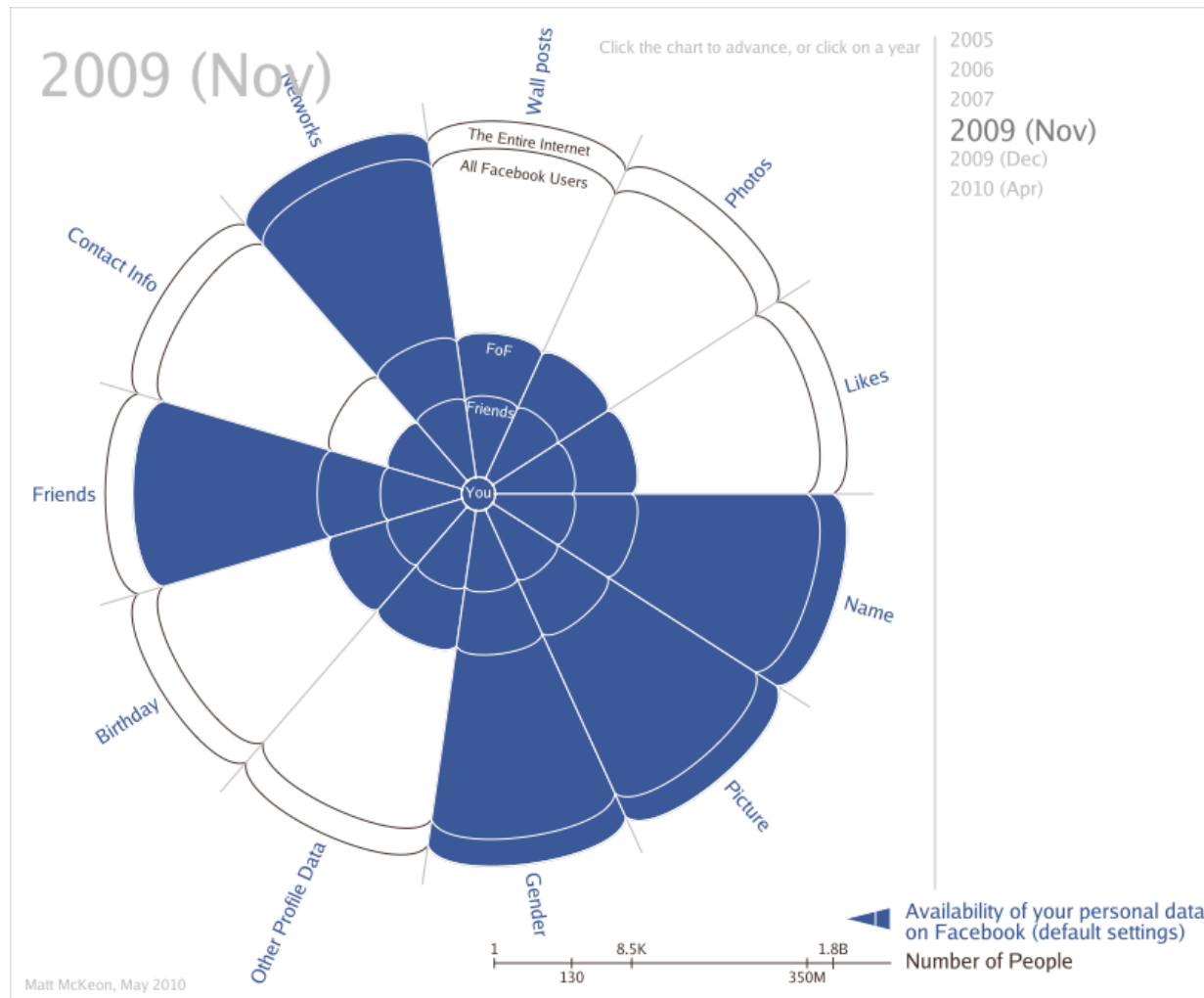
Facebooks Privacy-Evolution: 2006



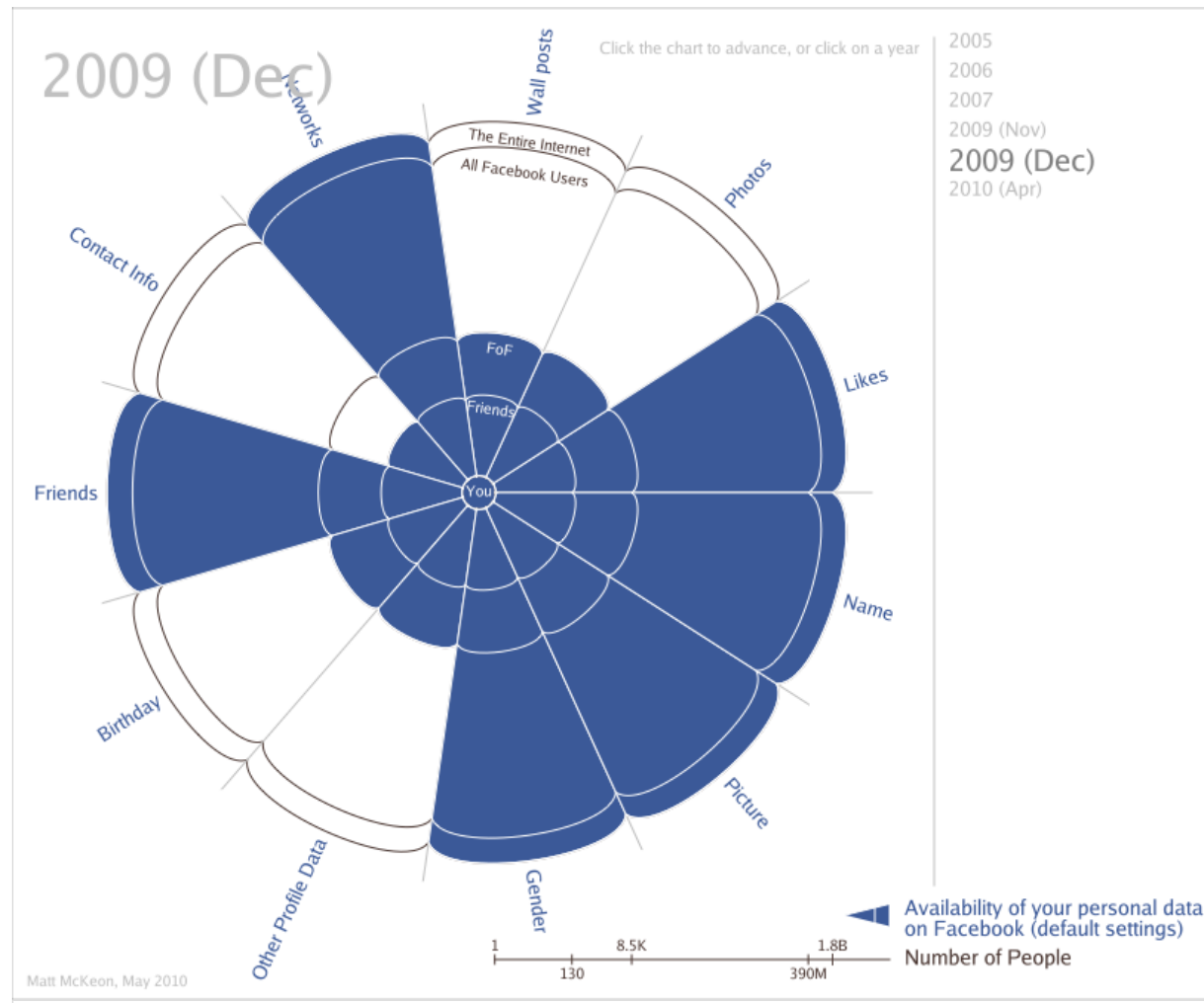
Facebooks Privacy-Evolution: 2007



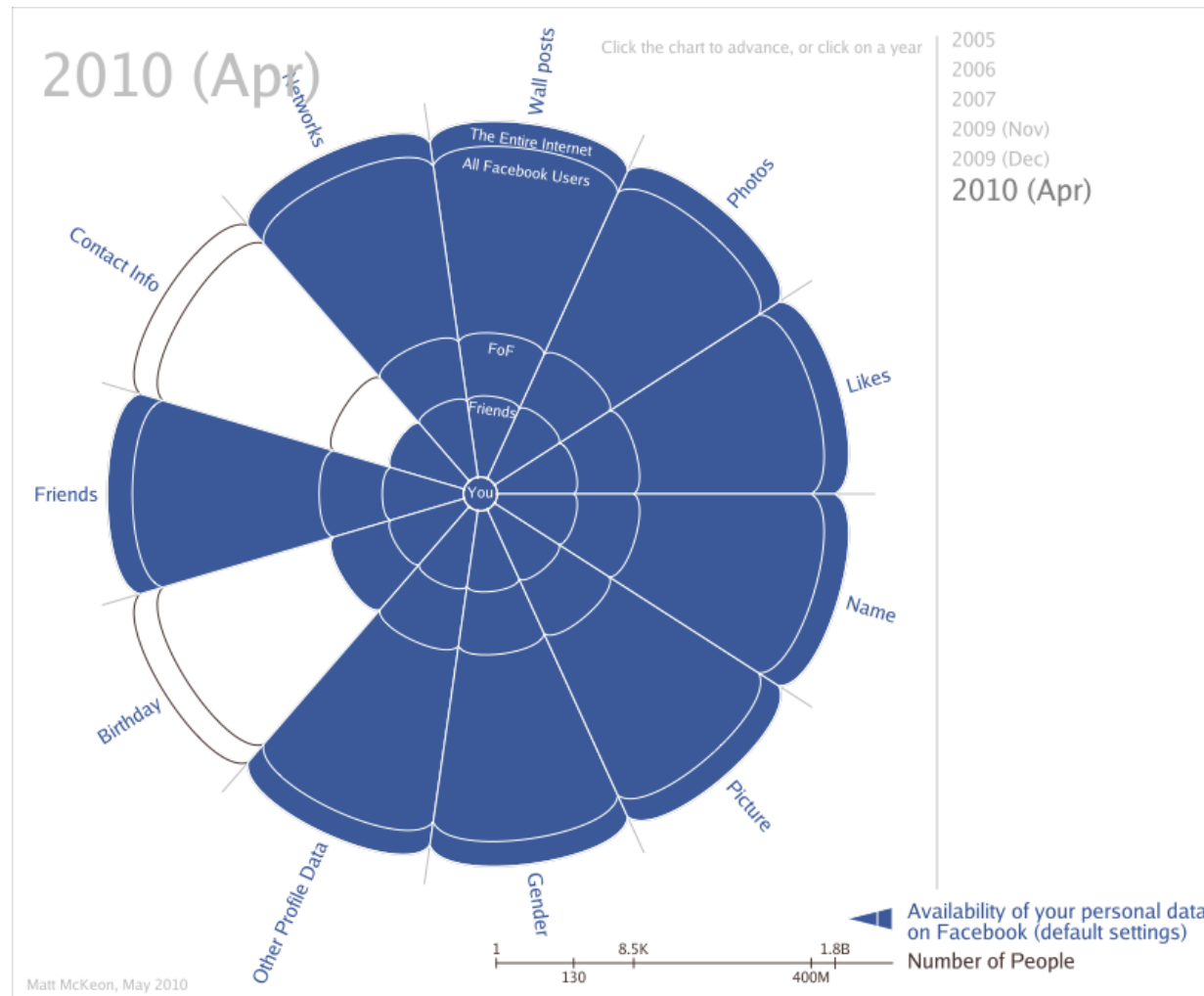
Facebooks Privacy-Evolution: 2009



Facebooks Privacy-Evolution: 2009



Facebooks Privacy-Evolution: 2010





- 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



2009 MADRID

April 20th-24th, 2009
18th Int. World Wide Web Conference

WELCOME ATTENDING PRESS ROOM STORE REGISTRATION

BECOME A SPONSOR

Track: Security and Privacy / Session: Web Security

All Your Contacts Are Belong to Us: Automated Identity Theft Attacks on Social Networks

Leyla Bilge, Thorsten Strufe, Davide Balzarotti, Engin Kirda
EURECOM
Sophia Antipolis, France
bilge@eurecom.fr, strufe@eurecom.fr, balzarotti@eurecom.fr, kirda@eurecom.fr

ABSTRACT

Social networking sites have been increasingly popular. Well-known sites such as MySpace and Facebook have been reporting growth rates of 10% per week [5]. Many social networking sites allow users to share photographs and videos with their friends and to keep in touch. In this paper, we investigate how easy it would be for a potential attacker to launch automated crawling and identity theft attacks against a number of popular social networking sites in order to gain access to a large volume of personal user information. The first attack we present is the auto-

1. INTRODUCTION

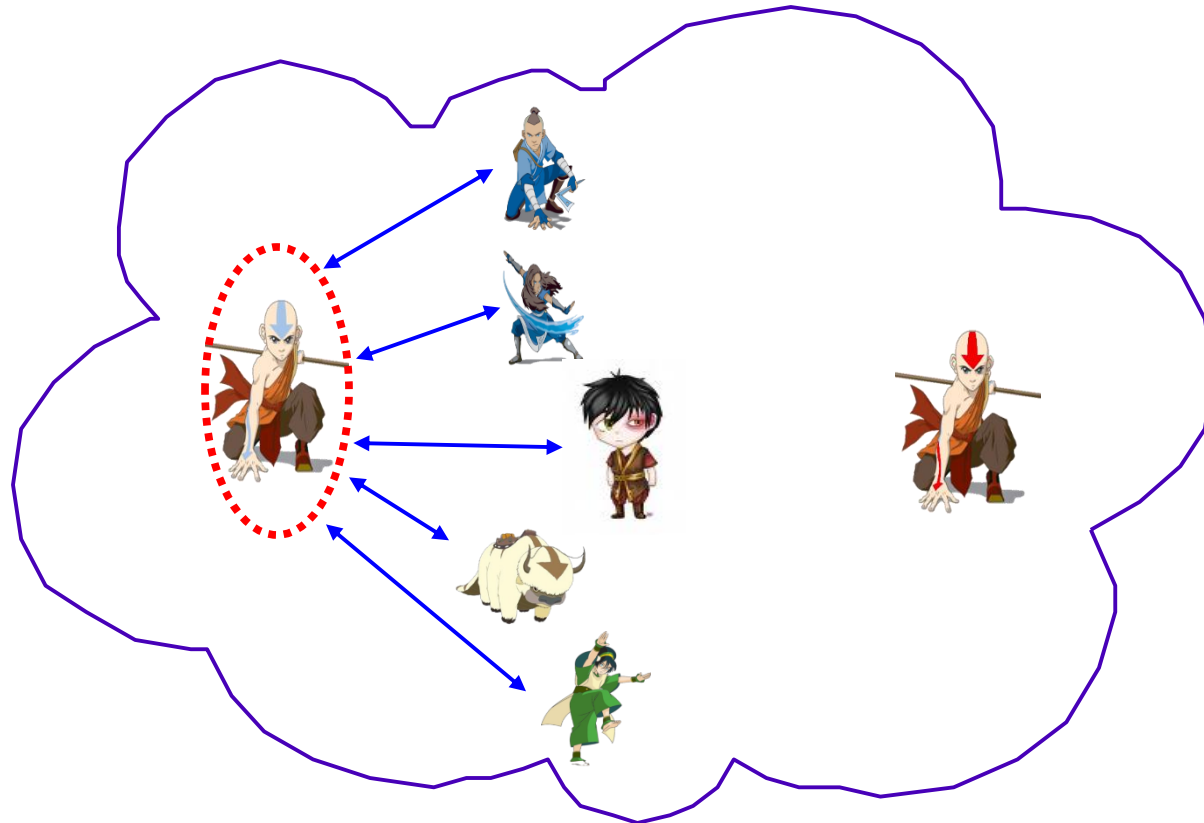
A social network is a social structure that is made up of nodes representing individuals or organizations. These nodes may be tied to each other by properties such as friendship, common values, visions, ideas, business relationships and general interests. Although the idea of social networks has been around for a long time (e.g., see [14]), social networking web sites and services are a relatively new phenomenon on the Internet. Business relationship-focused social networking sites such as XING [13] (previously known as OpenBC) and LinkedIn [6], as well as friendship-focused social networking sites such as Facebook [4], MySpace [8],

- 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

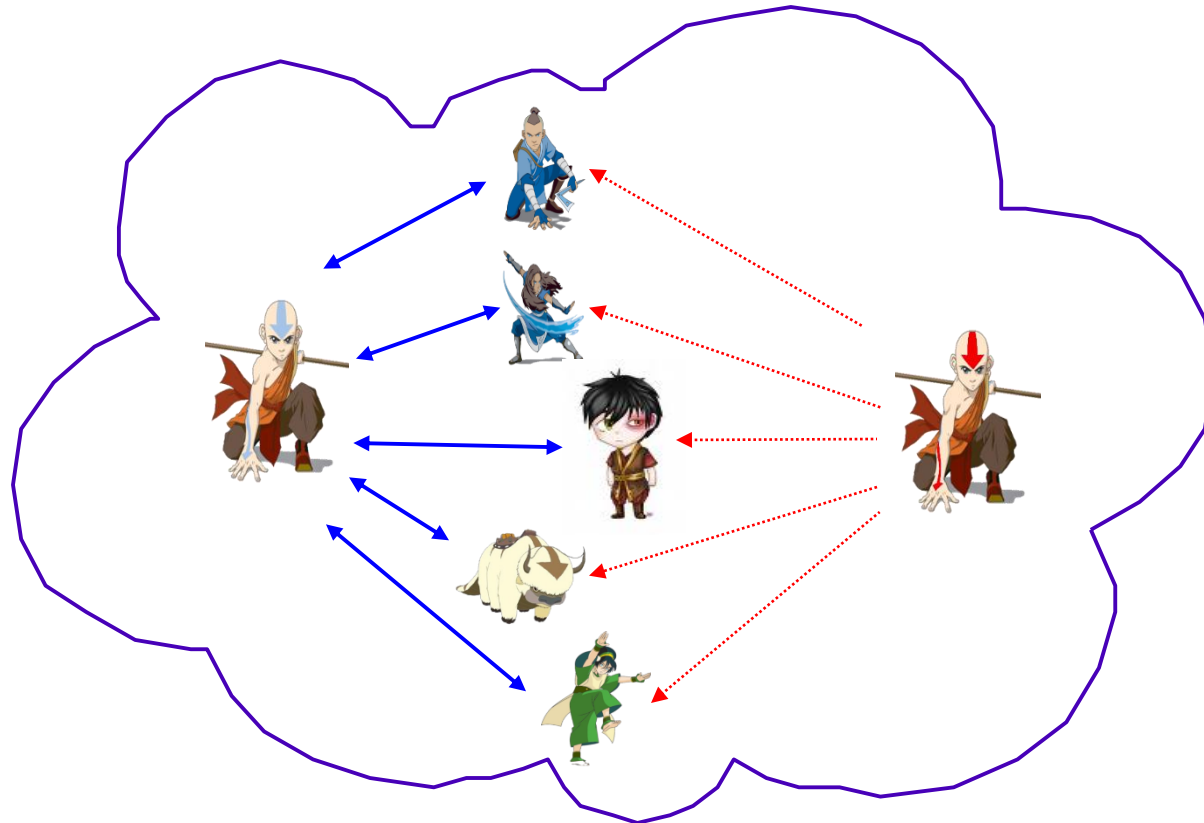


- Is it possible for an attacker to launch impersonation attacks on a large scale against a number of popular social networking sites?
 - Facebook (international)
 - XING (international)
 - LinkedIn (international)
 - MeinVZ (popular in Germany, Austria, Switzerland)
 - StudiVZ (popular in Germany, Austria,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

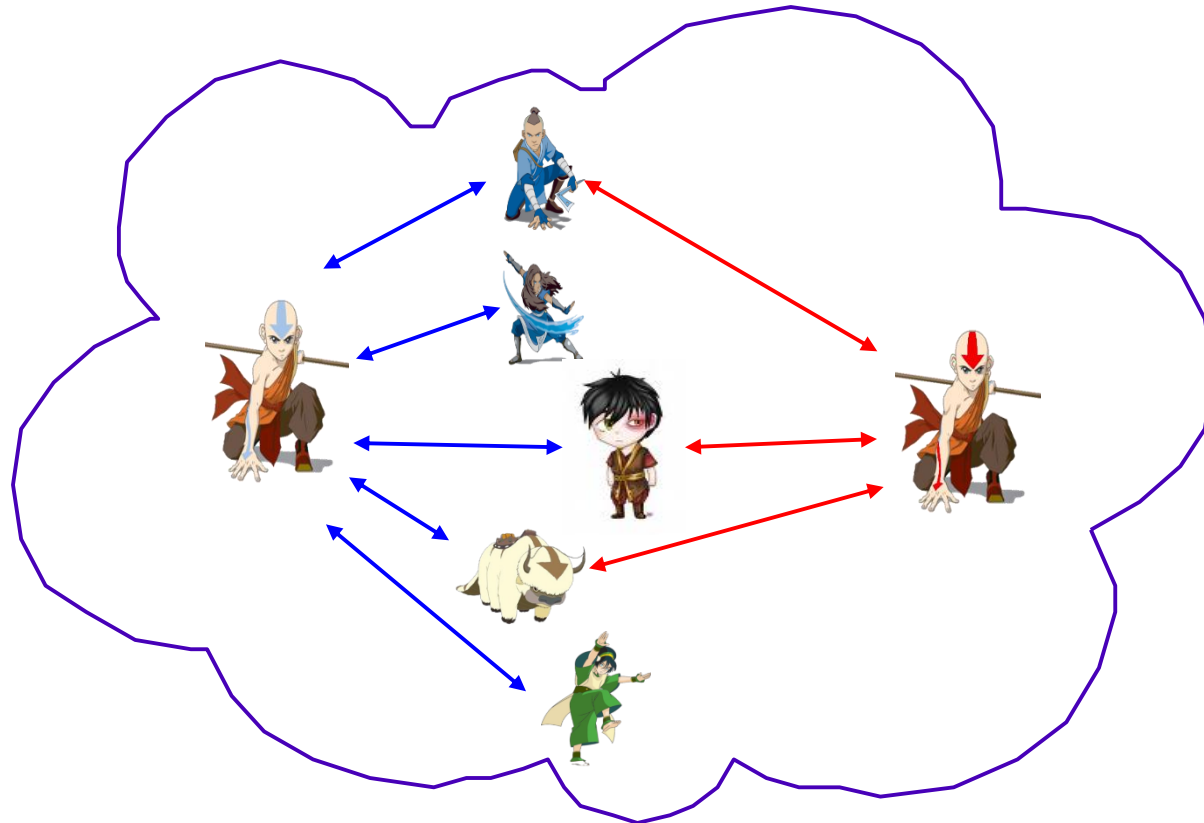
Automated Profile Cloning



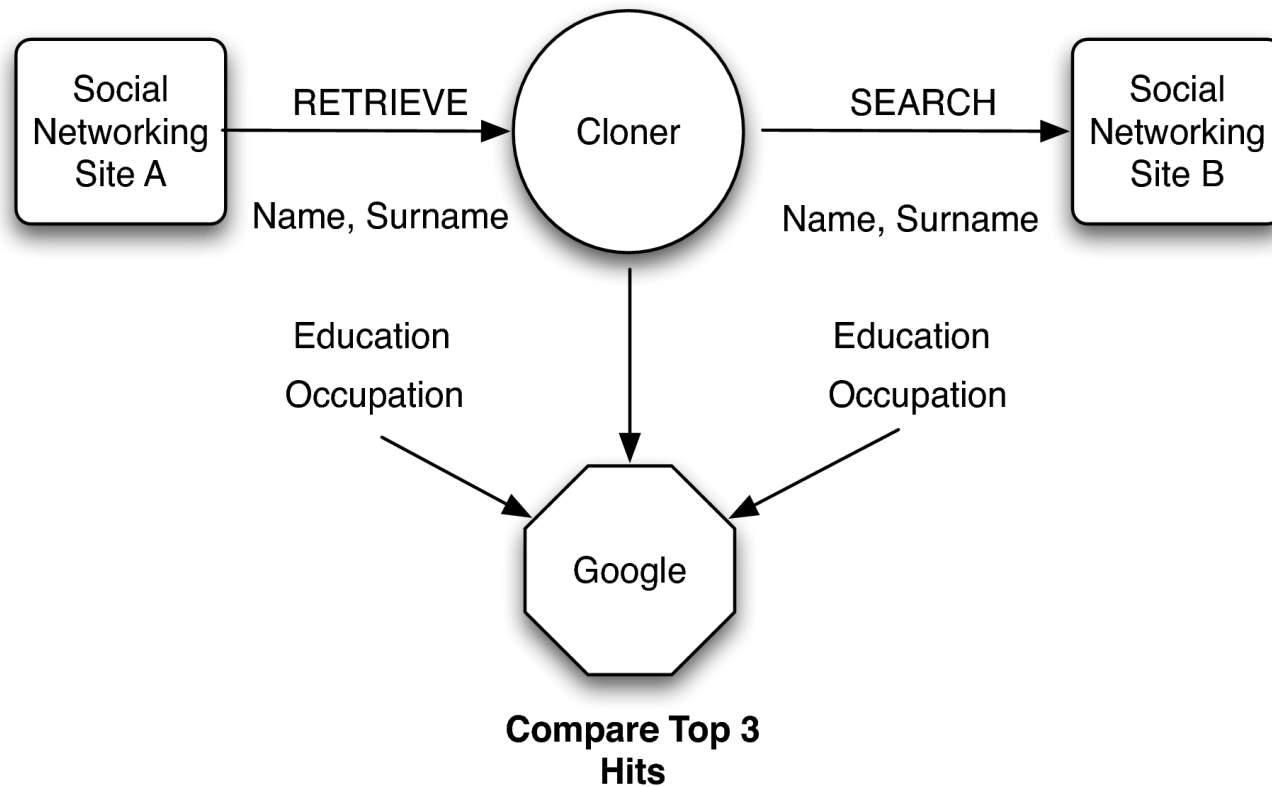
Automated Profile Cloning



Automated Profile Cloning



Cross-Site Profile Cloning

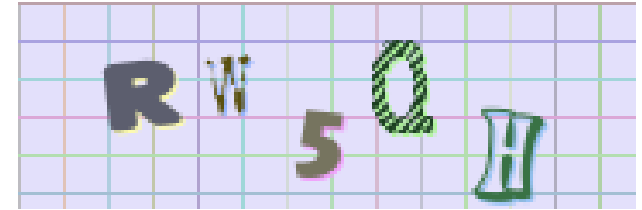




- CAPTCHA: Completely Automated Public Turing test to tell Computers and Humans Apart
- CAPTCHAs are employed to prevent automated programs from accessing and abusing the services
- In order to automate the attacks, a number of CAPTCHA breaking techniques were developed
 - “Quick and dirty”, techniques are not perfect
 - The aim is to break the CAPTCHAs efficiently enough to make automated attacks against several social networking sites possible



- 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%



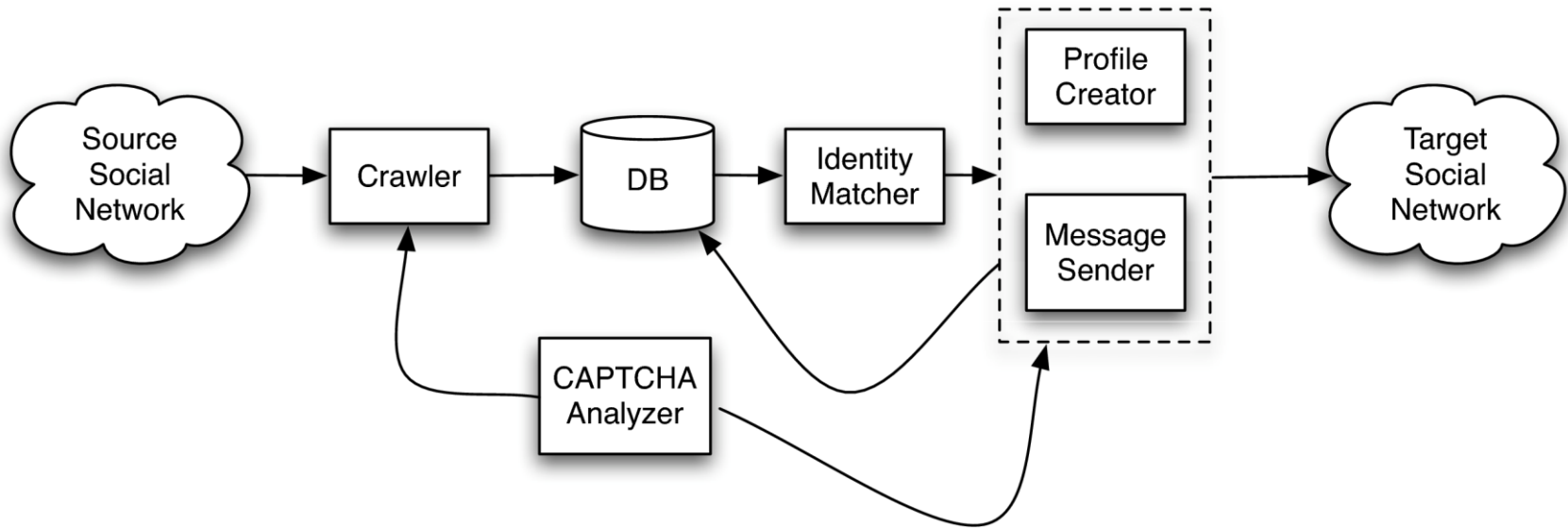
**Cracking the CAPTCHAs was done with serious amounts of help from Michael Roßberg/TU-Ilmenau*



- 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





- 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?



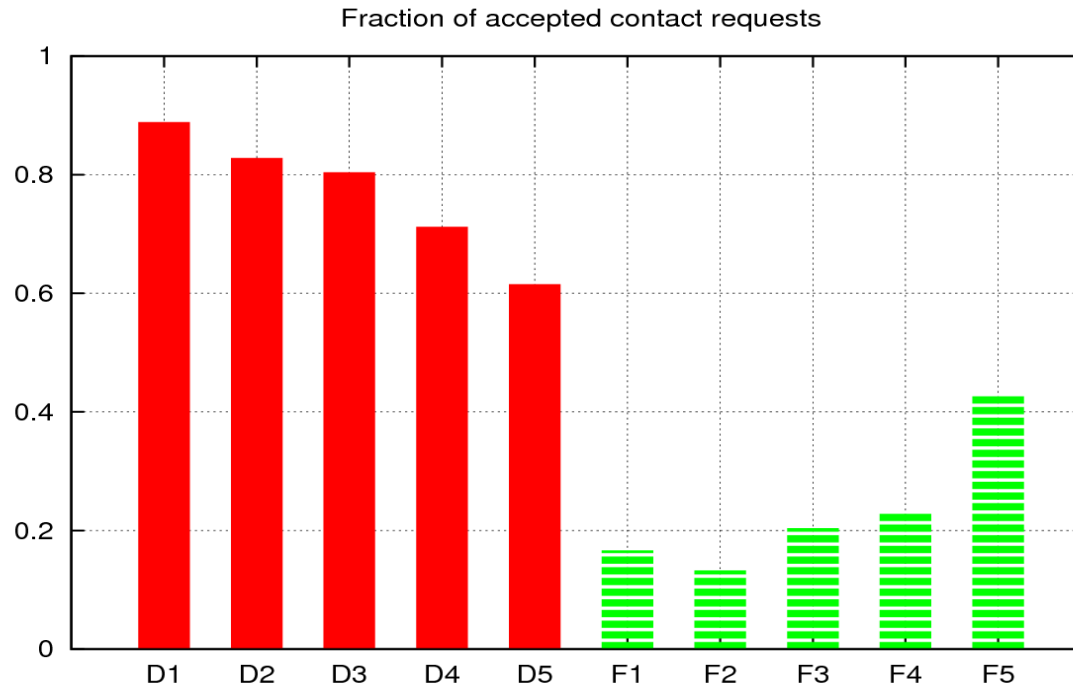
- StudiVZ and MeinVZ
 - Displays CAPTCHA if large number of requests come from one account
 - To collect as much information as possible, without being noticed, 16 accounts were created, and separately used for crawling
 - Collected 5M profiles with contact information, and 1.2M complete user profiles

- 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)





- 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!:

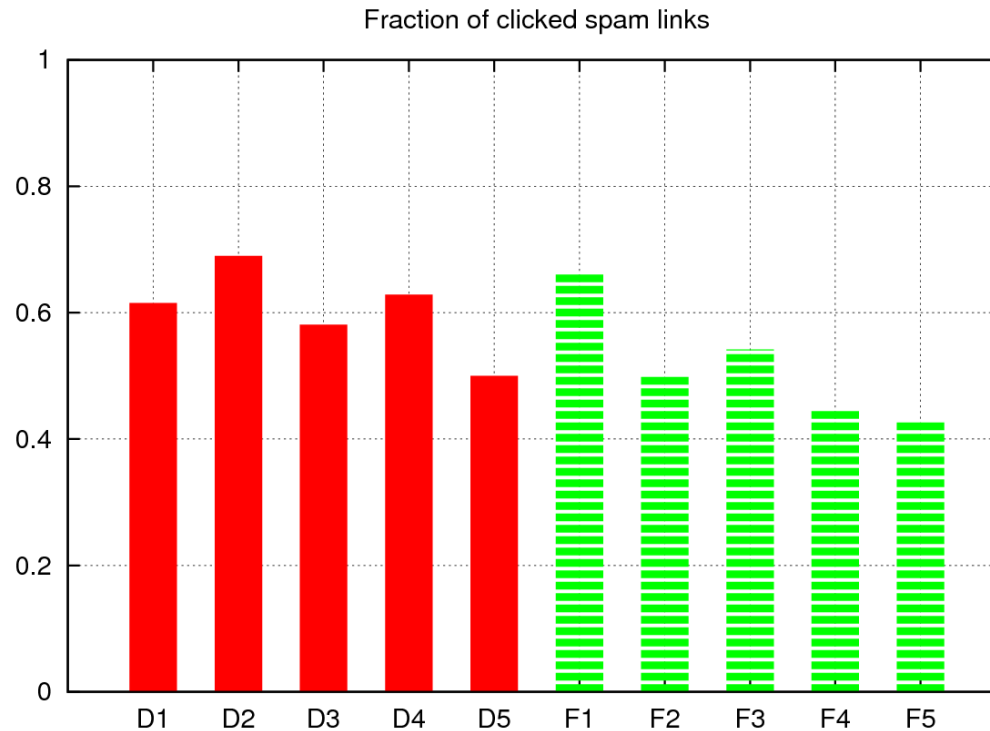
`http://193.55.112.123/userspace/pix?user=<account>
&guest=<contact>&cred=3252kj5kj25kjk325hk}`

Ciao, <account first-name>

Profile Cloning Experiments



- Click through rate for messages from duplicate / fictitious profiles



Cross-Site Profile Cloning Experiments



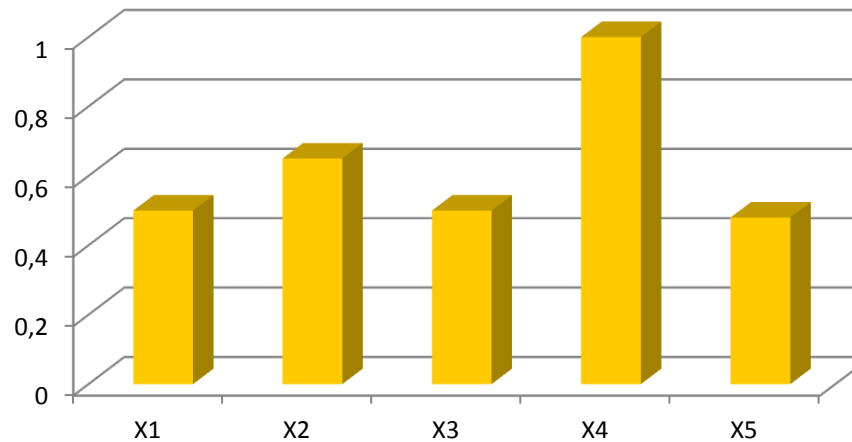
- Cloning profiles that exist on XING, but not on LinkedIn
- The success of the cross-site profile cloning depends on the number of users that have a profile in both of the networks
- From around 30.000 crawled profiles in XING, 3.700 were also registered in LinkedIn

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:

Fraction of accepted 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

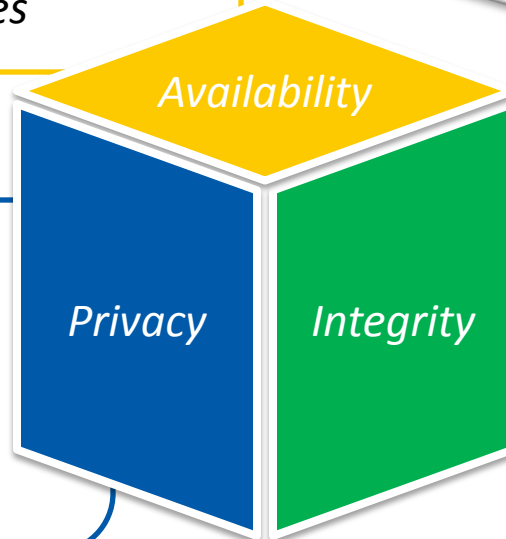
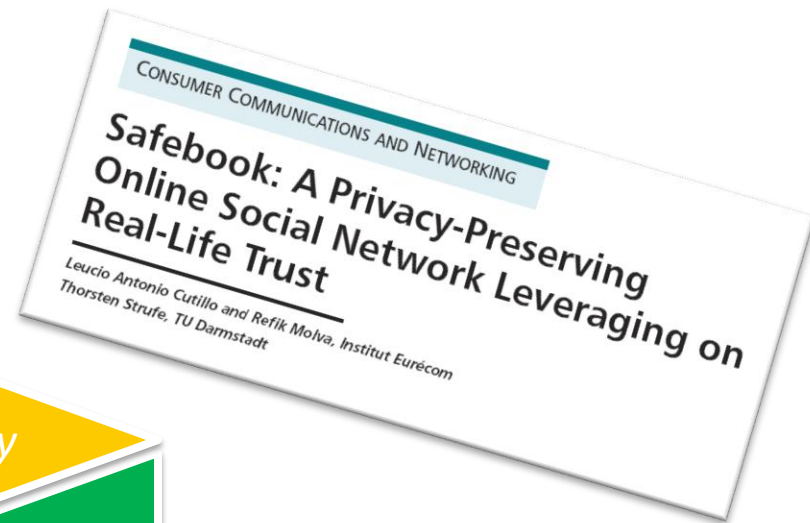
- Apply anomaly detection techniques to detect:
 - Crawling
 - Massive amount of friend requests or messages
 - Profiles with identical/similar information
 - Attempts to automatically solve CAPTCHAs

- ...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*



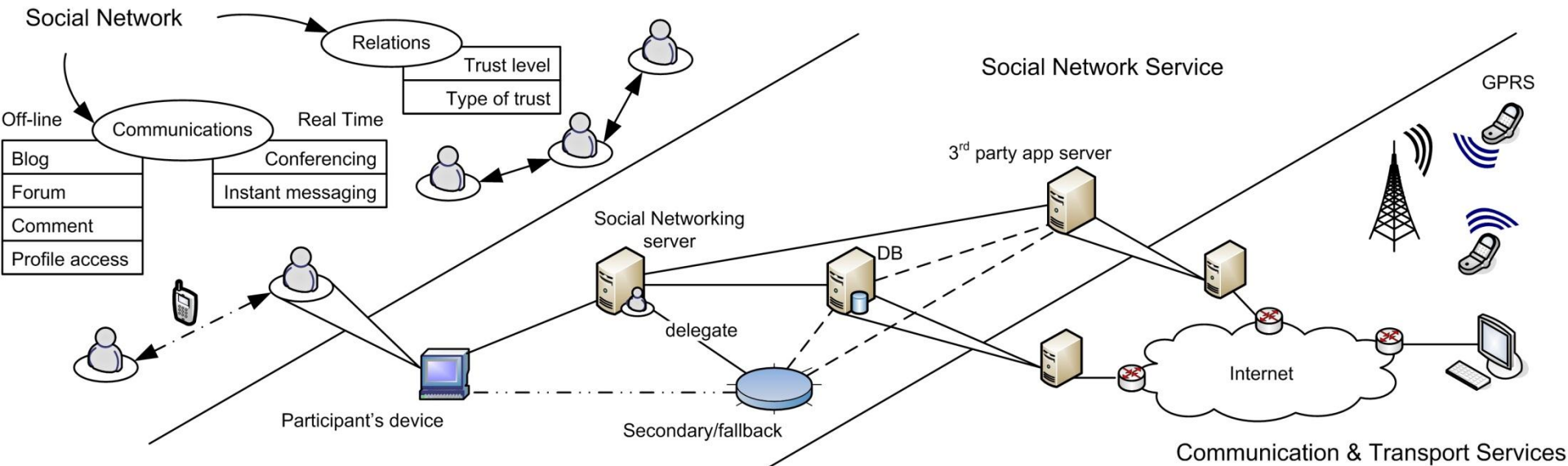
1. *Only trusted parties can access:*
 - *Personal Information*
 - *Communication*
2. *Possibility to hide any data about any user*

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

Generalized Architectural Model of SNS

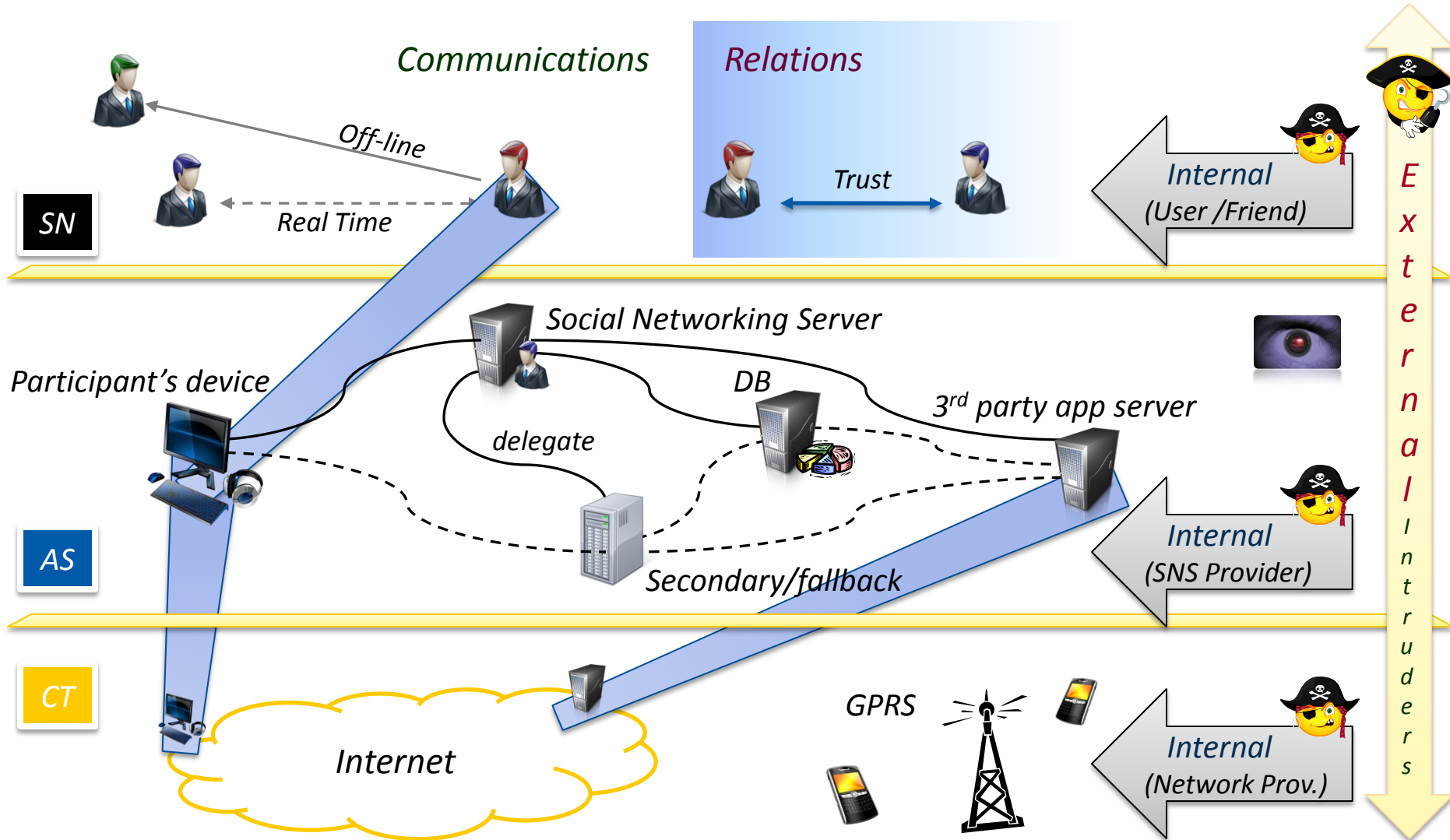


- Distributed Environment

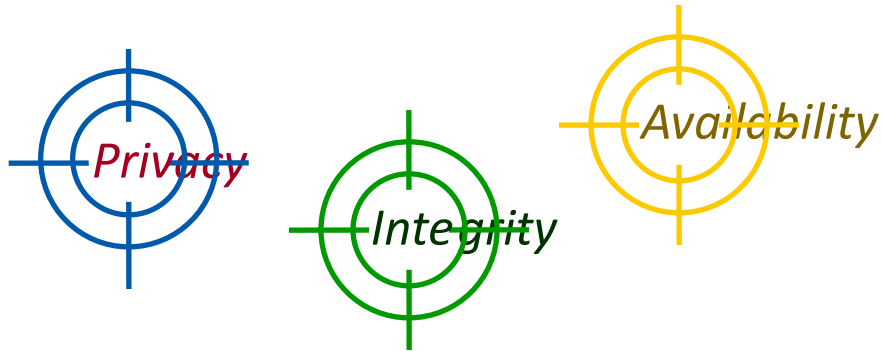


- Main SNS centralized
- Additional 3rd party application servers

Security Analysis of OSNs



Current Status of OSNs



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

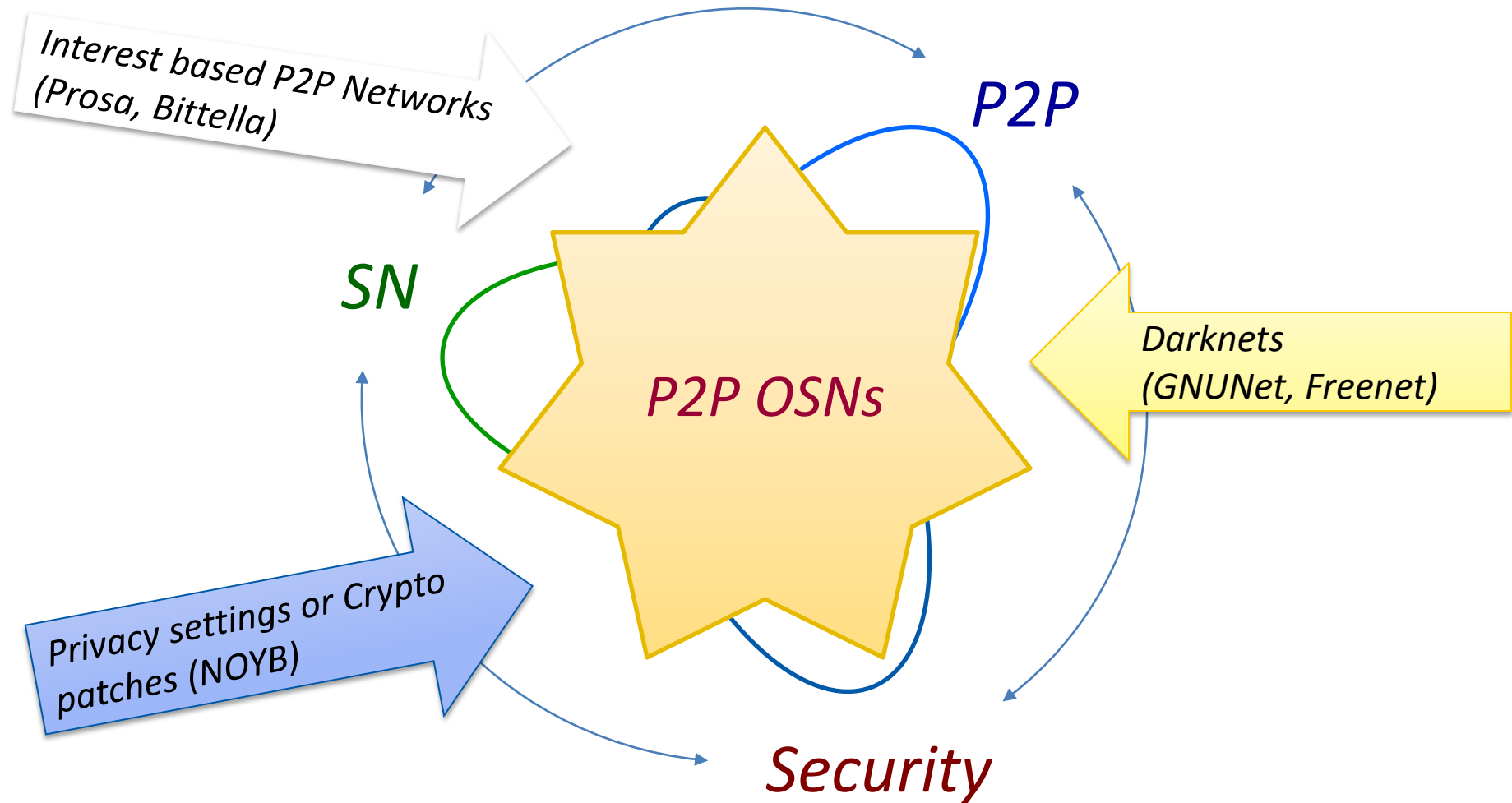
Attacks	P	I	A
Id Theft			
Profile Cloning			
Profile Porting			
Fake profiles			
Profiling			
Communication tracking			
Secondary Data collection			

Attaks	P	I	A
Image retrieval			
Fac... on			
...			
Defamation			
Censorship			
Collusion			

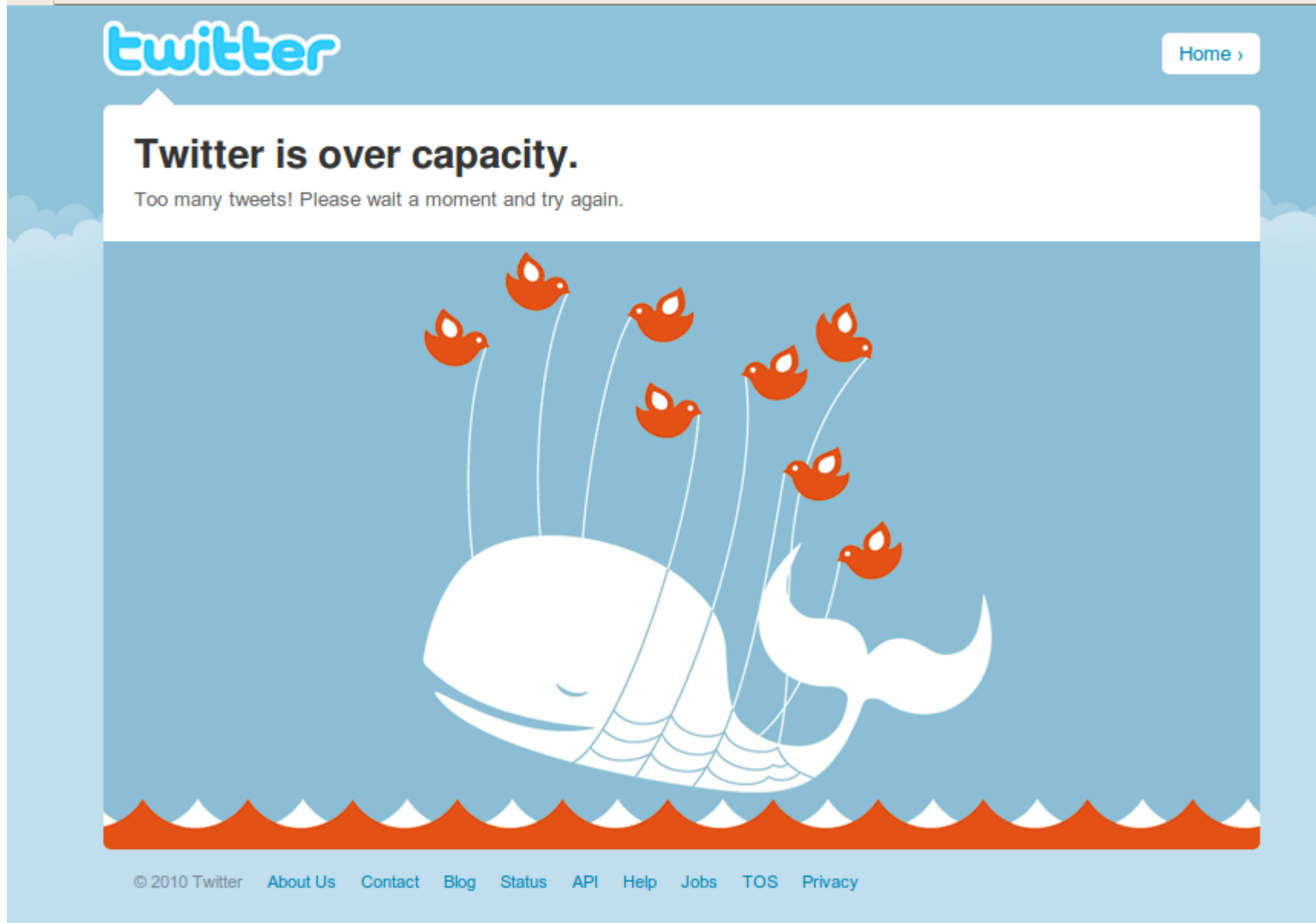
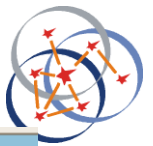


**SNS provider as a
“BIG BROTHER”**

Current solutions



Quest to Decentralize





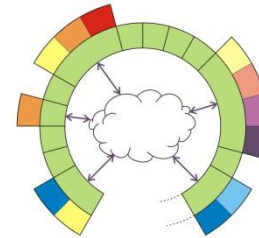
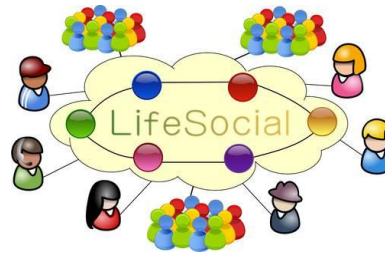
- FOSS
 - BuddyPress, CrabGrass, Cobs, DaisyChain, Diki, Elgg, FETHR, GNUNet, Gossple, Jappix, Lorea, Mycella, Movim, PeerScape, Pinax, StatusNet
- Commercial Approaches
 - diaspora, wuala, LifeSocial
- Academia
 - Friend-of-a-Friend, FriendStore, HelloWorld, LifeSocial, LotusNet (Likir), PeerSon, Safebook, 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





- 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



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



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

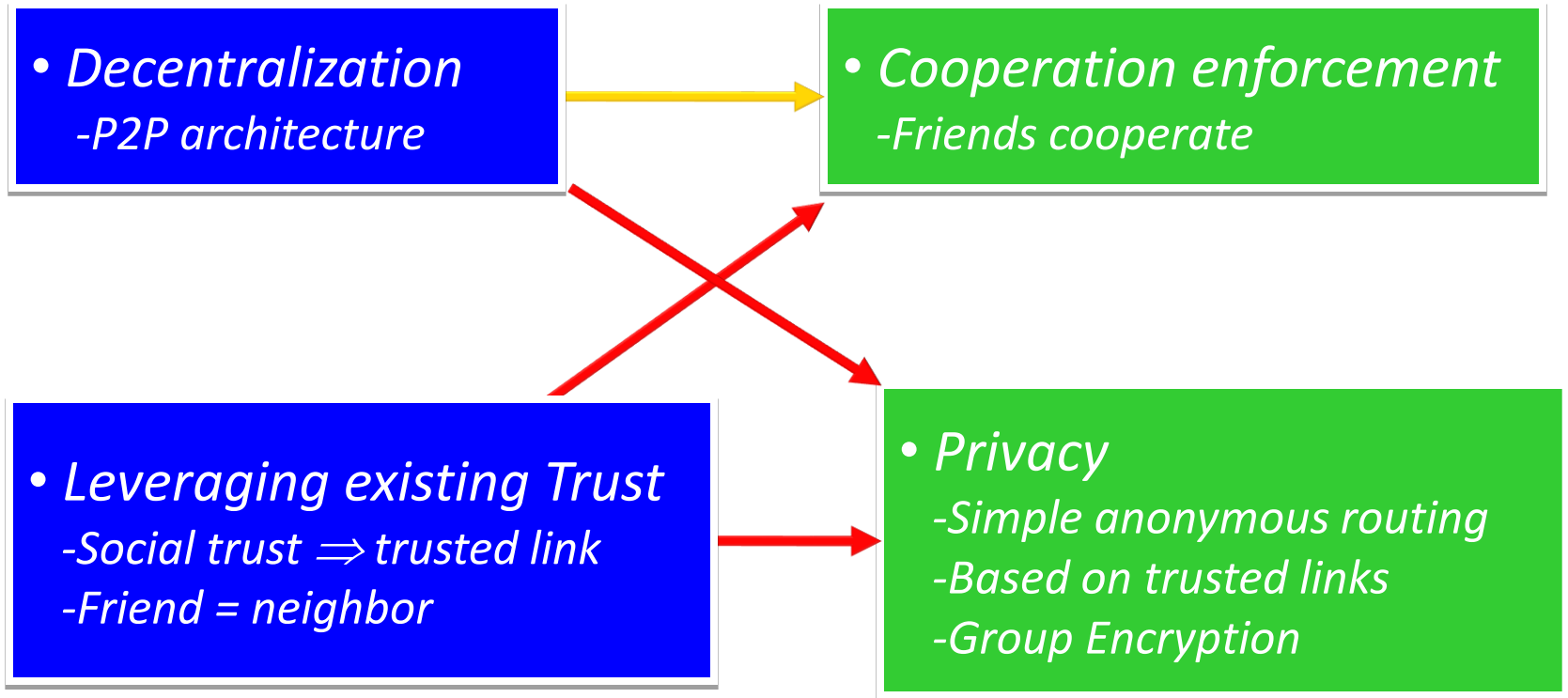
- Performance
 - Properties: P2P → filesharing! OSN: small, short lived objects
 - Extreme heterogeneity (friends, uploaded content, resources)
 - User behaviour / sessions (diurnal patterns, session- and inter session times)
 - Incentives
 - Mobile users vs. p2p / recursive routing, etc...

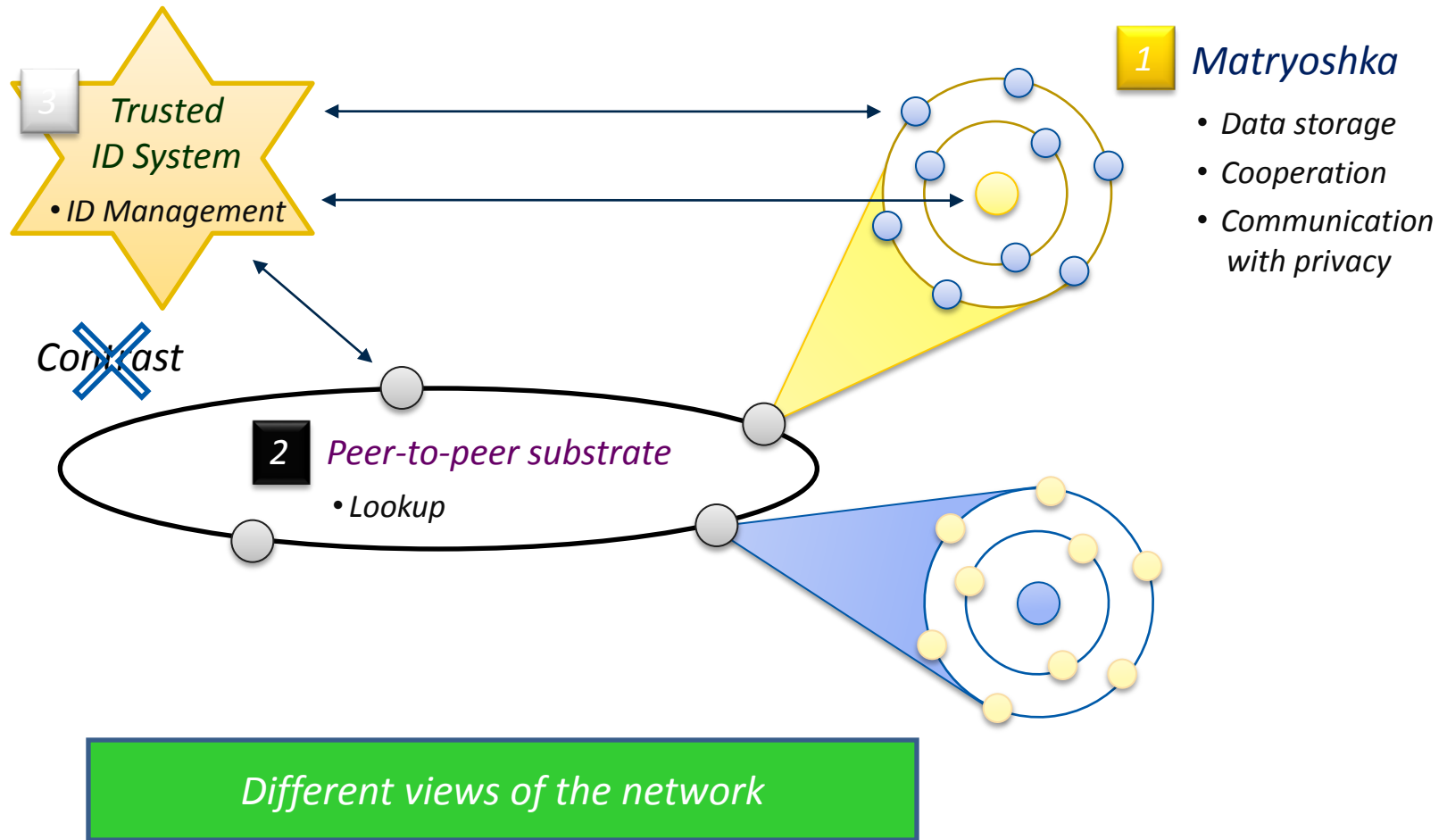
- Security
 - How can we „guarantee“ availability?
 - How do you search for somebody?
 - Key management!?
 - Identification services (we don't want fake accounts!)



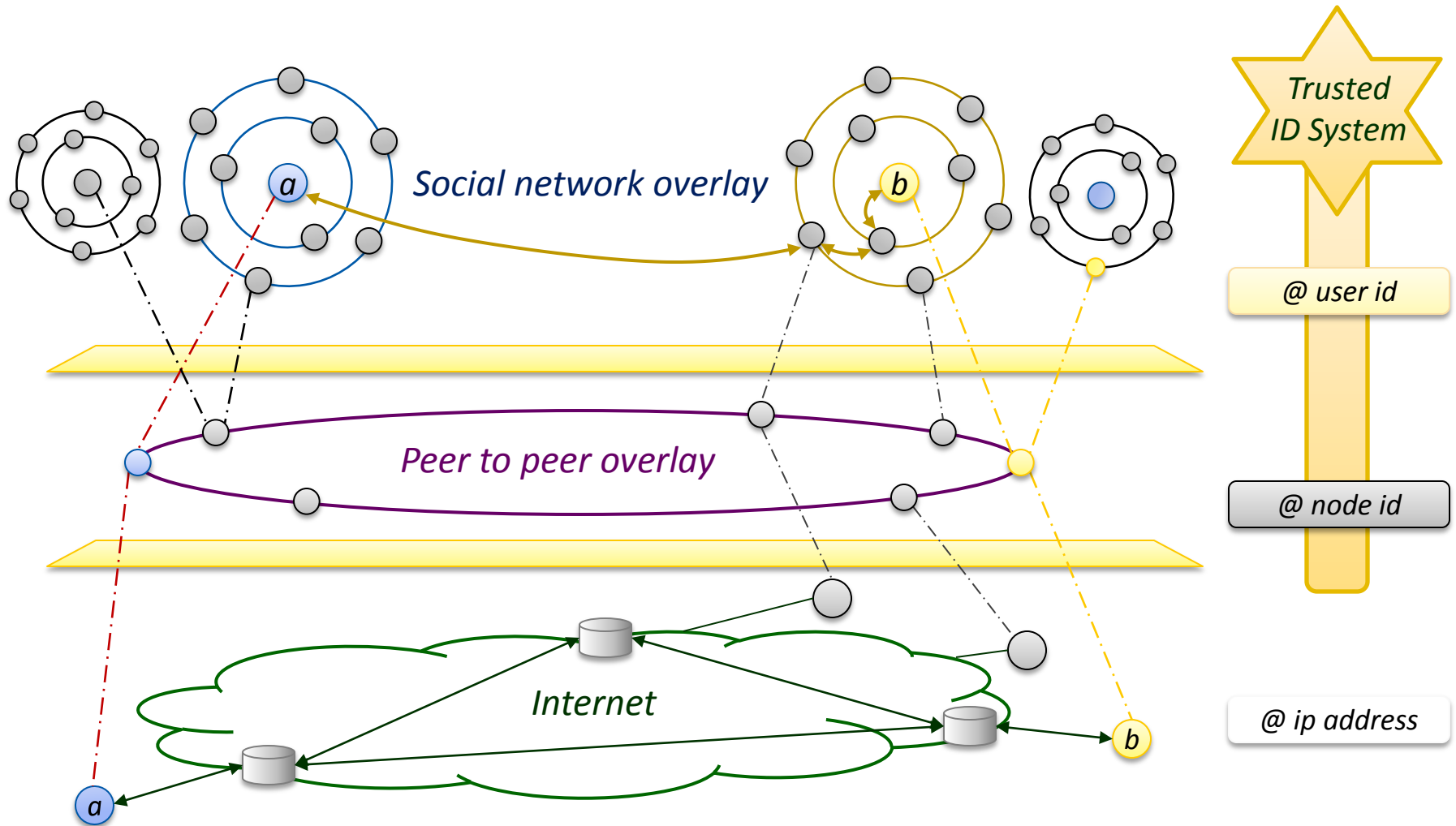
Safebook : Privacy-Preserving Online Social Networking

<http://www.safebook.us>

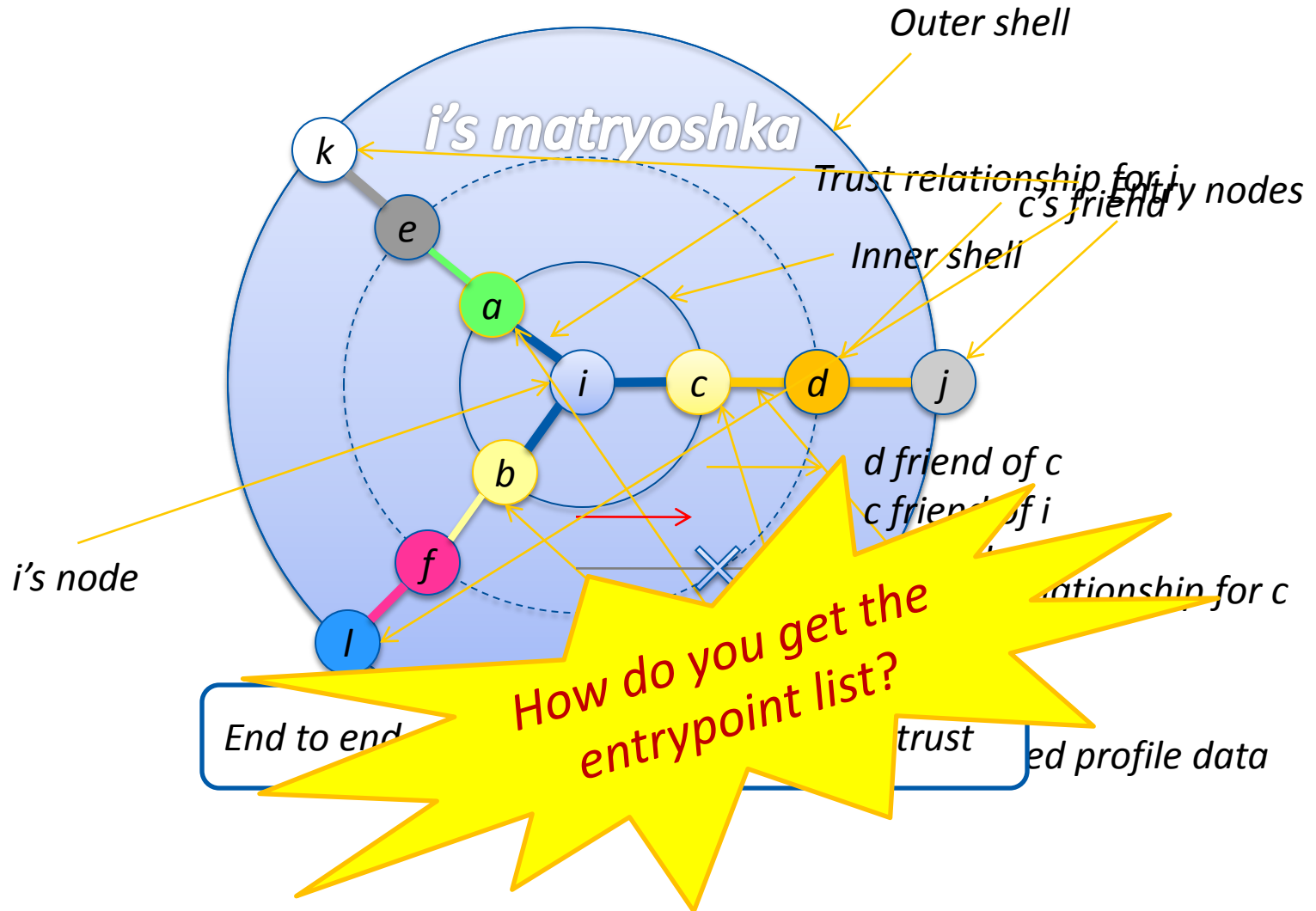




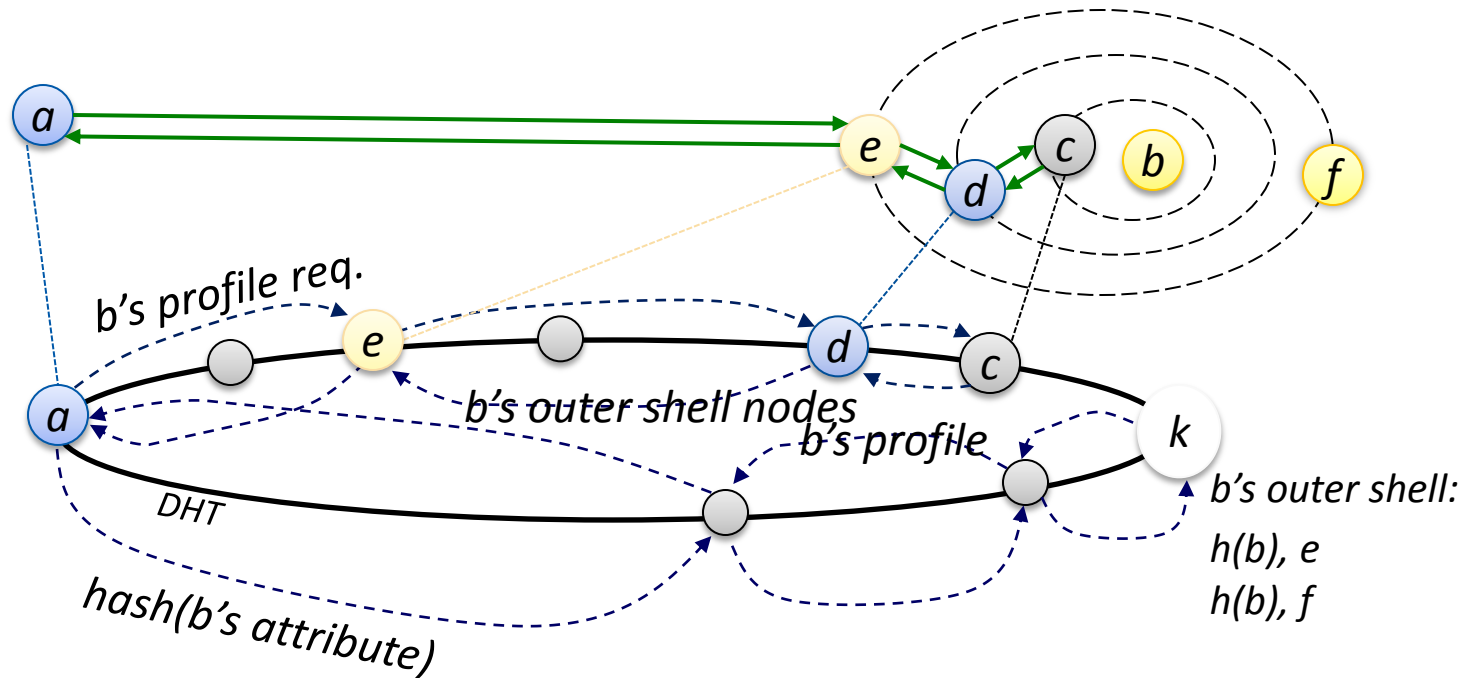
Network view



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

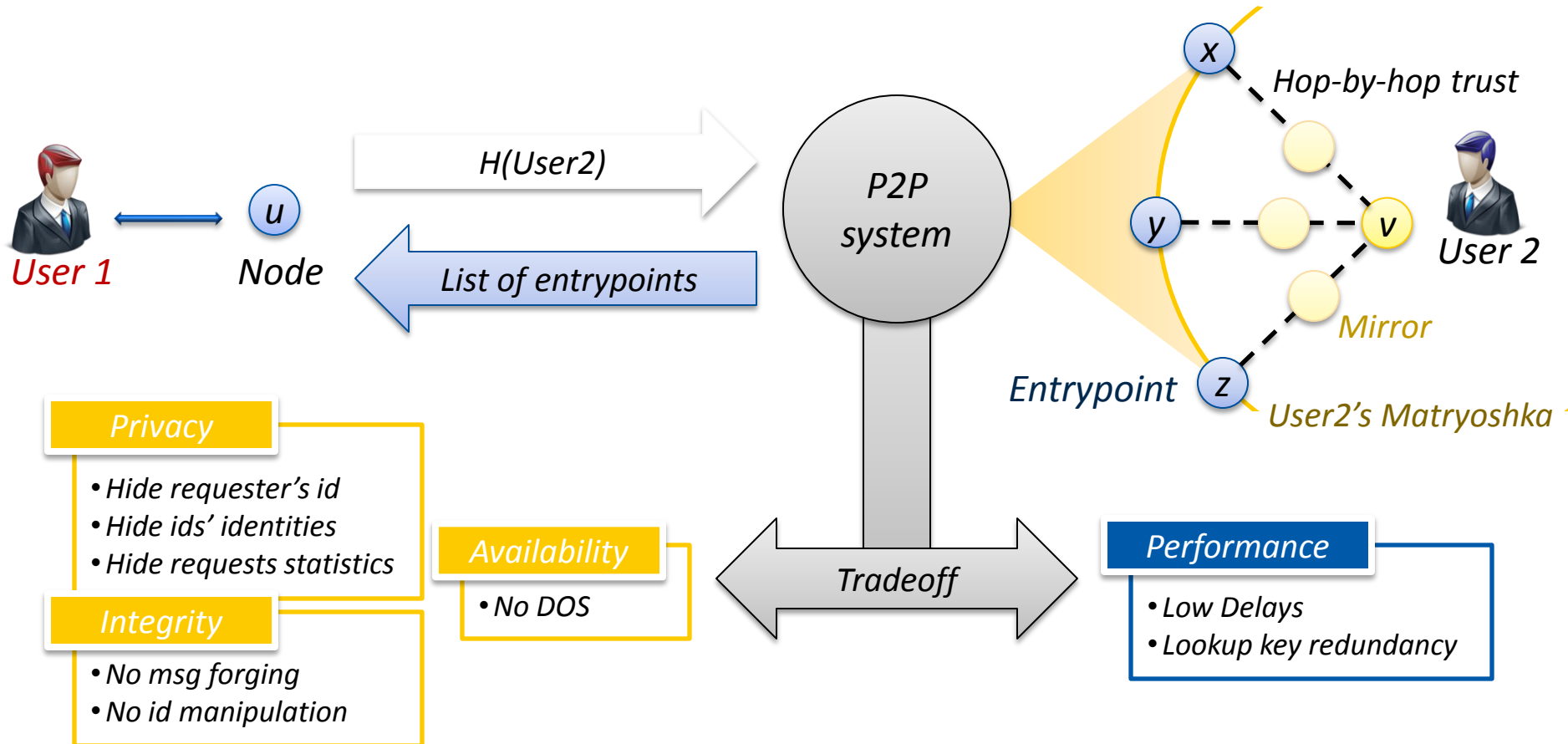
data request

- *a* sends profile data request to *a*'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



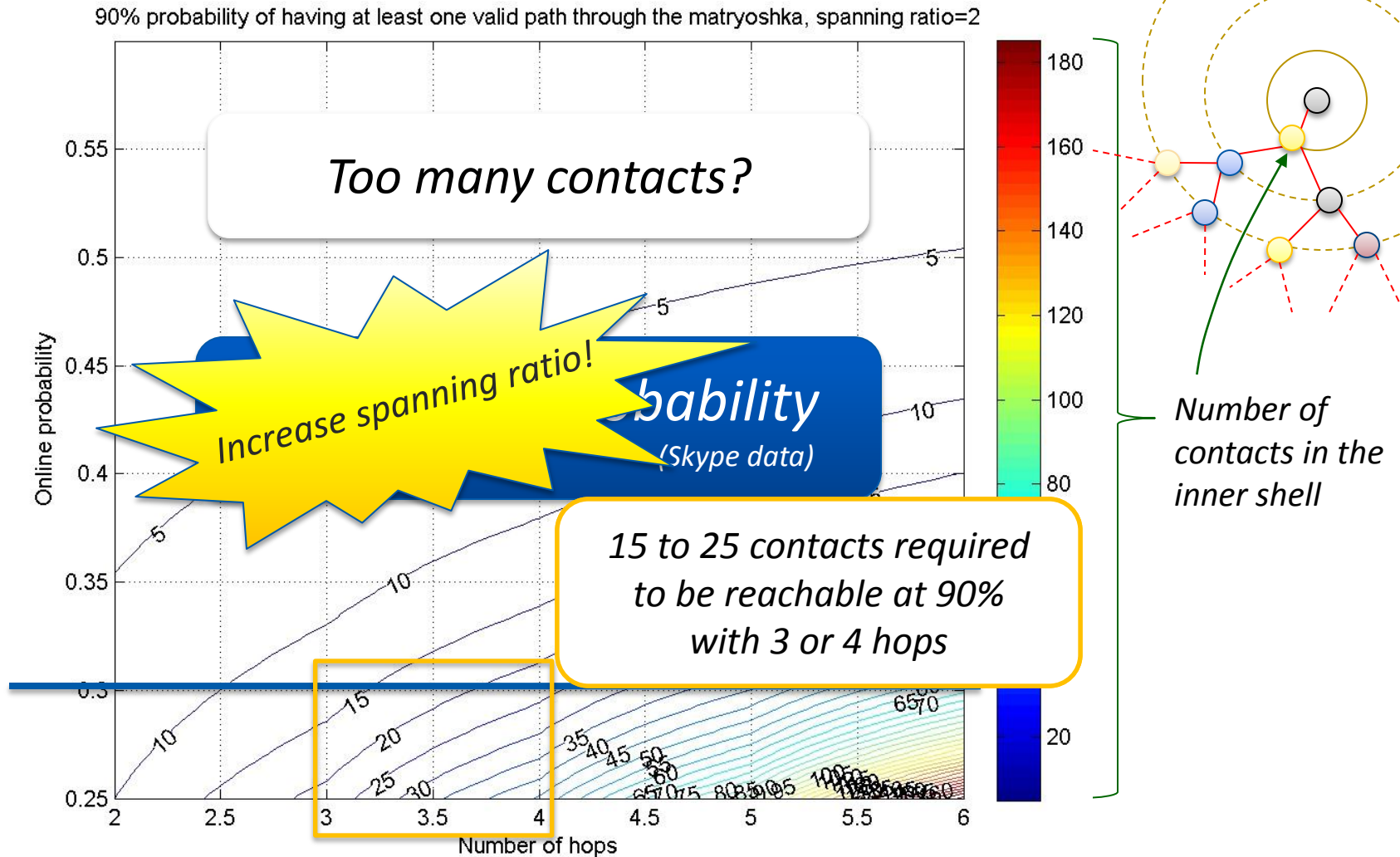
Key management

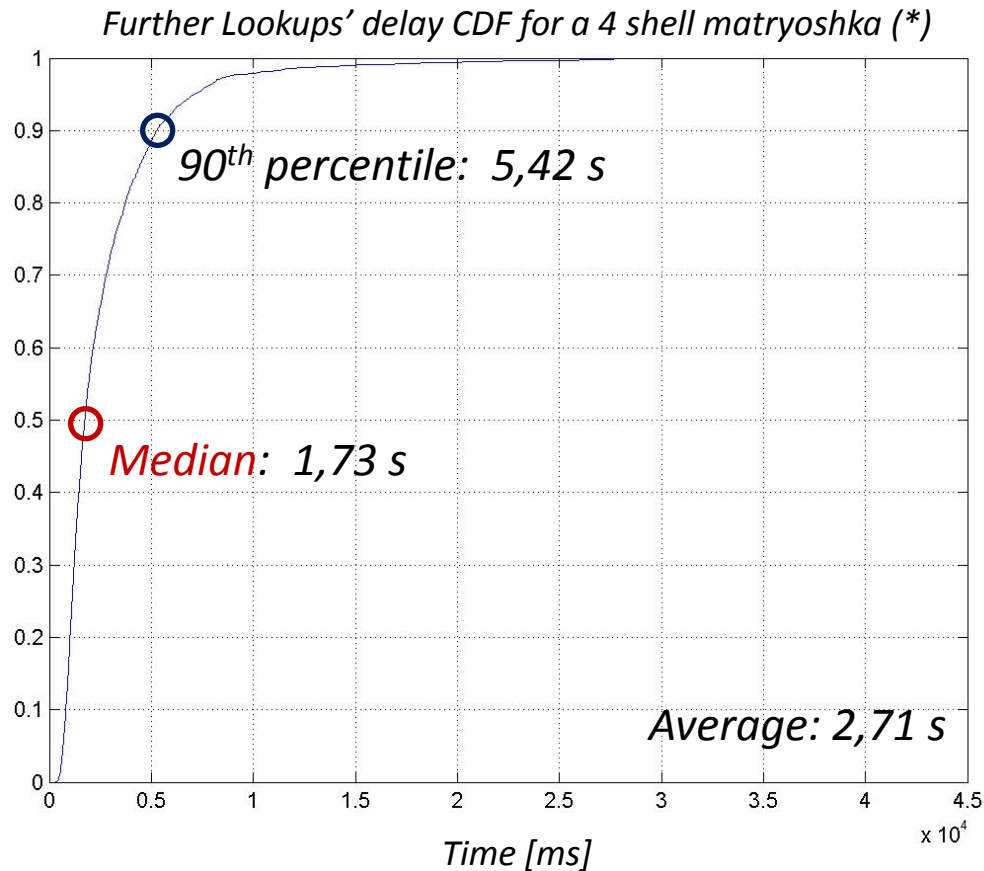
Availability



***Data replication** at friends' nodes*

Performance - Reachability





Total data lookup time:

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

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

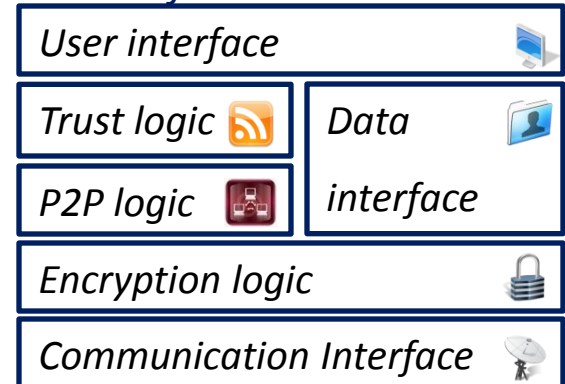
(*) Data computed by applying the Monte Carlo sampling technique on single hop delay measurements and on delay measurement for a successful DHT key lookup in KAD (Biersack..)

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



- 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