



# Peer-to-Peer Networks

Chapter 5: Applications on P2P, Online Social Networks  
Thorsten Strufe

# Chapter Outline



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

# 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

# A Short History of Networked Services...



```

11:16 03/07/09 datavase BBS
Last On: 03/07/09 Mode 1 Up 4 days 1
FirstOn: 31/01/08 Calls 822 of 278

List/Flag Files      File Area Selection      File Search
L List files in dir  J Jump to new file area  F Find text in descrip
N New file scan      * List directories       S Search for filename
E Extended file info /* List libraries

Upload/Download      { } # Select directory   Other Commands
D Download file      [ ] /# Select library    & File scan config
U Upload file        R Remove/edit file
/D Download from user
/U Upload to user
Z Upload to sysop
B Batch/Bi-dir xfers

Anytime | Ctrl-U Who's online
File 0:02:29 (2) Atari (13)
  
```

```

Legend of the
Red Dragon ..
Lunatic..
Telnet...

FREE ACCESS to ..
Tradewars 2002
Lord2....
FTP...

And Much More...

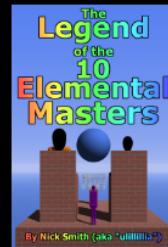
... FROGLAND BBS ...

Running On: Wildcat v5/winserver
Your Sysop: Frog Man
Telnet: bbs.frogland.net
WWW: http://www.frogland.net
Note: Check Out http://www.bbsdnet.com slashdot for BBS's!!!!!!
      Only One Account Per Person

Connected To node:2
Today's Date Is: 3/26/2002
The Time Is: 4:47 PM

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 area and guests going back urban I began it in late 2001. This area 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 [anonymo](#)

**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](mailto: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



### 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  
the world at the end (those'll be  
the '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..." – WNRC New York TV

half of 2009, be sure to add your 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

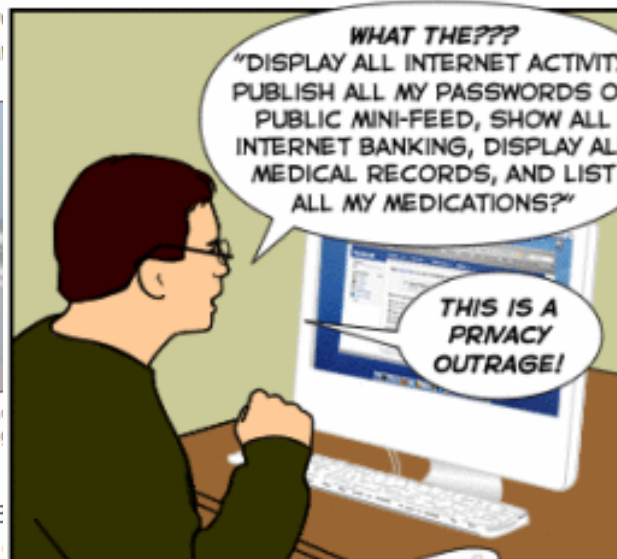
MUSIC LOVERS, SNL, France à l'ONU direct loser we love to hate,

# ...with calculated side effects...



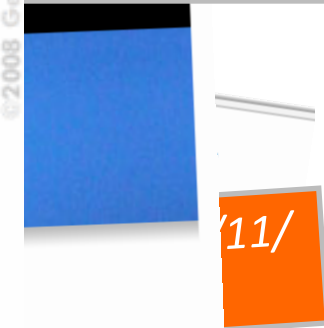
The Joy of Tech™

by Nitrozac & Snaggy



joyoftech.com

3/26/facebook-e-robbery/



robbery

ow you choose y  
reveal.

cent-seeming stat  
atment."

and called Fire De  
g was caught on c  
oked suspiciously

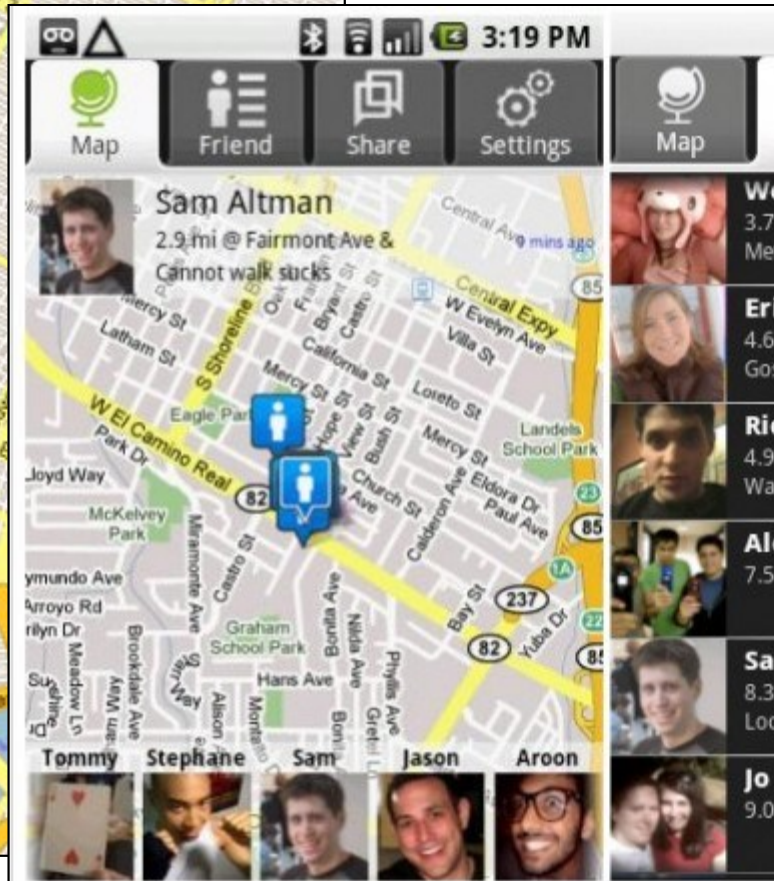
Friday's American



# ...increasingly immersive to daily life...



google



Loop



tweetspotting



# „Volkszählung“ 1987

- Penultimate public census in
- Scheduled for 1981 (delayed)
- Significant public opposition
  - Fear of a surveillance society
  - The transparent citizen („gläserner Mensch“)
  - 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?**  
nein, ja

**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.), Selbständige(r) mit bezahlten Beschäftigten, 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  
BVV Gera MfS 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, Vatersname, Geburtsname, weitere Namen

1.2. Geheimn., Pseudonym, Spitzname

1.3. erste Niederlassung und Tätigkeitsbeginn im Dienst seit  
Januar 1949

1.4. Deutschland, Ottendorf  
Geburtsort (Ermittlung v. Name)

1.5. männlich  
Geschlecht

1.6. deutsch  
Nationalität

1.7. BRD  
Geburtsland (bei Ausländern)

1.8. Weber, K. [redacted], Ehefrau  
Name der Ehefrau (bei Verheirateten)

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

BSIU  
0000

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  
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



- What is Social Media anyways?
  - ***Web-based*** (*user generated*) ***content*** sharing
  - Knowledge about the posting ***individual***
  - Plus the ***relationship*** between the users
- Relationship?
  - „Friend“, „Acquaintance“, „Idol“...?
  - Uni- vs. bi-directional („connect“ vs. „subscribe“)



- Content Communities
- Blogs/Microblogs
- Virtual Gaming Worlds
- Virtual „Social“ Worlds
- Collaborative Projects
- Online Social Networks  
(read: all of the above ;) )

- youtube
- Twitter
- WoW
- 2nd life
- Wikipedia
- FB, g+ (?),

*Are last.fm, pandora  
social media?*

*What about  
amazon/ebay,  
then?*



## 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





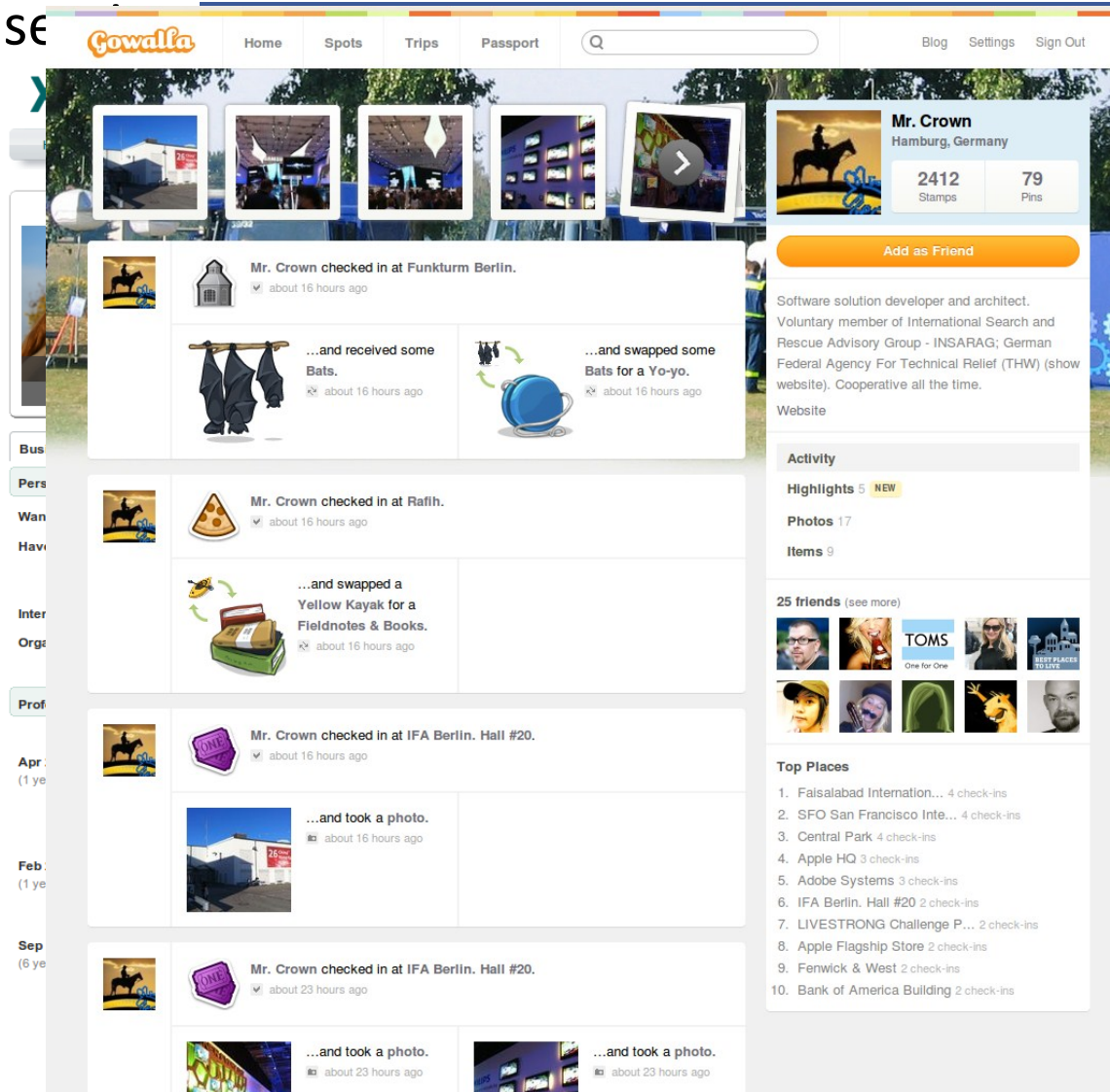
- Simplified, walled-garden version of „the Web“:
    - Easy to set-up pages („**profiles**“) of **individuals** (... and companies...)
    - Links reflecting **real-world relations** between individuals
    - Possibility to share user generated content
  - ...including messaging
    - “Guest book” /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









# Two words on: *Information Sovereignty*



- The Bible
- Newspapers
- Mass media
- The Internet (1.0)
- Web 2.0...



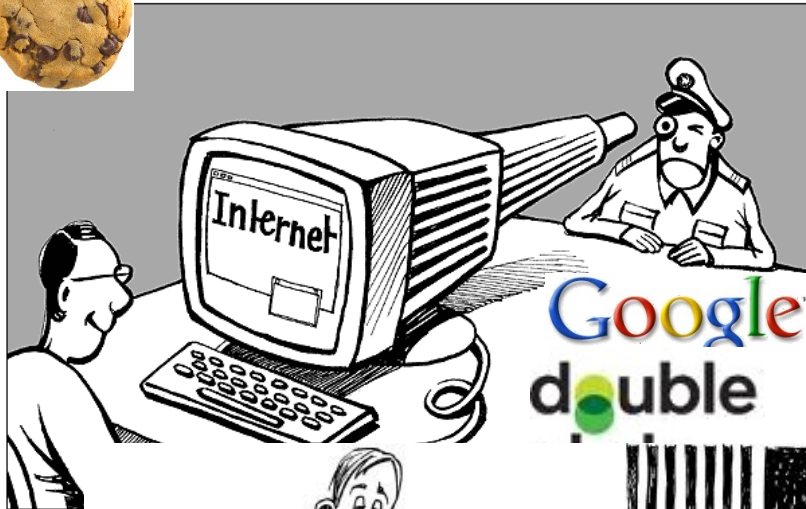
*i.o.w: to which effect? aka.  
„Power to the people!“ (O RLY?)*



**So what do you do?  
Go into the ammo business...**

*The deer now have guns...*

# The Ammo Business...



*„Facebook is and will always be free.“  
„At Facebook your data is yours.“*



*„The best minds of my generation are thinking about which ad to show to whom...“*

*„Myspace, after failing to meet the expected income, starts selling their users' data.“*

# The Deer May have Guns, but...



it's known **who they are**  
with **whom** they're **friends**  
whom they're **talking to**  
what they **think** and **want**  
  
even what they're **aiming** at...



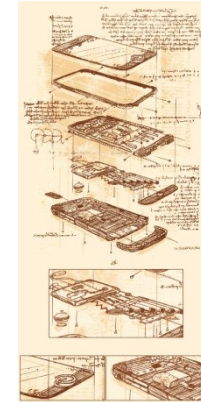
*...and quite fortunately, **their ammo can be removed,**  
**should they fail to comply***





# Can it Get Worse? (a little polemic)

Homogeneity and control! Provide their  
Hardware  
Operating System  
„Applications“ (controlled, be the gatekeeper!)  
Just license them, don't give them any ownership!  
Control/surveil their music  
Their movies  
Their social network...  
...their life...

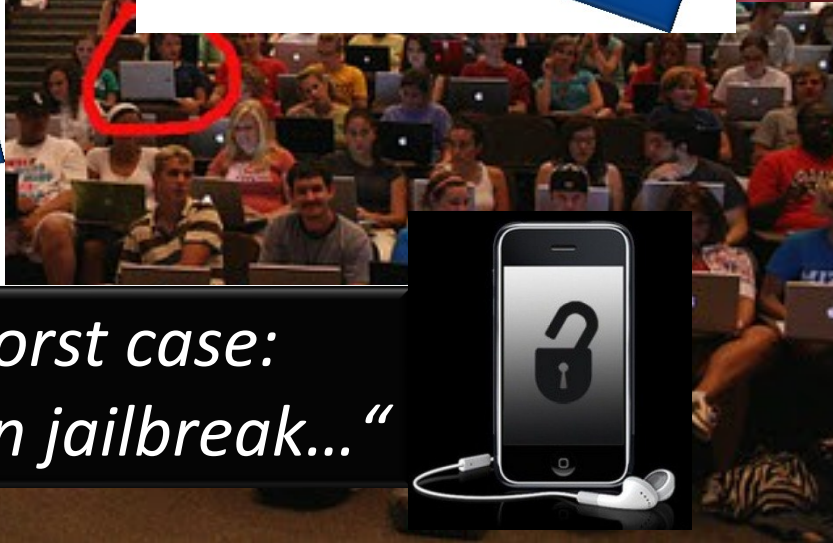




*„And it's a Unix  
under the hood!“*



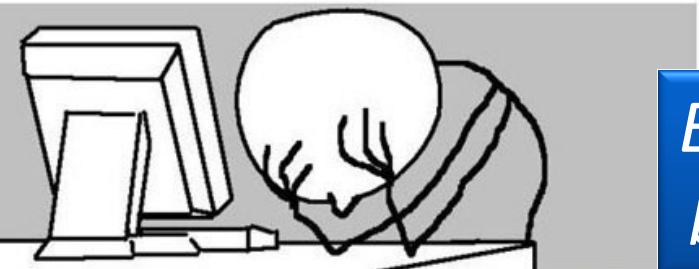
*„Yeah, they're so cool,  
no effort, no hassle!“*



*„Worst case:  
you can jailbreak...“*



# Think diffe



*Enough polemics,  
back to the topic!*



# Back to Social Networks – and: Why Bother?



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

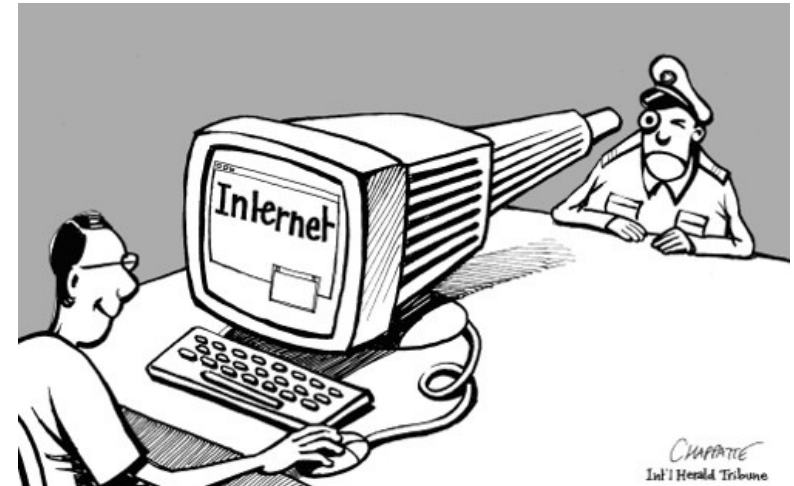
RANK	Brand	Unique Audience (000)	Time Per Person (hh:mm:ss)	MOM UA % Change	MOM Time % Change
1	Google	152,708	1:23:54	4.10%	-16.90%
2	Yahoo!	134,561	2:28:11	4.30%	-26.80%
3	Facebook	116,329	7:01:41	5.80%	9.70%
4	MSN/WindowsLive/Bing	109,425	1:25:22	1.20%	-18.10%
5	YouTube	99,525	1:02:27	7.60%	-10.30%
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** 0 0 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

**Business details** Confirmed contacts (83) **About me** Guestbook Applications

**Personal information**

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

**Professional experience**

+ Add entry

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

**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**

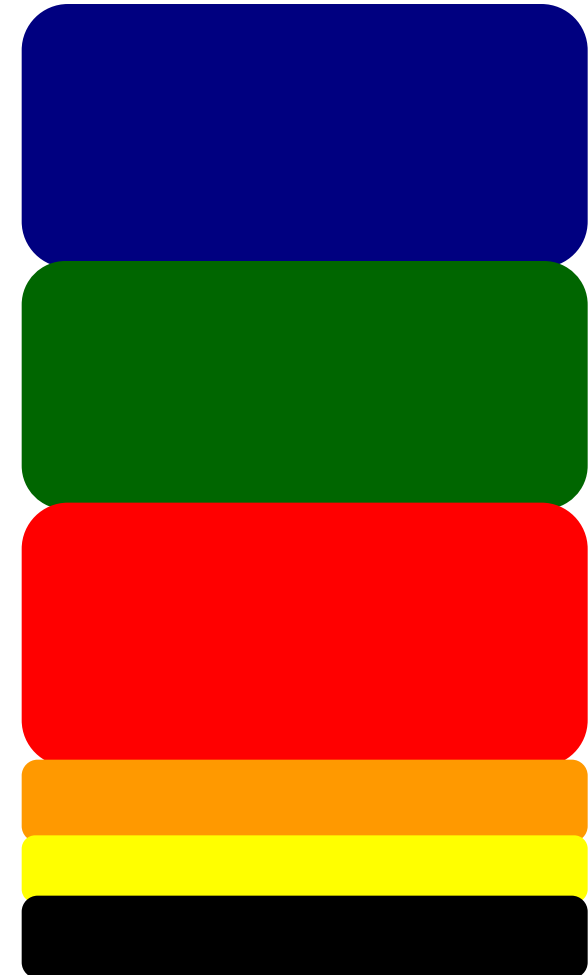
**Thorsten Strufe's statistics**

**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)





- 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 for 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: 

- Tahssin Asfour präsentiert das neue design von [www.pr-bote.de](http://www.pr-bote.de) (Today - Status update - Send message)
- Mario Melle joined the "Neuer Personalausweis (nPA) für ..." group. (Today)
- Detlef Mämpel joined the "Army Network" group. (Today)
- Dr. Thomas Rücker will be attending the "Netzwerkgruppe Innovation" event. (Today - Send message)
- Stella Pierl is now connected to Julian Hühnemann (Bigpoint GmbH). (Today)
- Only Premium Members joined the "Innovationsfinanzierung durch Business ..." group. (Yesterday)
- Only Premium Members has changed her private address. (Yesterday - Send message)
- Only Premium Members is now connected to Christoph Lühr (basilicom gmbh). (Yesterday)

 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)

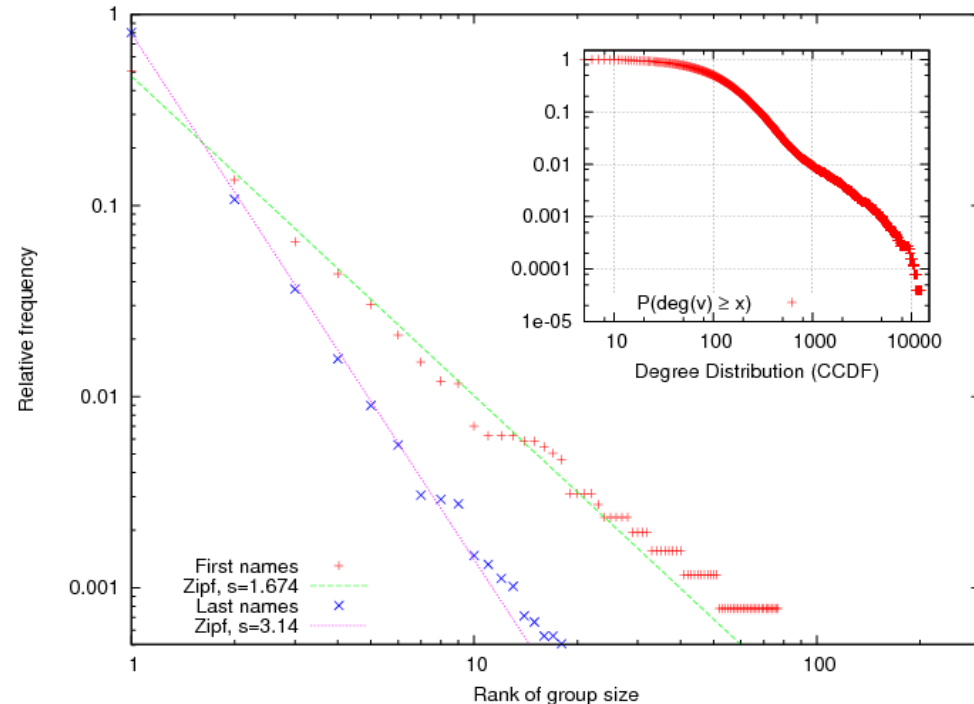
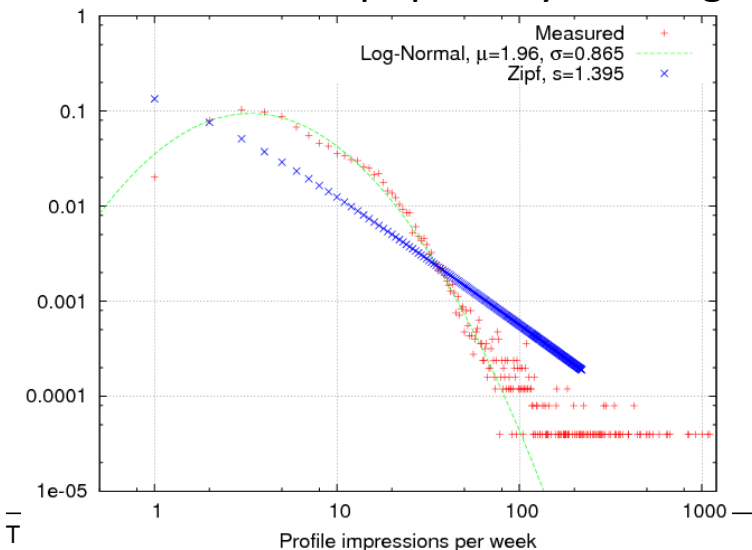
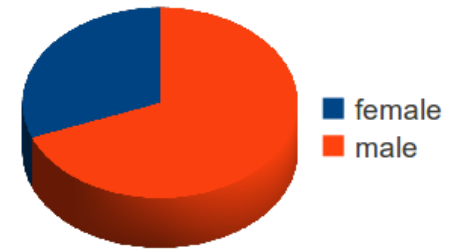


- Identifying Data
  - Name, Image, Gender
- CV
  - Current employment, universities attended, claimed spoken languages
- Interests
  - Interests as stated
  - Number of subscribed groups, subscribed groups, number of members in groups, number of messages in respective groups, languages of group
- Contact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





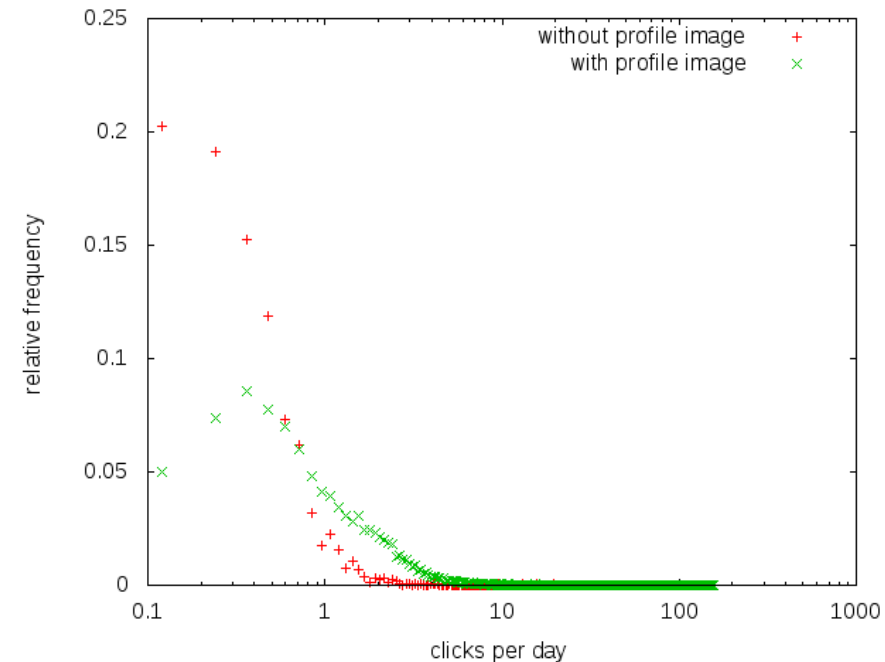
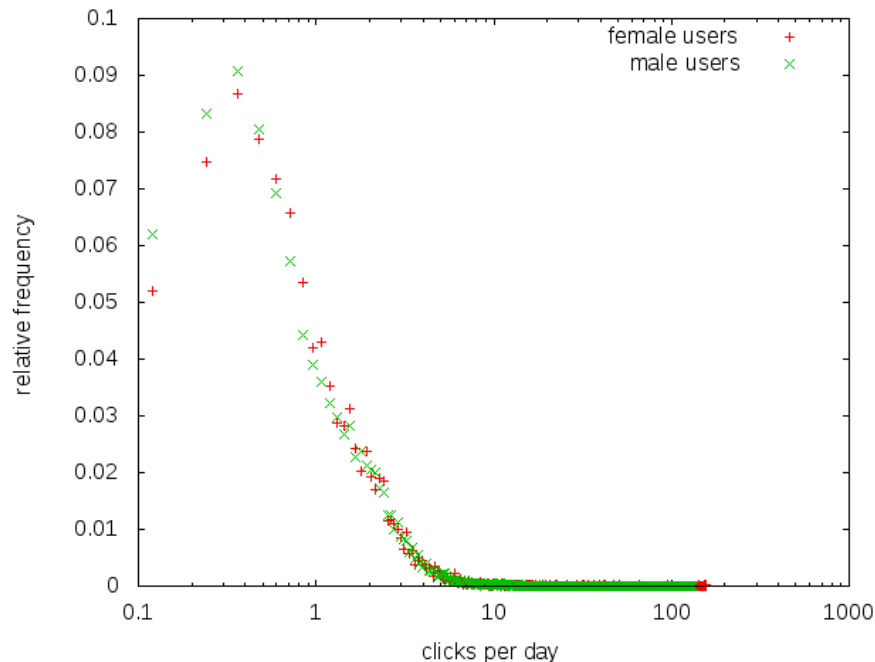


- Activity
  - “Activity meter” in profiles very coarse grained
  - Derived “profile alteration frequency” as alternative
  - Men are slightly more active than women (to both metrics)
  - Profiles without image belong to inactive users
  
- Membership in Groups
  - Wide range of group membership
  - 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 $r$ )



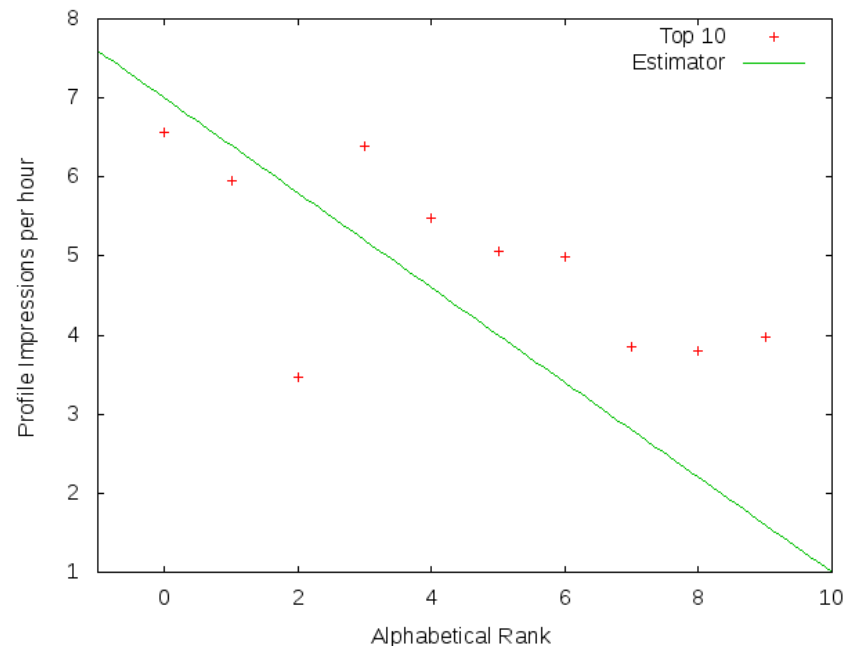
- *“The profiles of active users are more attractive”*
  - Activity as given on the profiles (**“activity meter”**)
  - $r \approx 0.17$ , no noteworthy correlation
  - Activity measured in **group memberships**
  - $r \approx 0.37$  (higher for men, lower for women)
  - Activity meter is very coarse 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:  $0.74 < r \leq 0.75$
- ***In retrospective: Combination somewhat 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!

# Trust



What is „**trust**“, in the first place?

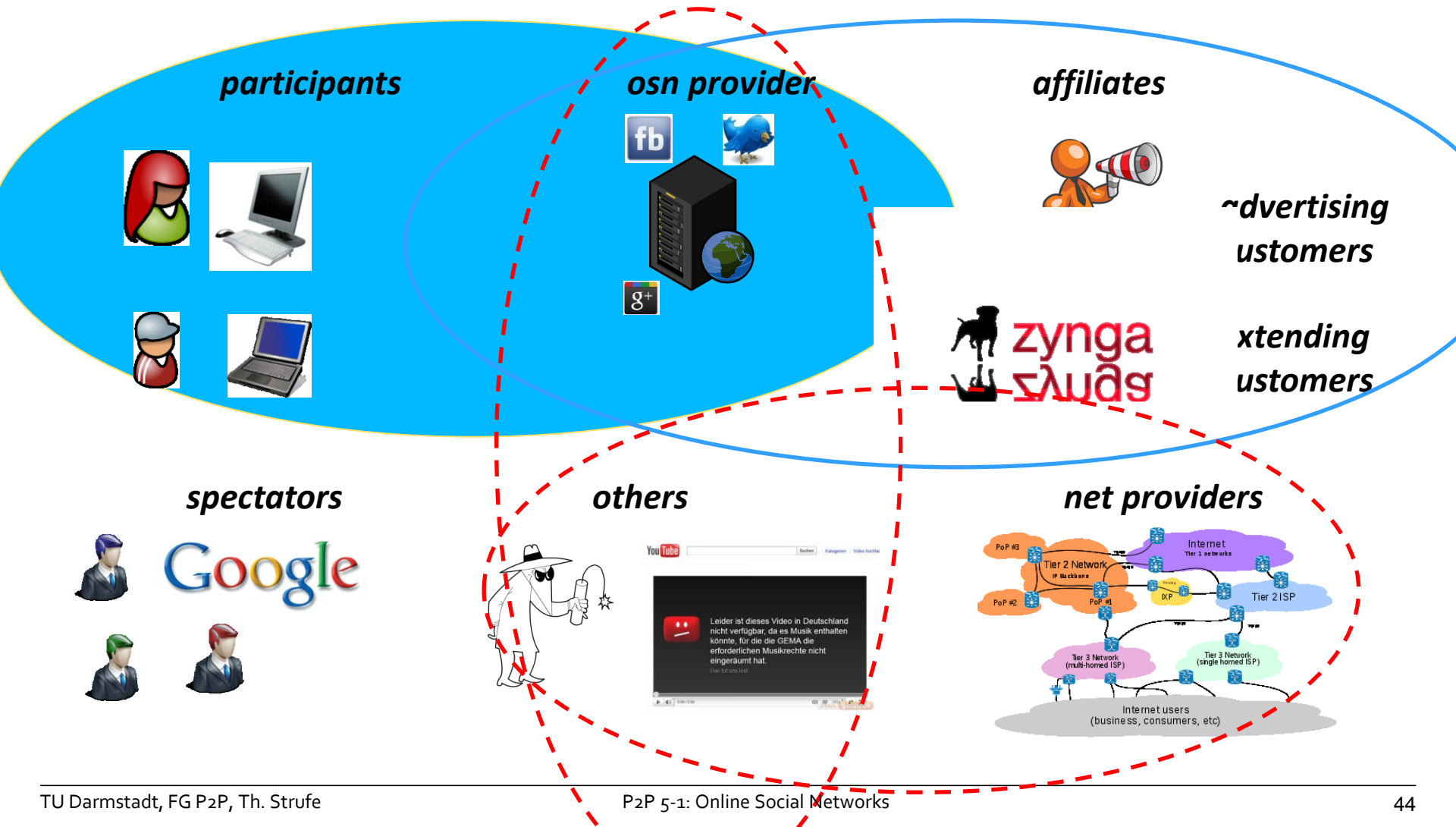
- *Trust is the certainty that another party (a subject) is going to act, operate, behave **as expected**.*

What is so special about trust in OSN?

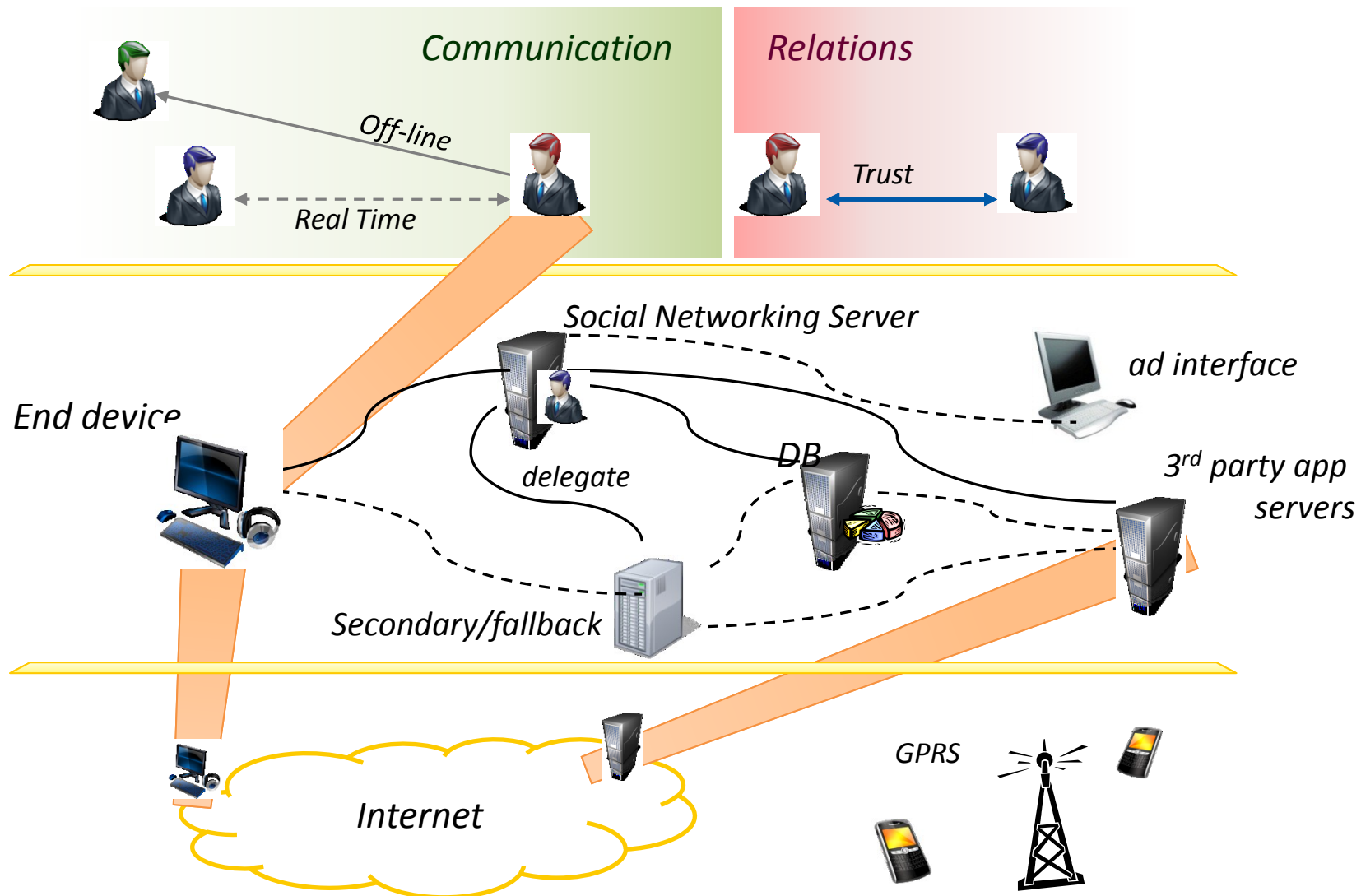
- Service focuses on **individuals** (whom the user knows)
- and their **relations** (*seeming* correspondence to RL trust)
- **Object** of trust:
  - wealth of personally identifiable information (PII)
  - RW reactions to seemingly personal communications



# Stakeholders in OSN


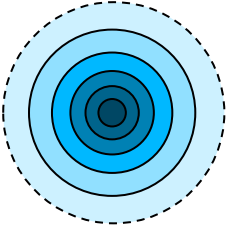







# A Little Model



# Access Model



- user 
- Grantable 
  - specific contact(s) 
  - contacts 
  - contacts of contacts 
  - service subscribers 
  - public 

## • Implicit

- SNP  

*Everything the installing user can see*

## • Affiliated

- Extended
- Advertiser



*Not much (aggregates)  
Unless they pay really well*

## • ISP



*Everything their subscribers see/write  
(until Nov 21st '12)*

*Ignoring downstream abuse...*

# Types of Information



## ■ Explicit

- created content (+self descriptive)
- annotations/comments
- preferences/structural interaction (contacts, +1, etc)



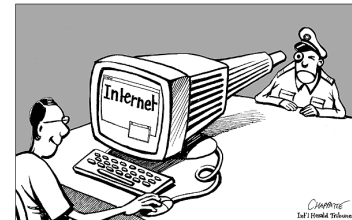
## ■ External

- interest/preferences (clickstreams through adnetworks, fb-connect)



## ■ Implicit

- inferable from environment (homophily)
- observable
  - **session artifacts** (time of actions), **interest** (retrieved profiles; membership in groups/ participation in discussions), **influence** (users)
  - **clickstreams**, ad preferences, exact sessions, **communication** (end points, type, intensity, frequency, extent), **location** (IP; shared; gps coordinates), **udid**



## ■ Processed/derived

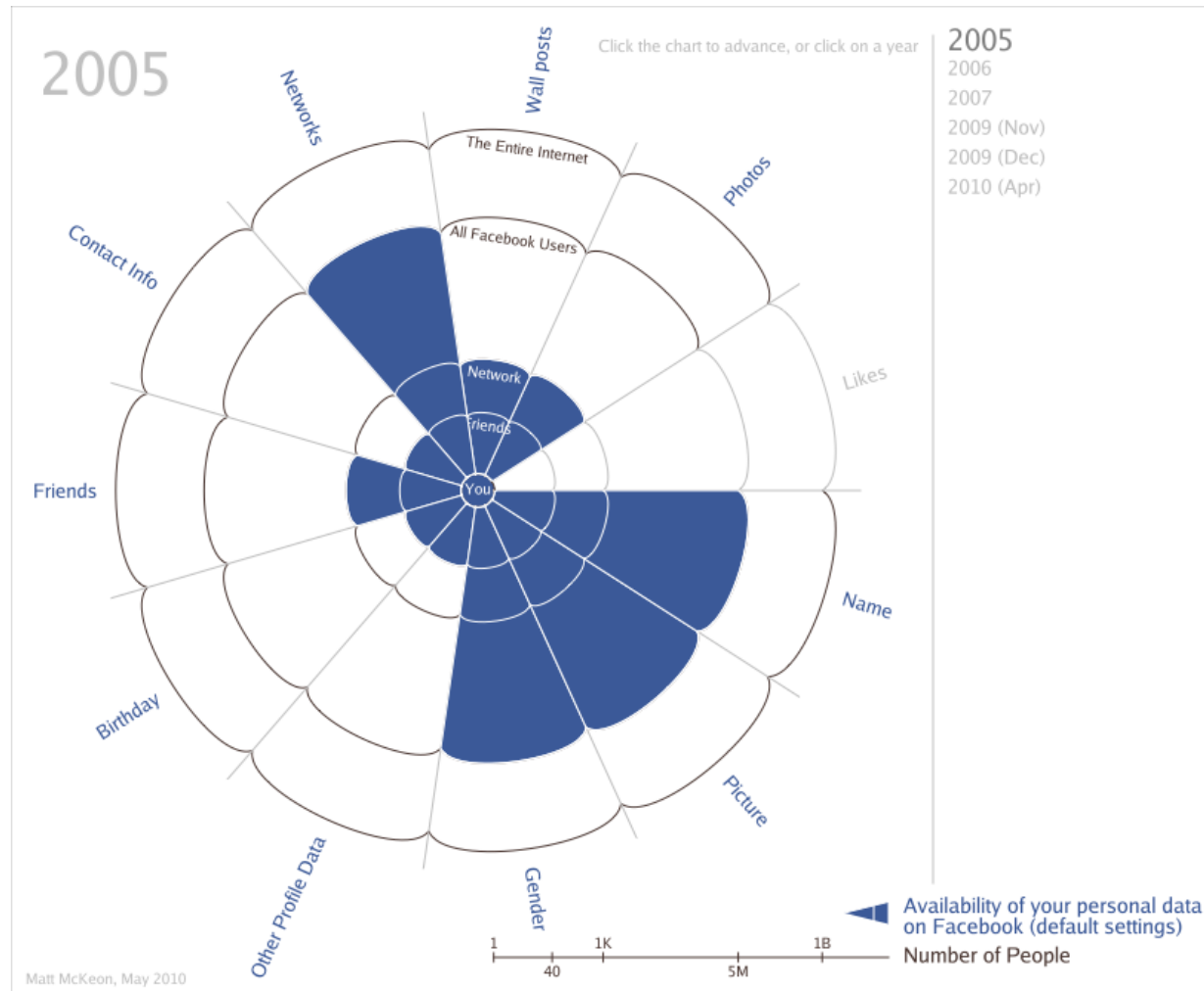
- preference models
- **image recognition models**
- aggregates (characteristic properties/ descriptors of user groups)



- 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!)

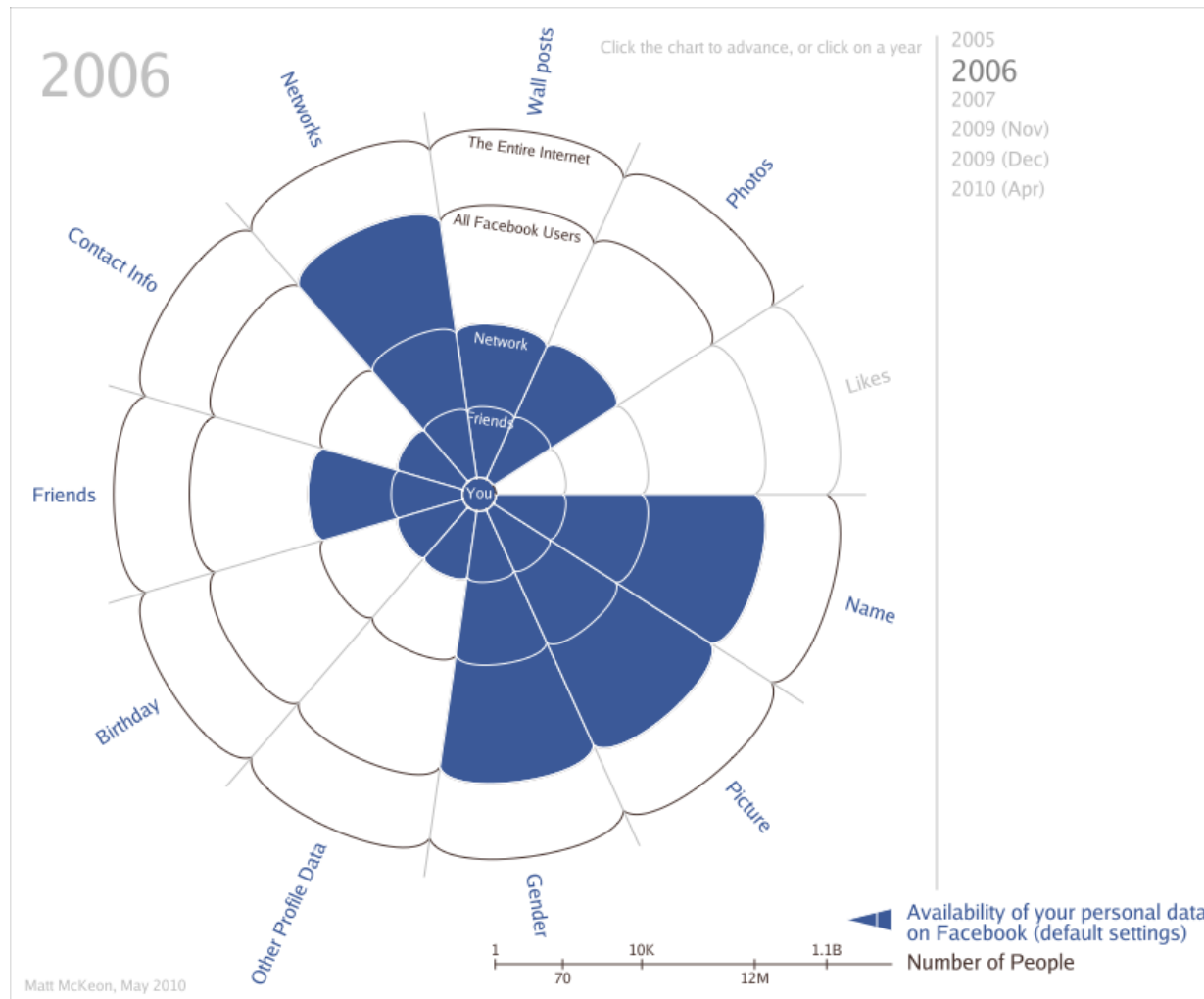


# Facebooks Privacy-Evolution: 2005

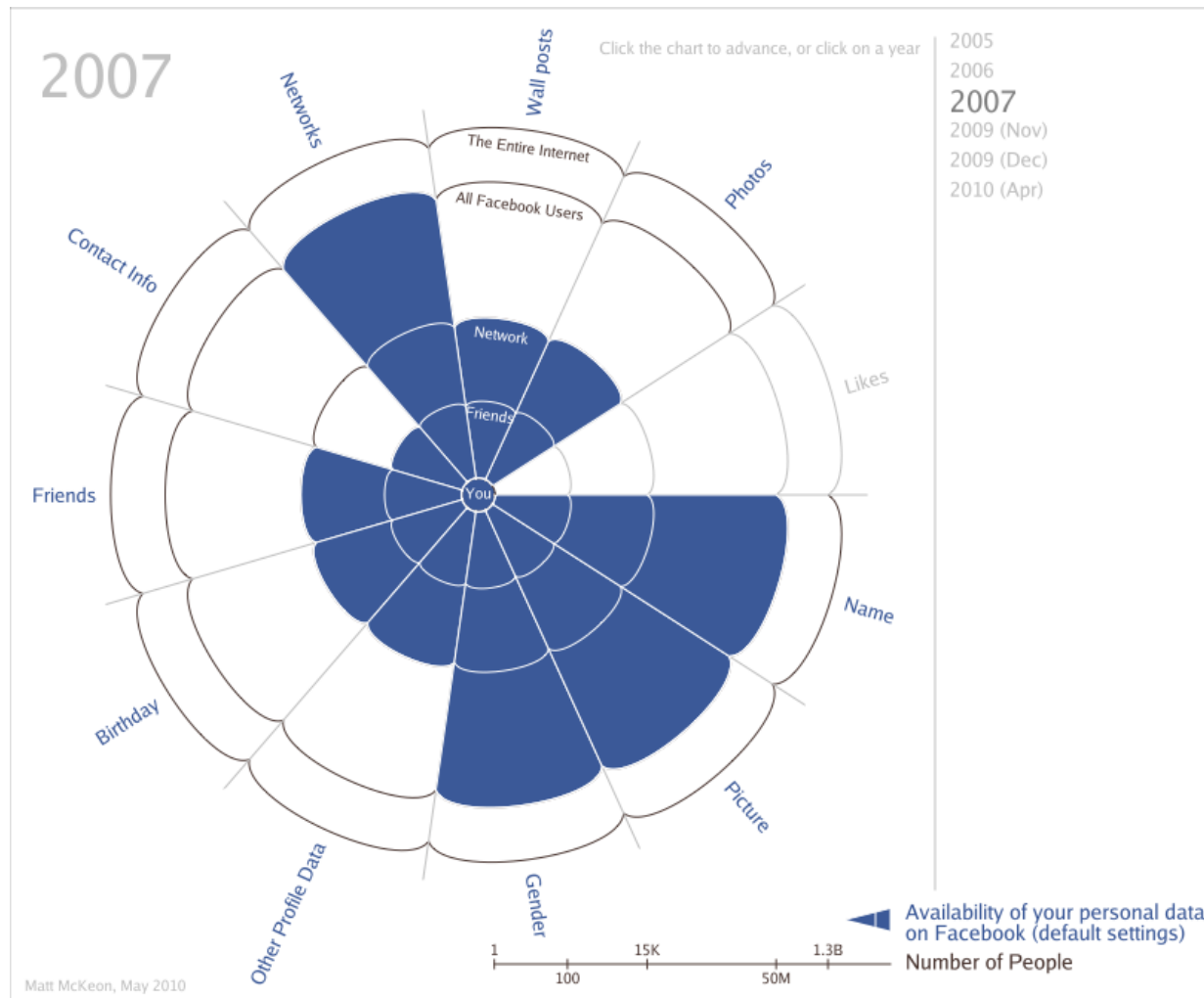


Source: McKeon

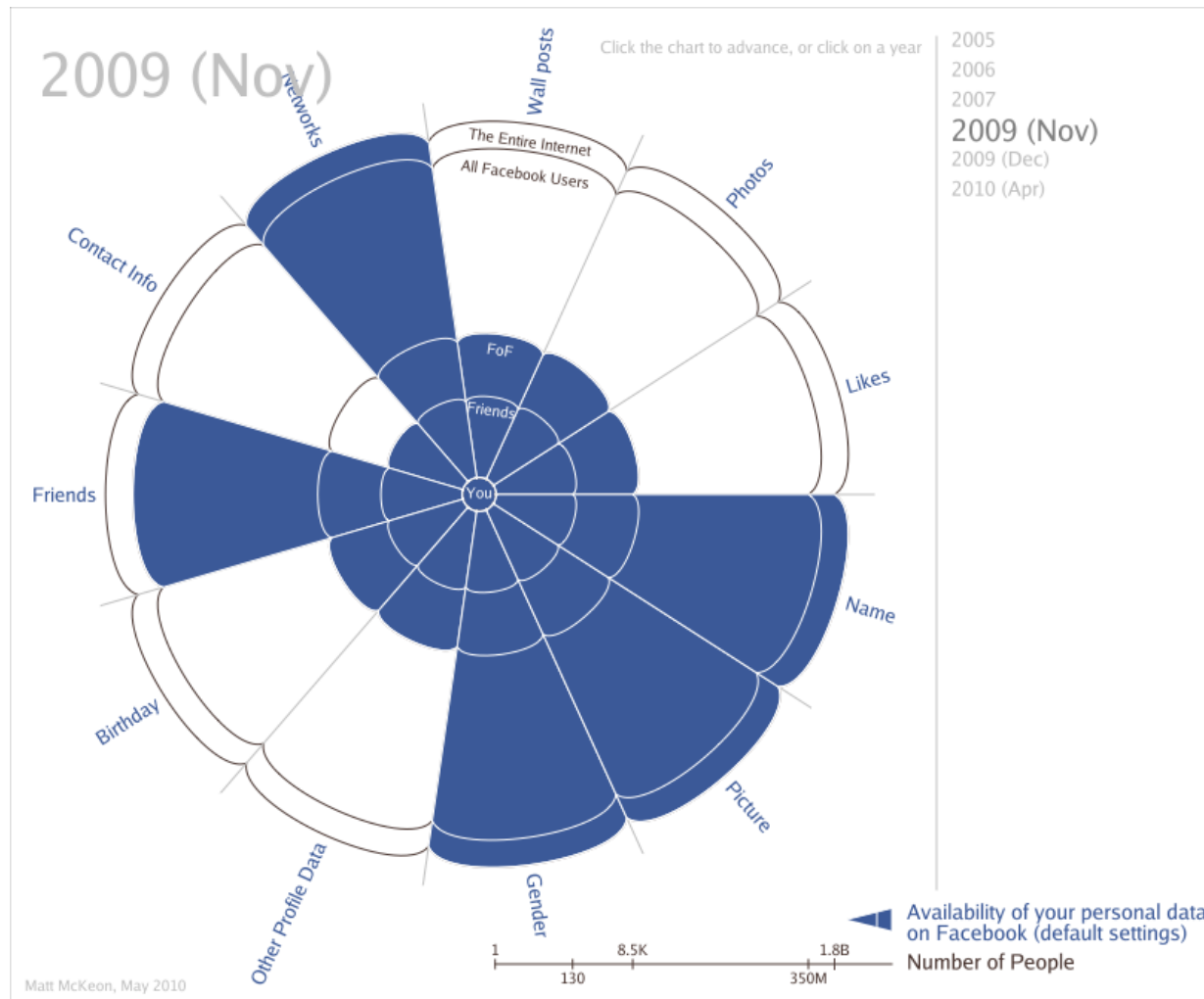
# Facebooks Privacy-Evolution: 2006



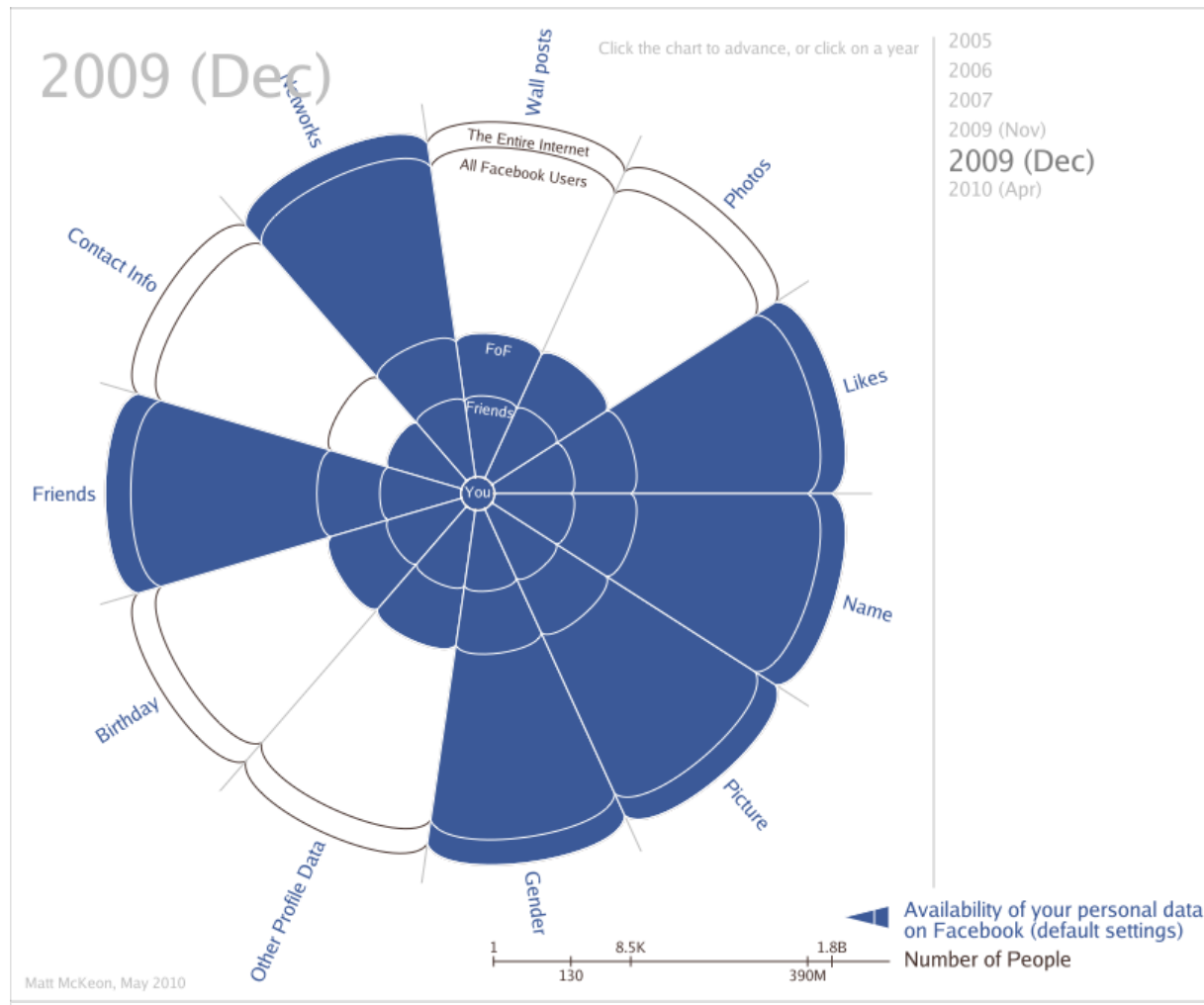
# Facebooks Privacy-Evolution: 2007



# Facebooks Privacy-Evolution: 2009

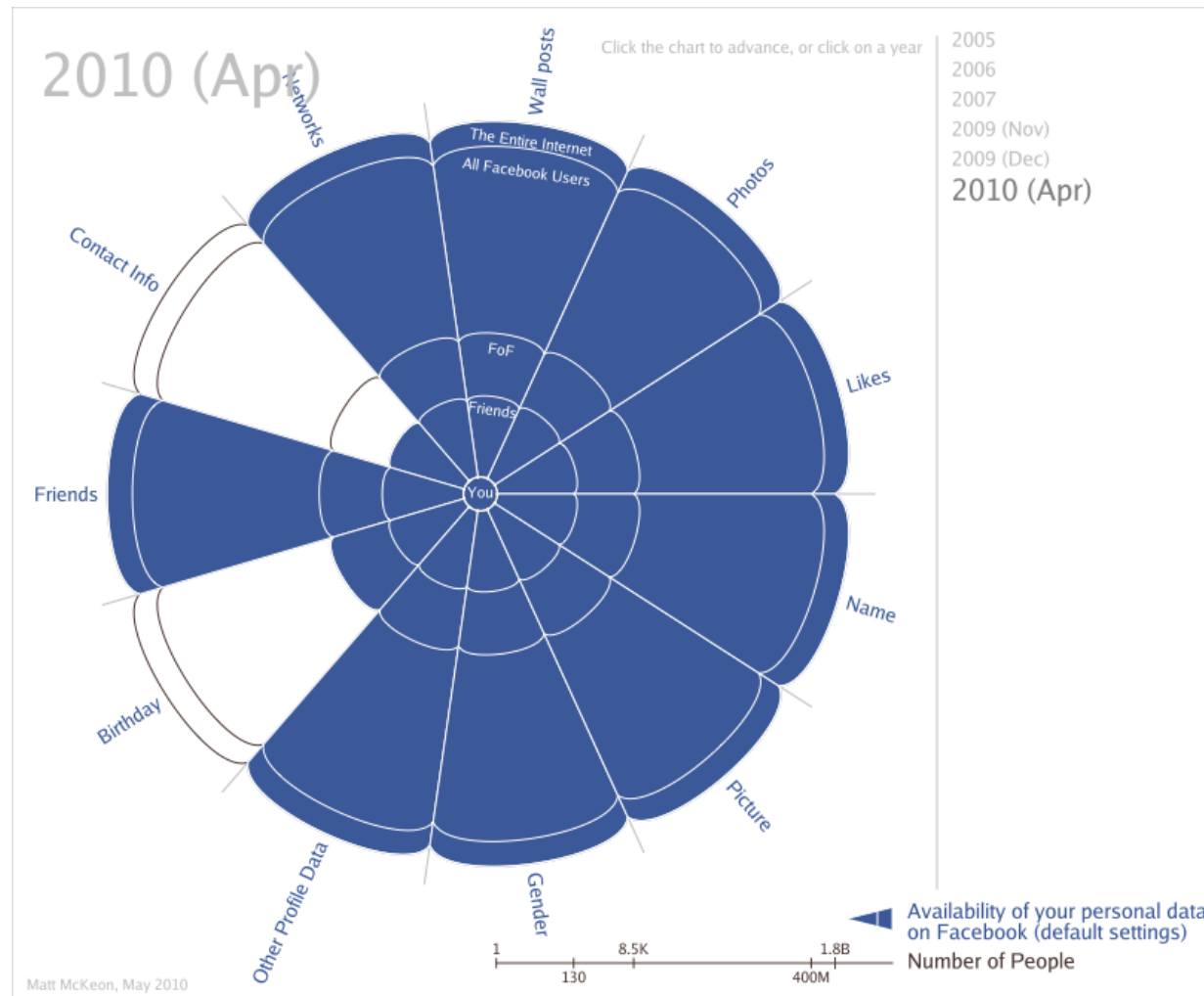


# Facebooks Privacy-Evolution: 2009





# 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

DAILY PROGRAM

PROCEEDINGS

Papers and Presentation

Track: Security and Privacy / Session: Web Security

BECOME A SPONSOR

**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 become a part of our daily life. Well known examples are MySpace, Facebook, XING, and LinkedIn. These sites provide a platform for users to connect with their friends and colleagues. In this paper, we present a new type of attack on social networking sites called 'Automated Identity Theft'. We show how an attacker can clone a user's profile and use it to steal sensitive information from the user's contacts.

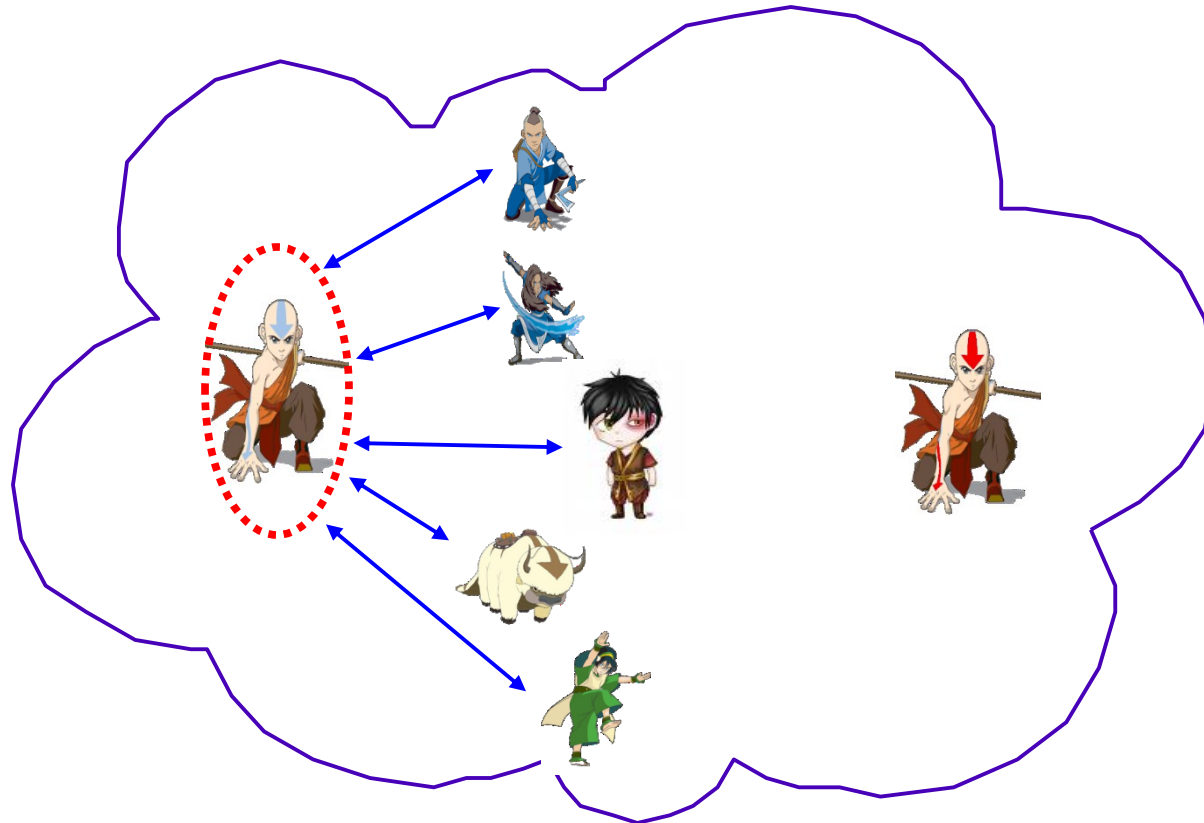
**1. INTRODUCTION**  
A social network is a social structure that is made up of nodes representing individuals or organizations. These nodes may be tied to each other by properties such as friendship, common values, visions, ideas, business relationships and general interests. Although the idea of social 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], and Friendster [9] have gained significant popularity in recent years.

- 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



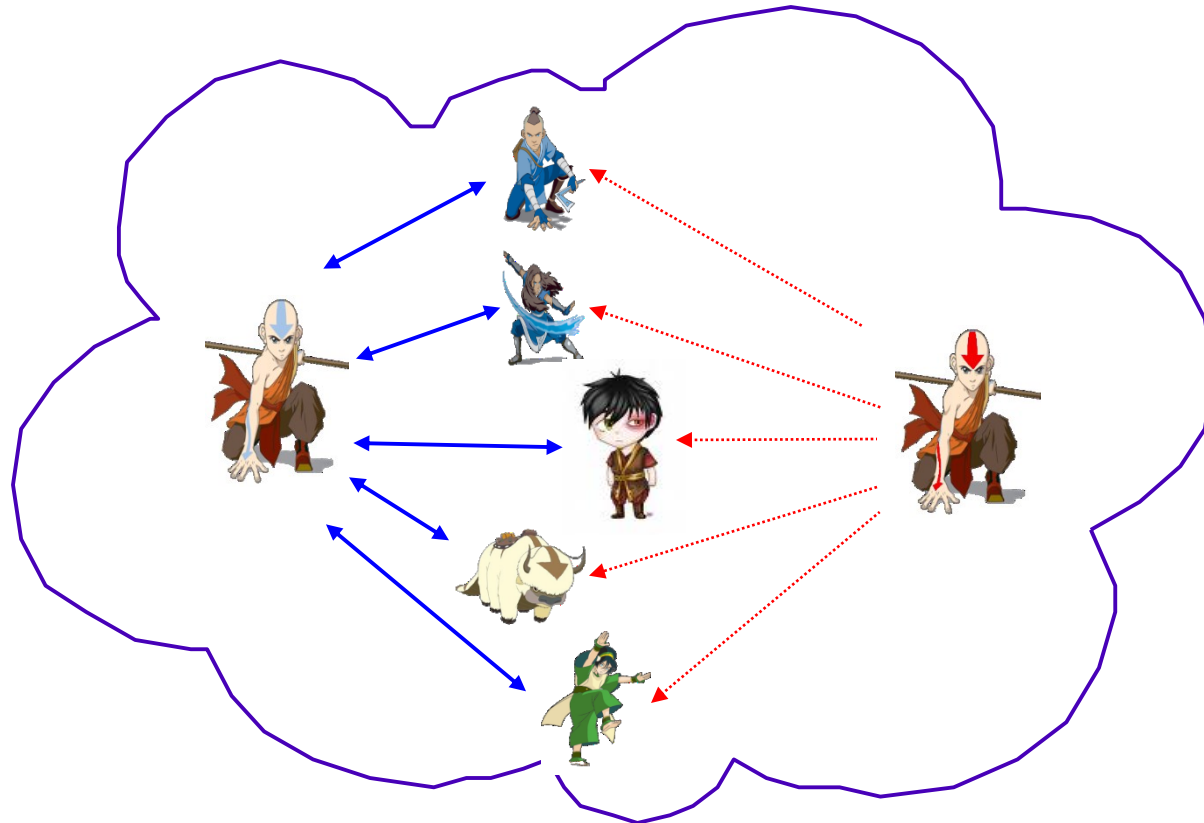
- Is it possible for an attacker to launch impersonation attacks on a large scale against a number of popular social networking sites?
  - Facebook (international)
  - XING (international)
  - LinkedIn (international)
  - MeinVZ (popular in Germany, Austria, Switzerland)
  - StudiVZ (popular in Germany,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

# Automated Profile Cloning

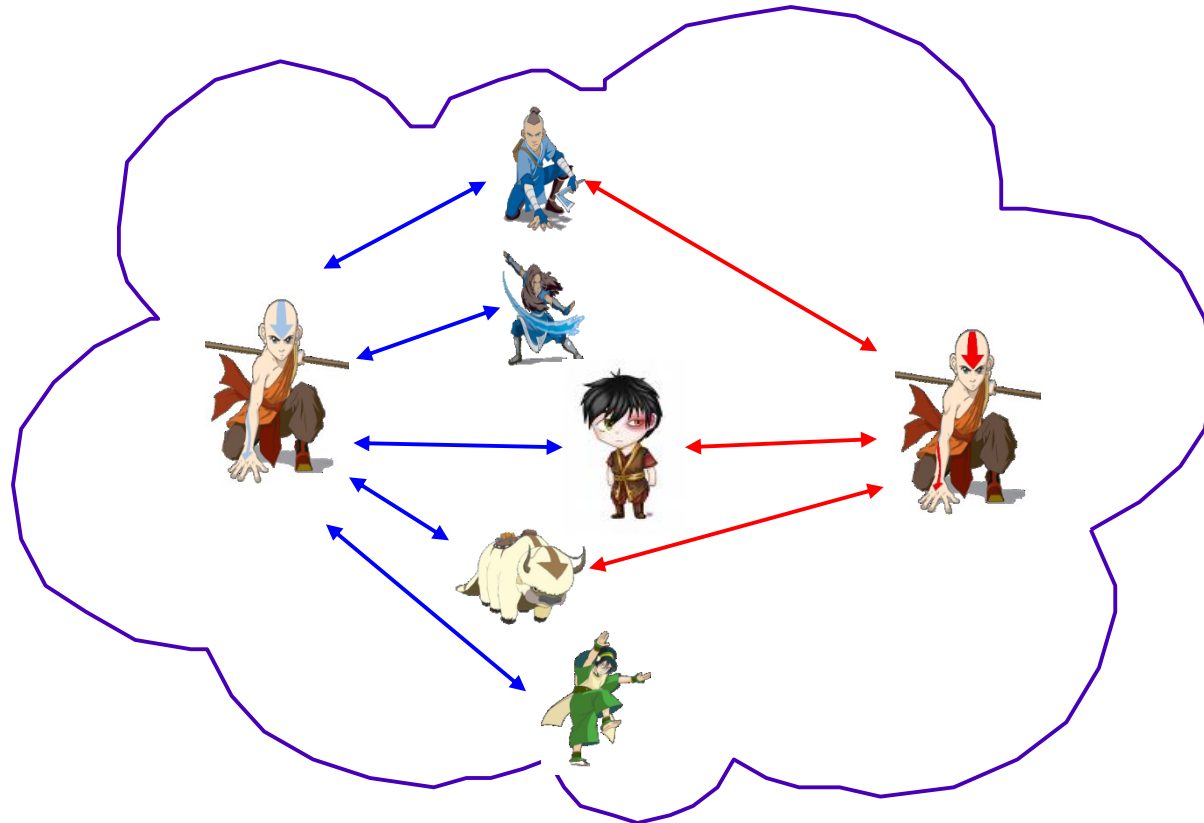




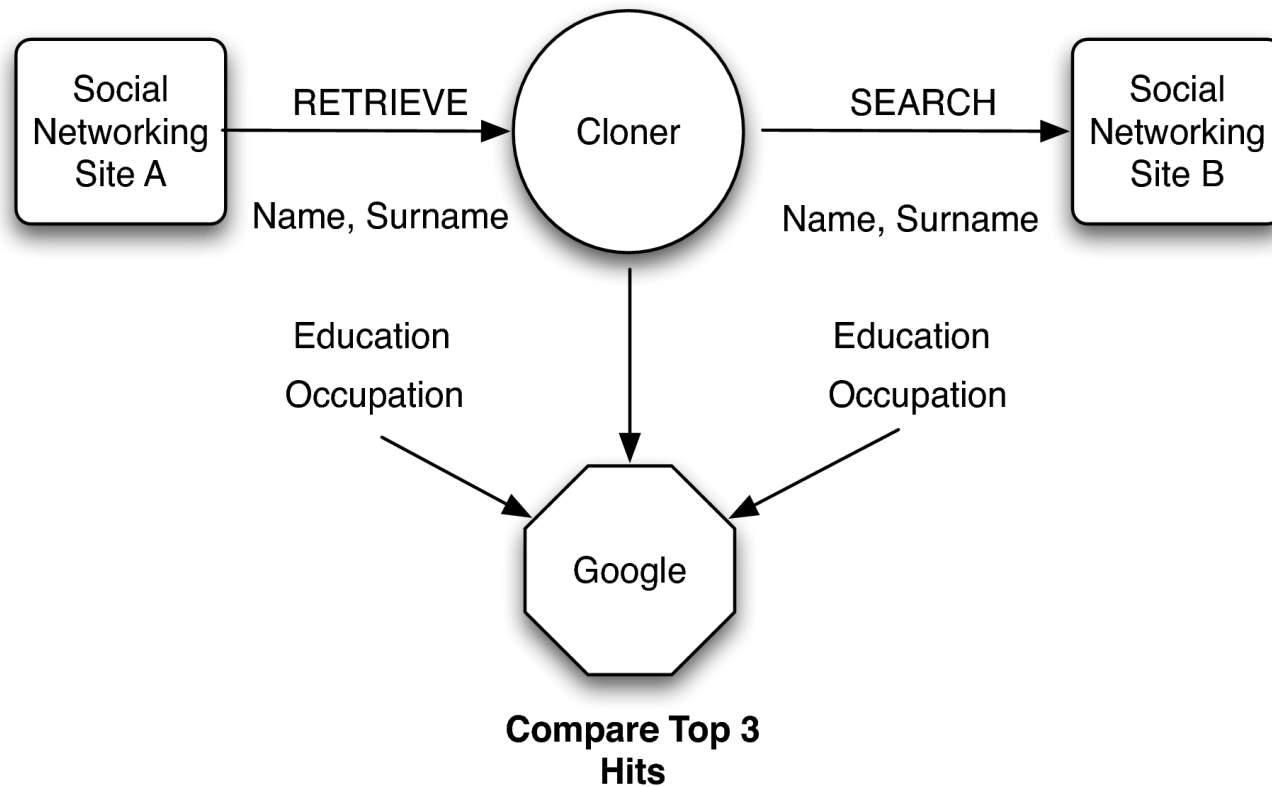
# Automated Profile Cloning



# Automated Profile Cloning



# Cross-Site Profile Cloning

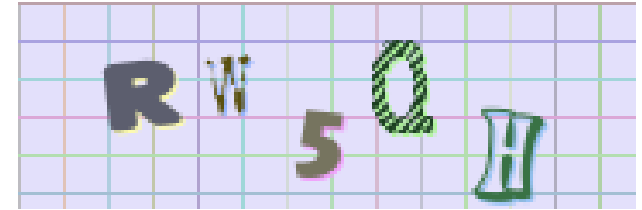




- CAPTCHA: Completely Automated Public Turing test to tell Computers and Humans Apart
- CAPTCHAs are employed to prevent automated programs from accessing and abusing the services
- In order to automate the attacks, a number of CAPTCHA breaking techniques were developed
  - “Quick and dirty”, techniques 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*

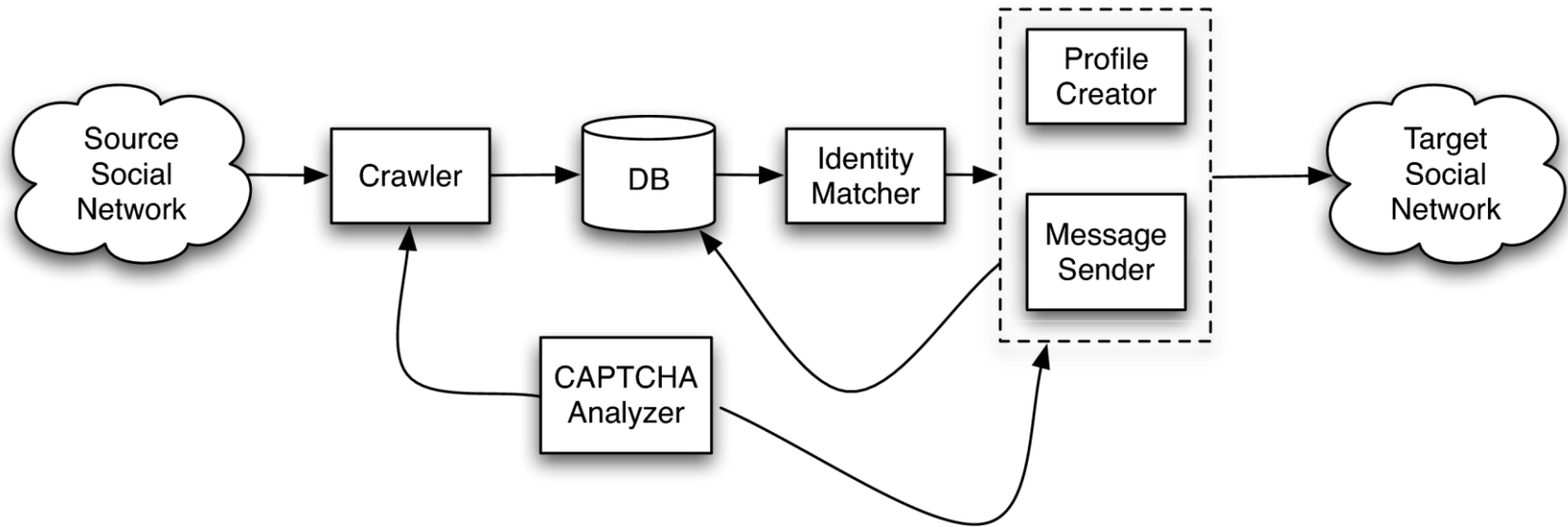


- Adopts ReCAPTCHA technique
  - Asks words that are encountered while digitizing books that cannot be correctly recognized by the OCR program
  - By solving the CAPTCHAs, the user contributes to the effort to increase the accuracy of the text of the digitized book
- ReCAPTCHA asks meaningful words. 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?

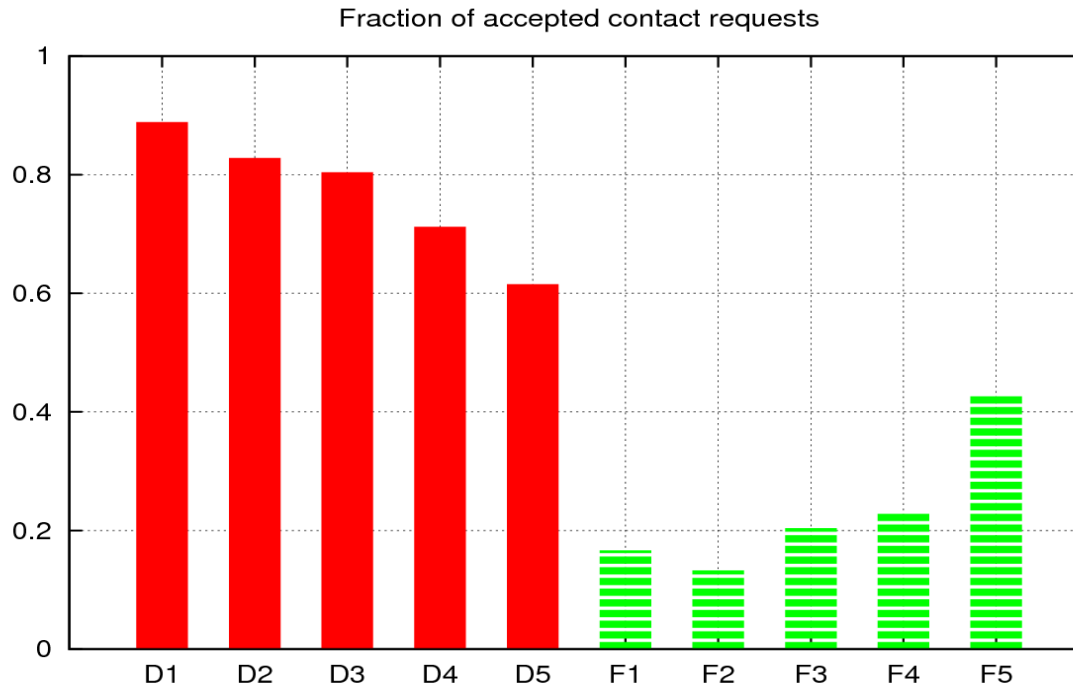


- StudiVZ and MeinVZ
  - Displays CAPTCHA if large number of requests come from one account
  - To collect as much information as possible, without being noticed, 16 accounts were created, and separately used for crawling
  - Collected 5M profiles with contact information, and 1.2M complete user profiles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

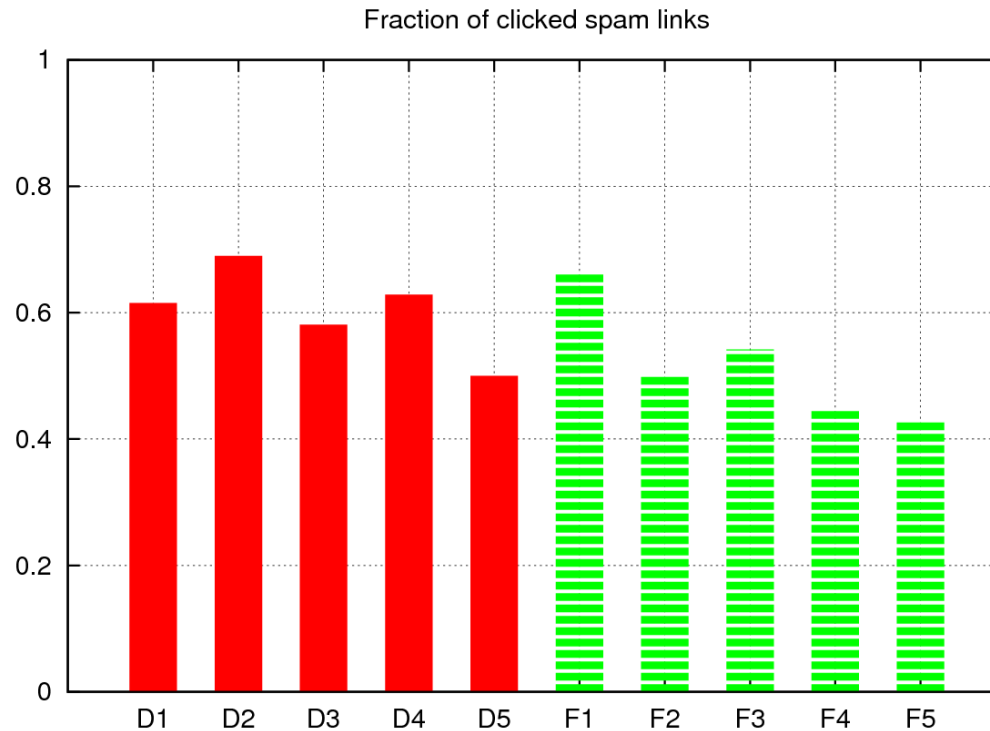
`http://193.55.112.123/userspace/pix?user=<account>  
&guest=<contact>&cred=3252kj5kj25kjk325hk}`

Ciao, <account first-name>

# Profile Cloning Experiments



- Click through rate for messages from duplicate / fictitious profiles





# Cross-Site Profile Cloning Experiments



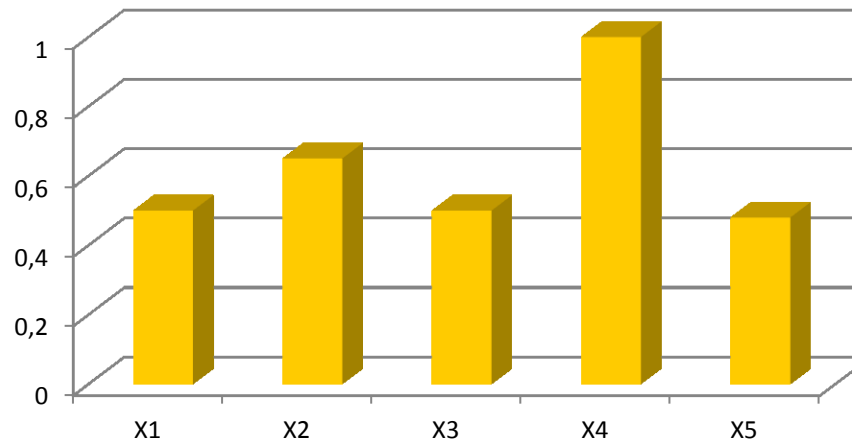
- Cloning profiles that exist on XING, but not on LinkedIn
- The success of the cross-site profile cloning depends on the number of users that have a profile in both of the networks
- From around 30.000 crawled profiles in XING, 3.700 were also registered in LinkedIn

# 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?

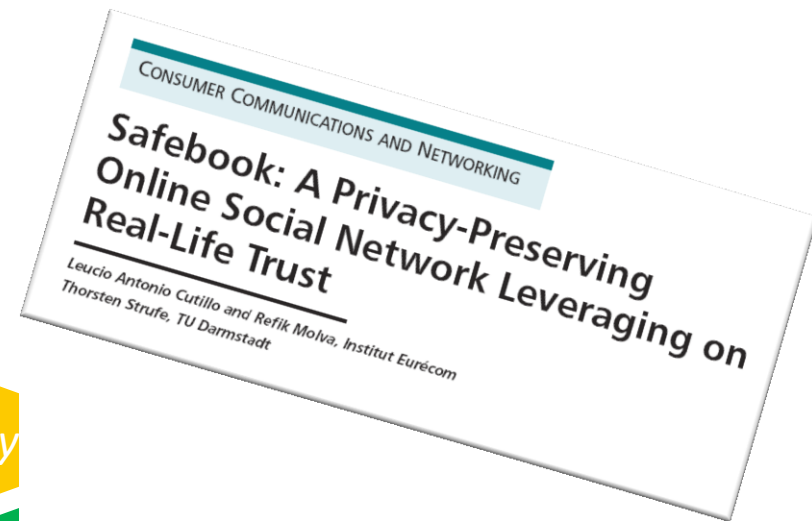


- Make the users aware of the risk
- Make privacy controls easier to use (and show consequences!)
- Provide a stronger way to authenticate users
  
- Increase defense (rate ctrl against massive automatic harvesting):
  - Make CAPTCHAs more difficult to break
  - Reduce response rates
  - Limit number 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*



*Availability*

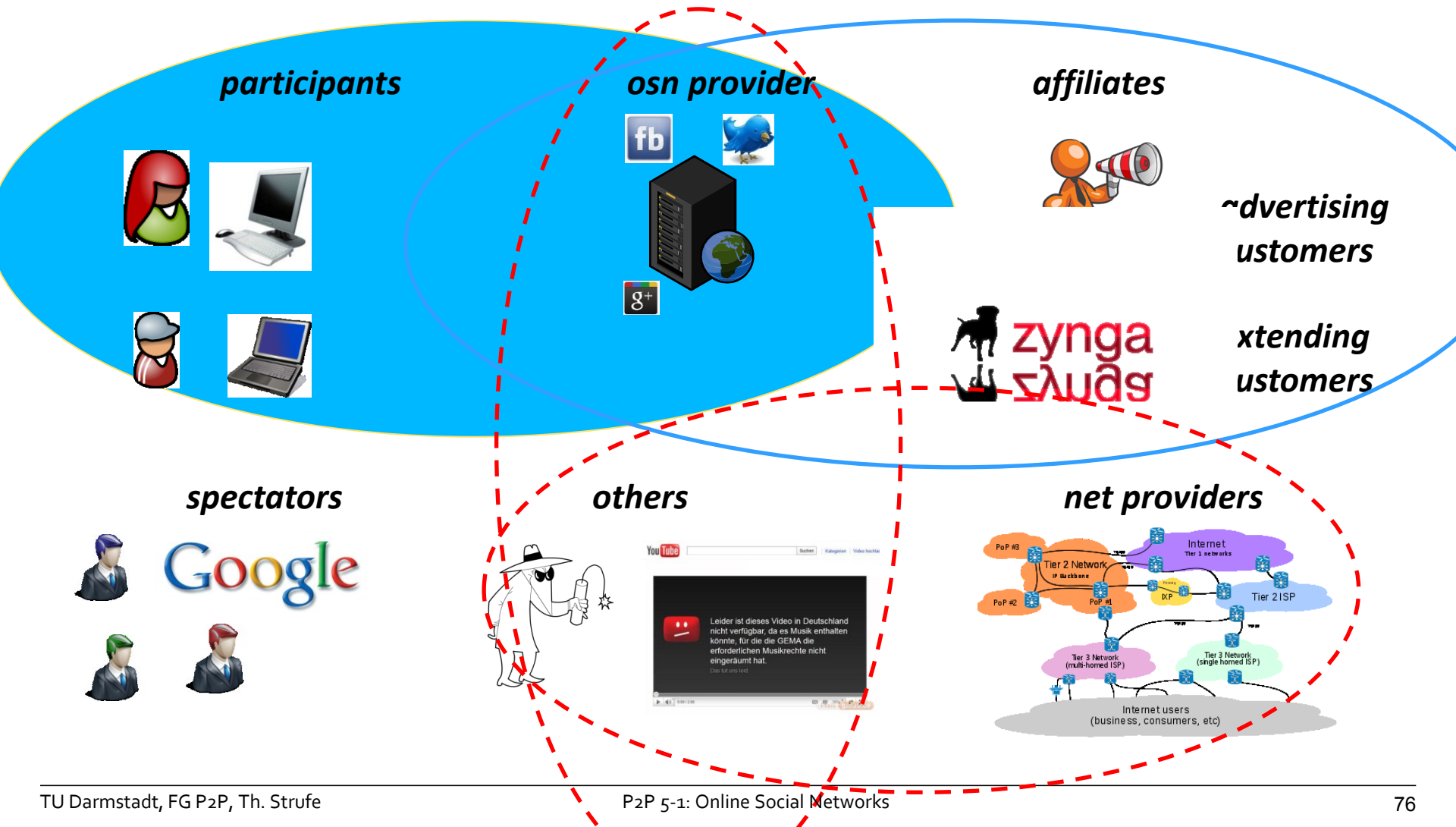
1. *Only trusted parties can access:*
  - *Personal Information*
  - *Communication*
2. *Possibility to hide any data about any user*

*Privacy*

*Integrity*

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

# Stakeholders in OSN





## ■ Explicit

- created content (+self descriptive)
- annotations/comments
- preferences/structural interaction (contacts, +1, etc)



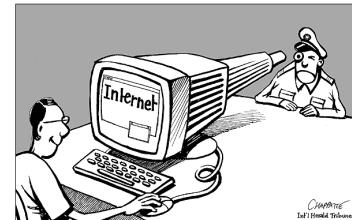
## ■ External

- interest/preferences (clickstreams through adnetworks, fb-connect)



## ■ Implicit

- inferable from environment (homophily)
- observable
  - **session artifacts** (time of actions), **interest** (retrieved profiles; membership in groups/ participation in discussions), **influence** (users)
  - **clickstreams**, ad preferences, exact sessions, **communication** (end points, type, intensity, frequency, extent), **location** (IP; shared; gps coordinates), **udid**



## ■ Processed/derived

- preference models
- **image recognition models**
- aggregates (characteristic properties/ descriptors of user groups)





- **Storage:**  
clear text at the provider
- **Authorization:**  
audience selection by owner  
through SNP functions (per  
object, per subject/group)
- **Authentication:**  
login to account at SNP
- **Access Control:**  
through SNP upon content  
request/presentation to  
authorized subjects only
- **User/Identity Mgmt:**  
Account creation (email),  
SNP functions for contacting  
user, group/list management



## Explicit

- user 
- Grantable 
  - specific contact(s) 
  - “friends” 
  - friends of friends 
  - service subscriber 
  - The public 

## Implicit

- SNP  

*Everything the installing user can see*

- Affiliates
  - Extenc
  - Advert



*Not much (aggregates)  
Unless they pay really well*

- ISP

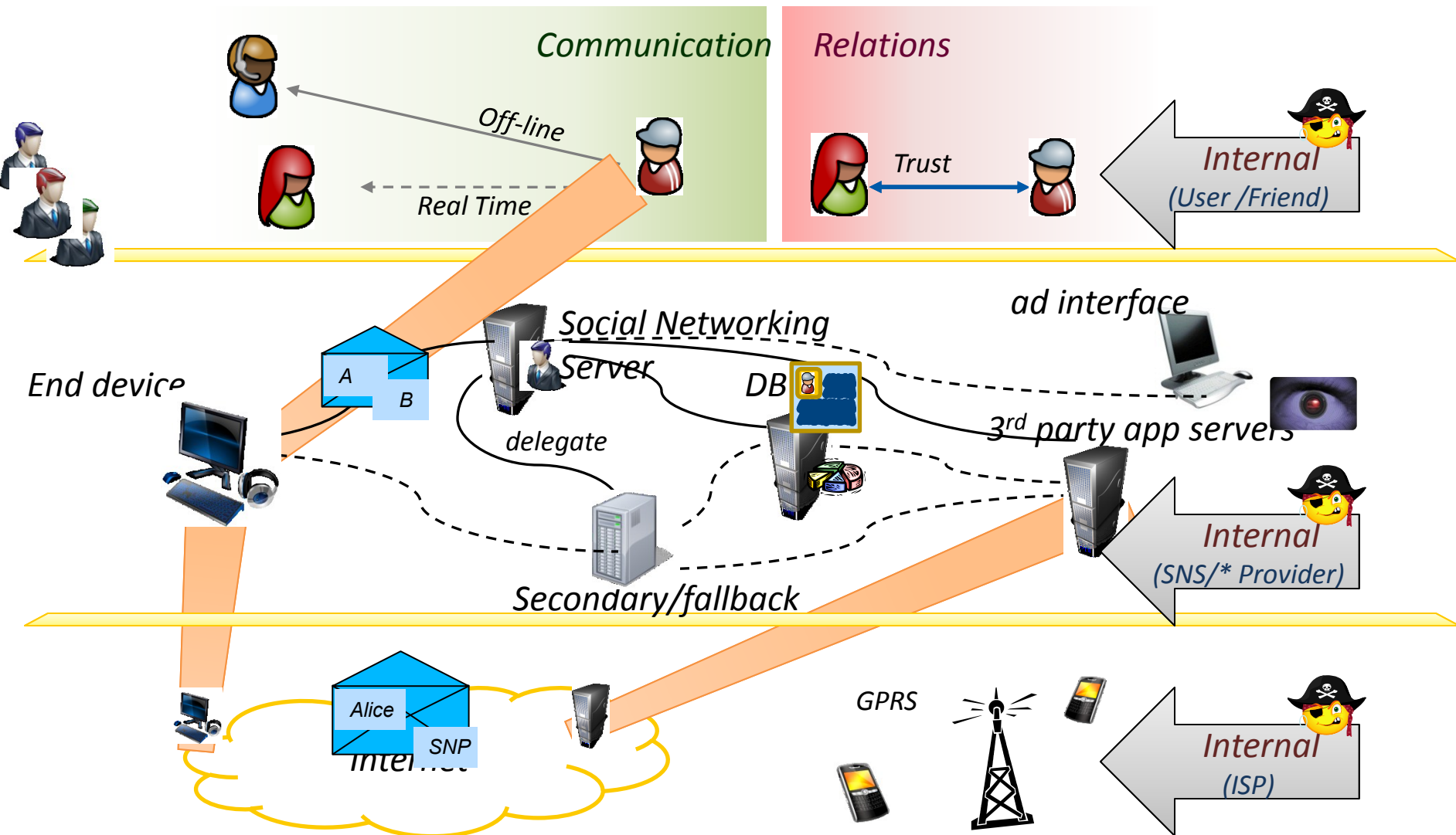


*Everything their subscribers see/write  
(until Nov 21st '12)*

*Ignoring downstream abuse...*



# The Model Revisited: Potential Adversaries



# Privacy Concerns

*Which disclosure are users concerned about?*



# Because They Don't Realize...



- Extent of collected information (data)
- Increasing ubiquity of monitoring
  - Facebook.com
  - -> places/foursquare
  - -> google maps
  - -> browser location sharing

*Loss of benefits*

*Illegal access,  
stalking*

*Cooperation with  
Intel services*

*Price discrimination*

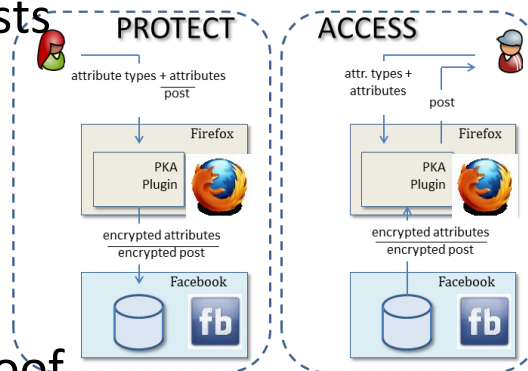


# Immediate Defenses – vs. Crawlers and Confusion



## Crawler Defense [1]

- Access trade-off (oversharing):
  - Crawlers/Adversaries: **deny!**
  - Friends: **grant** w/ low overhead
- Partial-knowledge based Access
- Use attribute sets as credentials
- Shamir shares stored in profile
- Encrypt posts



- Security proof
- Analysis on real world data:
  - Security comparable to passwords
- Firefox extension for download

## FB Privacy Settings [2,3,4]

- Default settings too weak and **complicated** to grasp
- Comprehensible controls:
  - Ease of use, direct feedback

### Über mich

#### Allgemeines

Geschlecht:	Männlich	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
Geburtstag:	22. November 1978	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
Beziehungsstatus:	Verlobt	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
Interessiert an:	Frauen	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
Auf der Suche nach:	Freundschaft Kontakte knüpfen	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
Derzeitiger Wohnort:	Hamburg, germany	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
Heimatsstadt:	Dortmund, Germany	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>

- Implementation: Firefox-Extension
- Comprehensive user studies (TK)
- Download: „Privacy Ampel“ (50k)

[1] WWW '13 [3] HCI WS '11  
[2] SocInfo '12 [4] WWW '12



# Immediate Defenses – Crypto Schemes



## *PMS [1,2]*

- Plain content stored at provider
- Comprehensive AC scheme using crypto
- Simple group management
- Encrypt on attribute level
- Shared key vs. Broadcast Enc.
- Proof of confidentiality, unlinkability/anonymity

## *BroadMask [3]*

- Implementation of PMS
- Chrome/FF extension
  - PBC, Crypto++, GnuPG
  - Picasa plugin
  - Bitmap wrapping



[1] FC/RLCSP '11

[2] LNCS

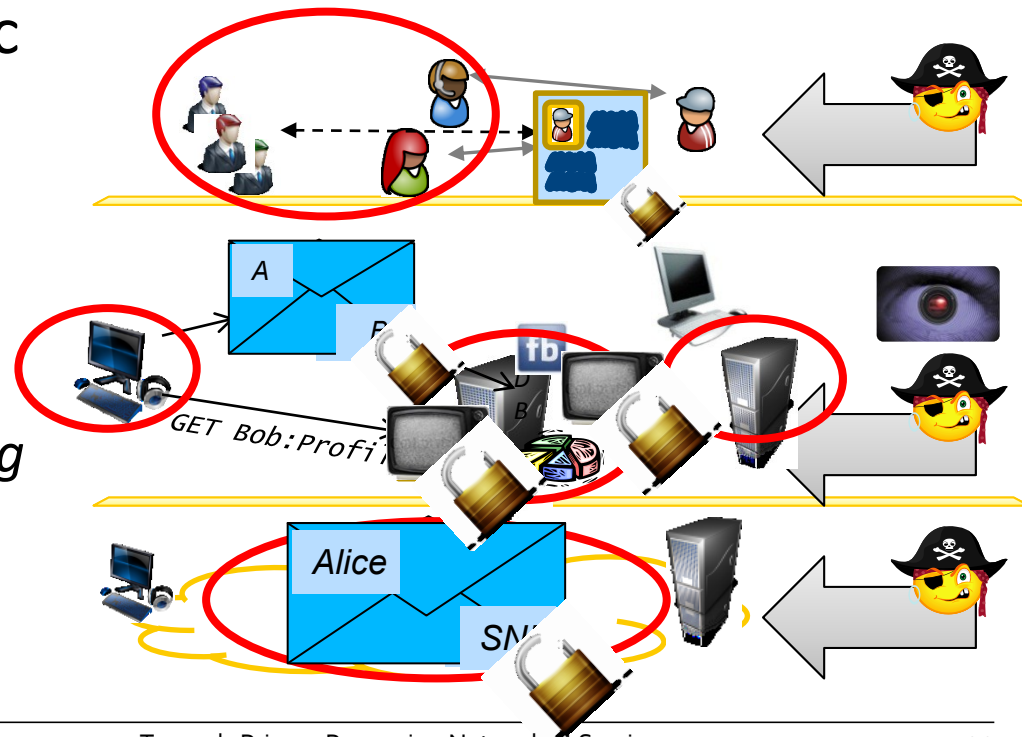
[3] WWW'13

# Potential Solution Classes



- Trust „everybody“
- Suspect Network
  - Transport Layer Security
- Suspect subscribers, public
  - Trust provider (& affiliates)
  - Apply OSN Access Control
- Suspect affiliates/browser
  - *Access abuse, unsolicited msg*
  - *Web security, Sandboxing..*

- Suspect provider & affiliates
  - Aim: Content confidentiality
  - **Crypto Schemes** (Scramble, NOYB)



# What if: you don't want to be observed?



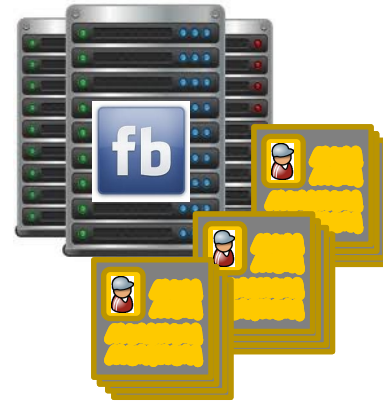
- Provider still has the power to
  - Intercept and drop messages
  - Observe who is active
  - What else the profiles are interested in (cookies)
  - Who is communicating with whom
    - OSN identities
    - Network layer addresses (actual individuals with post addresses)
  - ...

# Potential Adversaries - 2

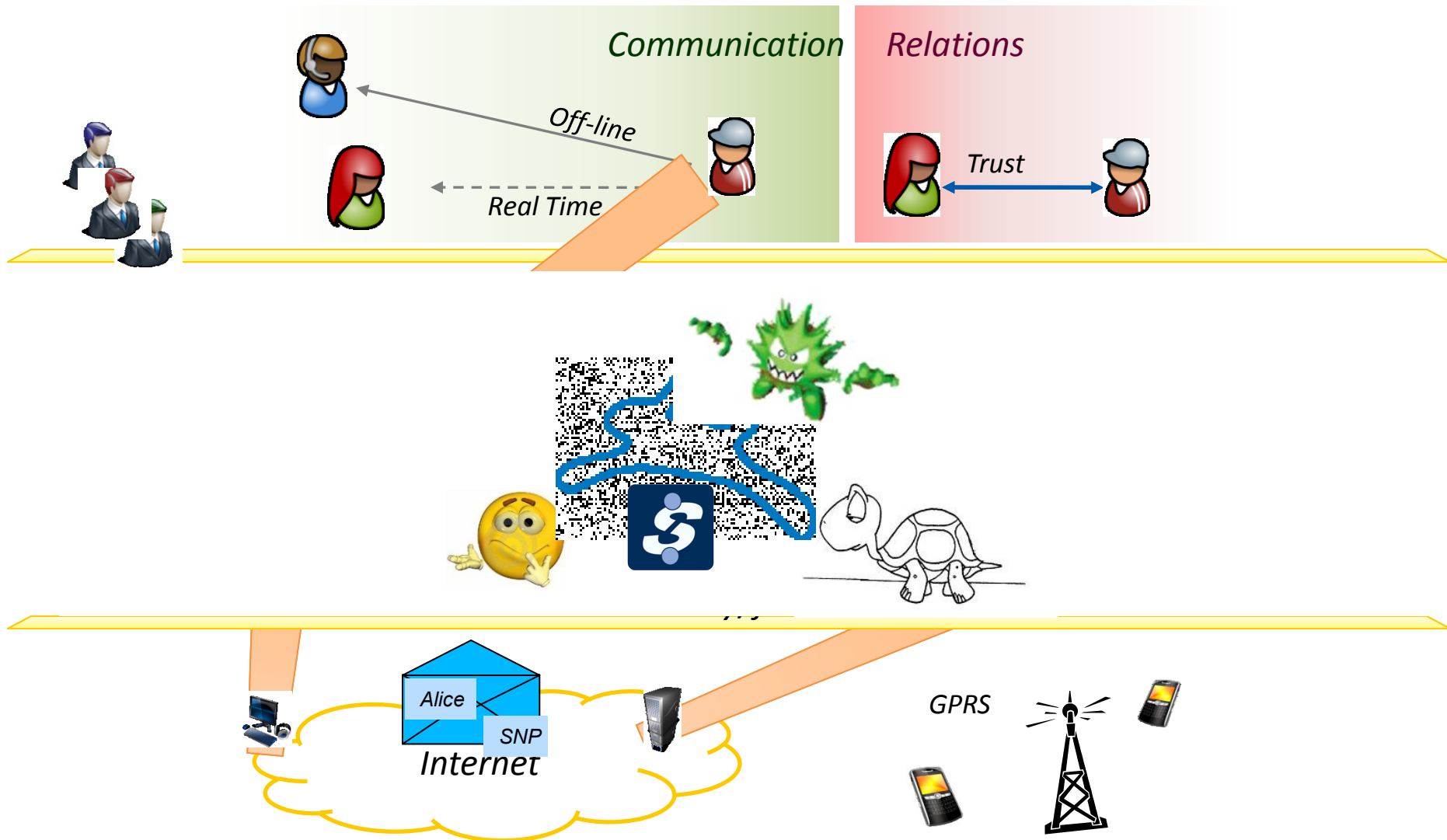
## (and solution classes)



- Suspect provider and affiliates
  - Threat: anonymity, behavior
    - Decentralization
  - Threat: o|
    - Darknets



# The Model Revisited: Potential Adversaries



# Solution Classes Summarized



- Web Security
- Transport Layer Security  
(https)



- **OSN Access Control**
  - *(plus Usability)*
- Web Security 2
  - (Sandboxing)

- **Crypto Schemes**
  - *(with different properties)*



- **Decentralization**



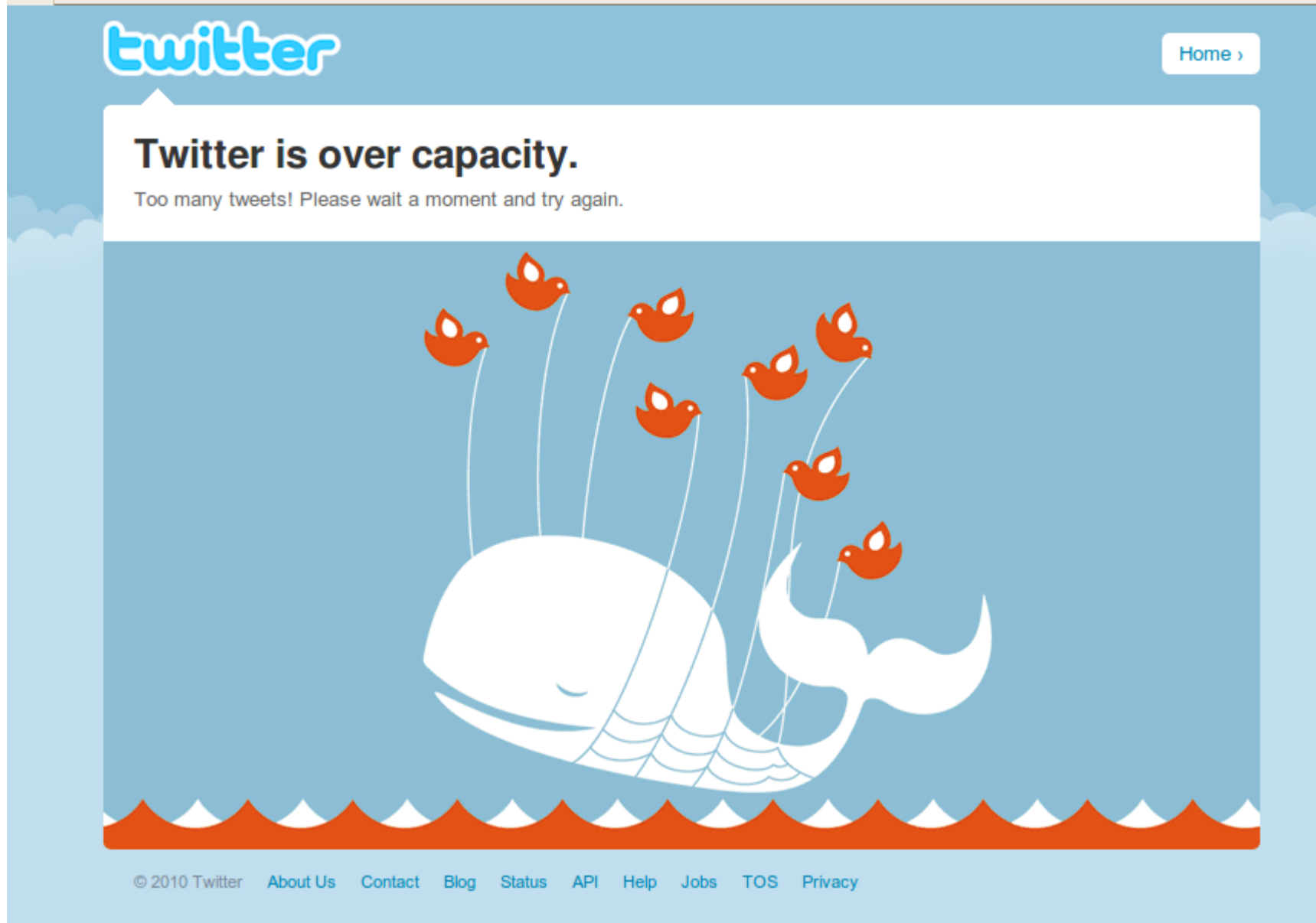
- **Darknets**

- **Awareness!**

- *Then again...*



# Quest to Decentralize







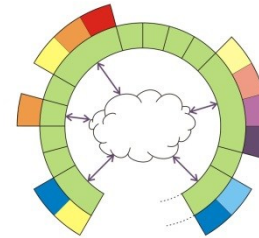
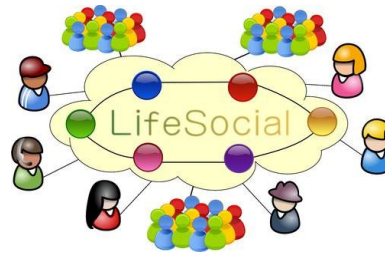
- FOSS
  - BuddyPress, CrabGrass, Cobs, DaisyChain, Diki, Elgg, FETHR, GNUNet, Gossple, Jappix, Lorea, Mycella, Movim, PeerScape, Pinax, StatusNet
- Commercial Approaches
  - diaspora, wuala, LifeSocial
- Academia
  - Friend-of-a-Friend, FriendStore, HelloWorld, LifeSocial, LotusNet (Likir), PeerSon, 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>
<b><i>diaspora</i></b>	web-based	complete	external	Premium
<b><i>FoaF</i></b>	web-based	complete	external	
<b><i>vis-à-vis</i></b>	Cloud	complete	external	
<b><i>LotusNet</i></b>	P2P	split	stand alone	
<b><i>PeerSon</i></b>	P2P	split	external	
<b><i>Safebook</i></b>	P2P	complete	stand alone	Social Coop
<b><i>LifeSocial</i></b>	hybrid	split	stand alone	Premium

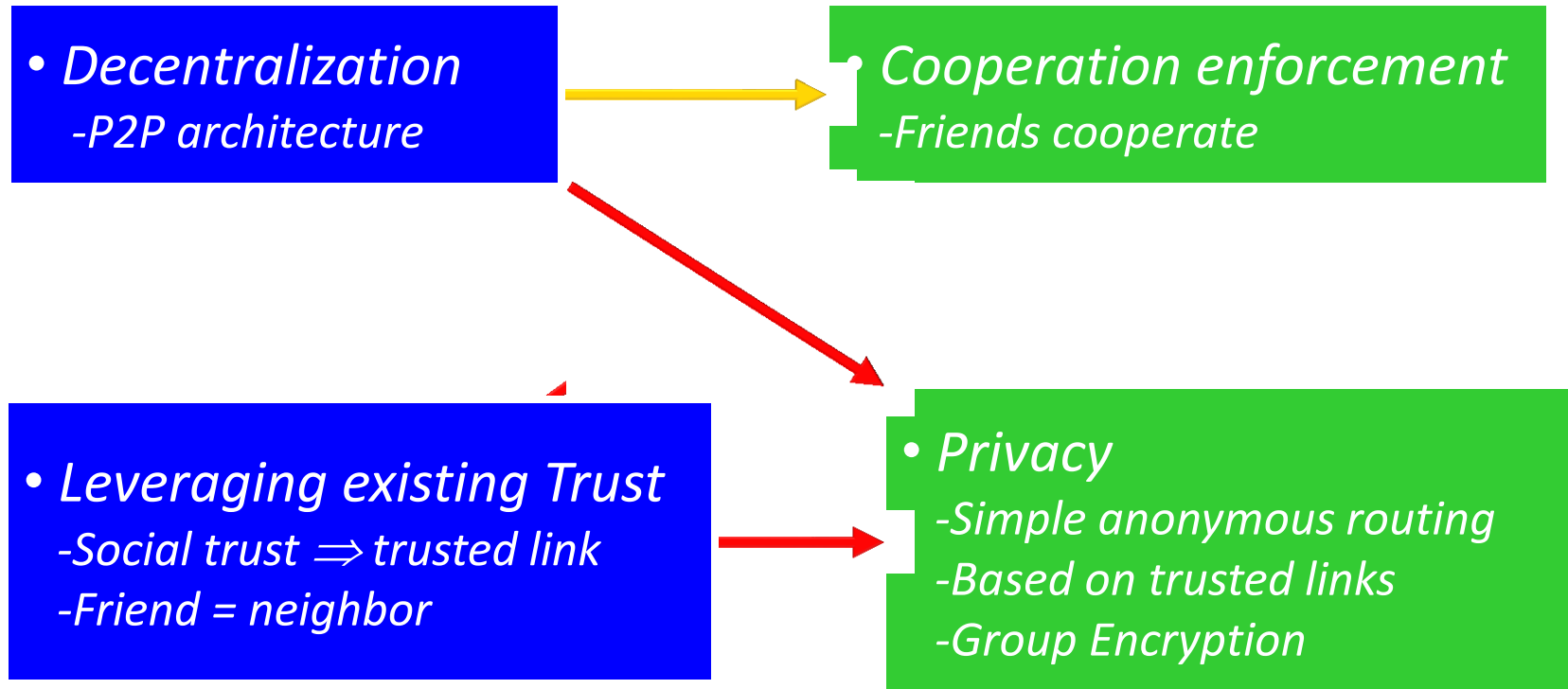


- Let's use P2P to increase privacy in OSN!
- *How could we proceed?*
- *Which challenges are we about to encounter?*
  
- Performance
  - Properties: P2P → filesharing! OSN: small, short lived objects
  - Extreme heterogeneity (friends, uploaded content, resources)
  - User behaviour / sessions (diurnal 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!)

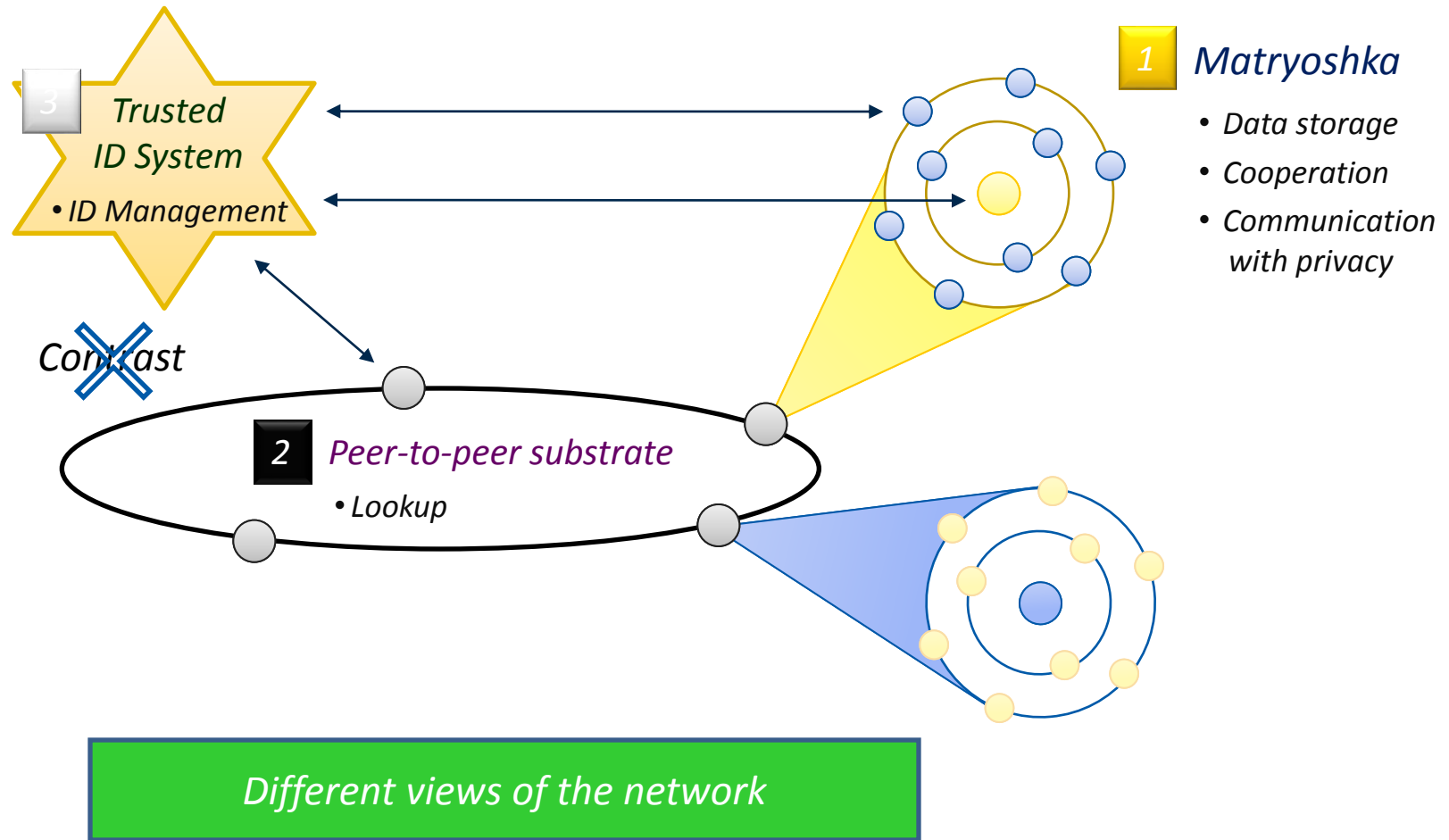


**Safebook** : Privacy-Preserving Online Social Networking

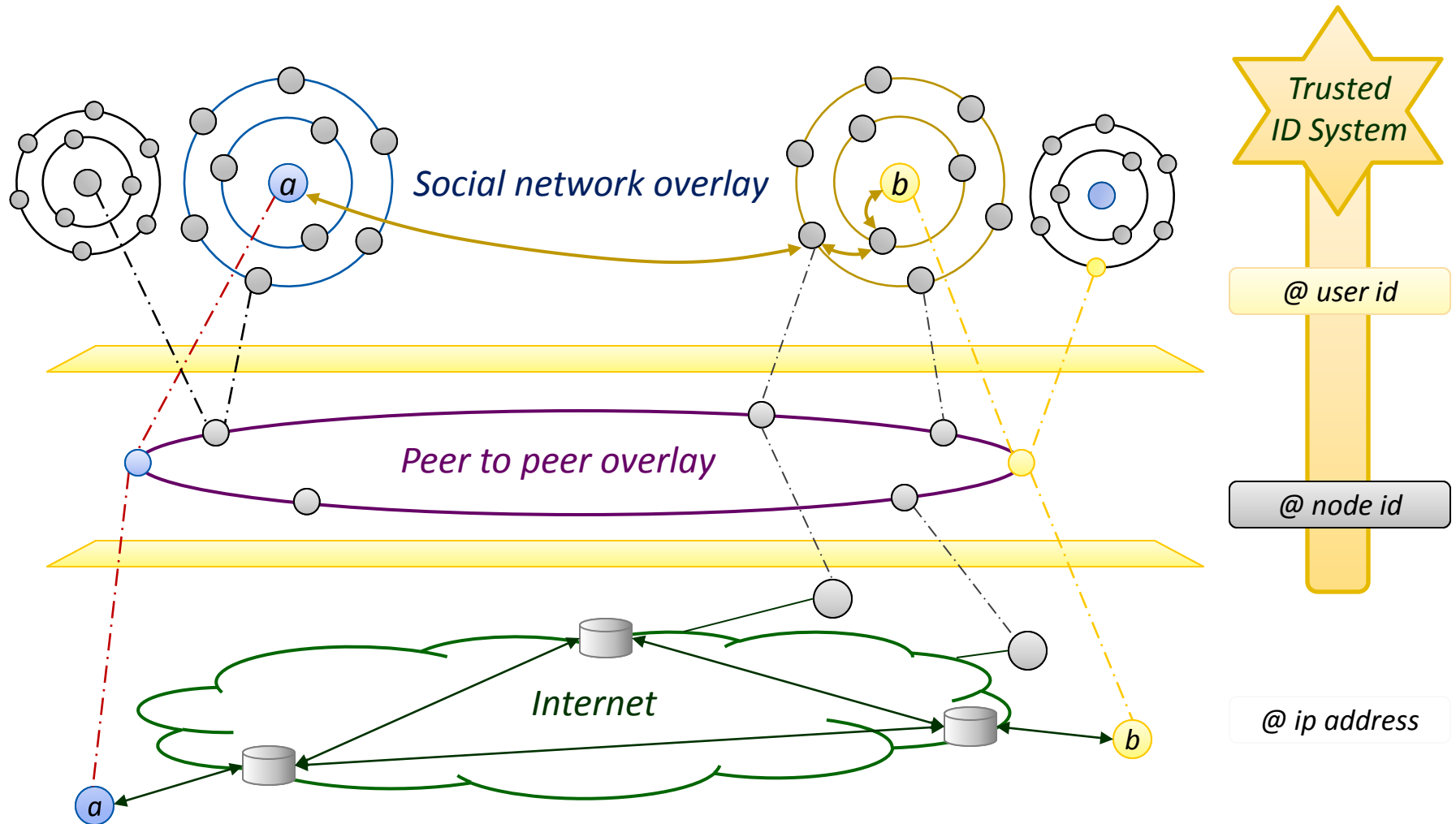
<http://www.safebook.us>



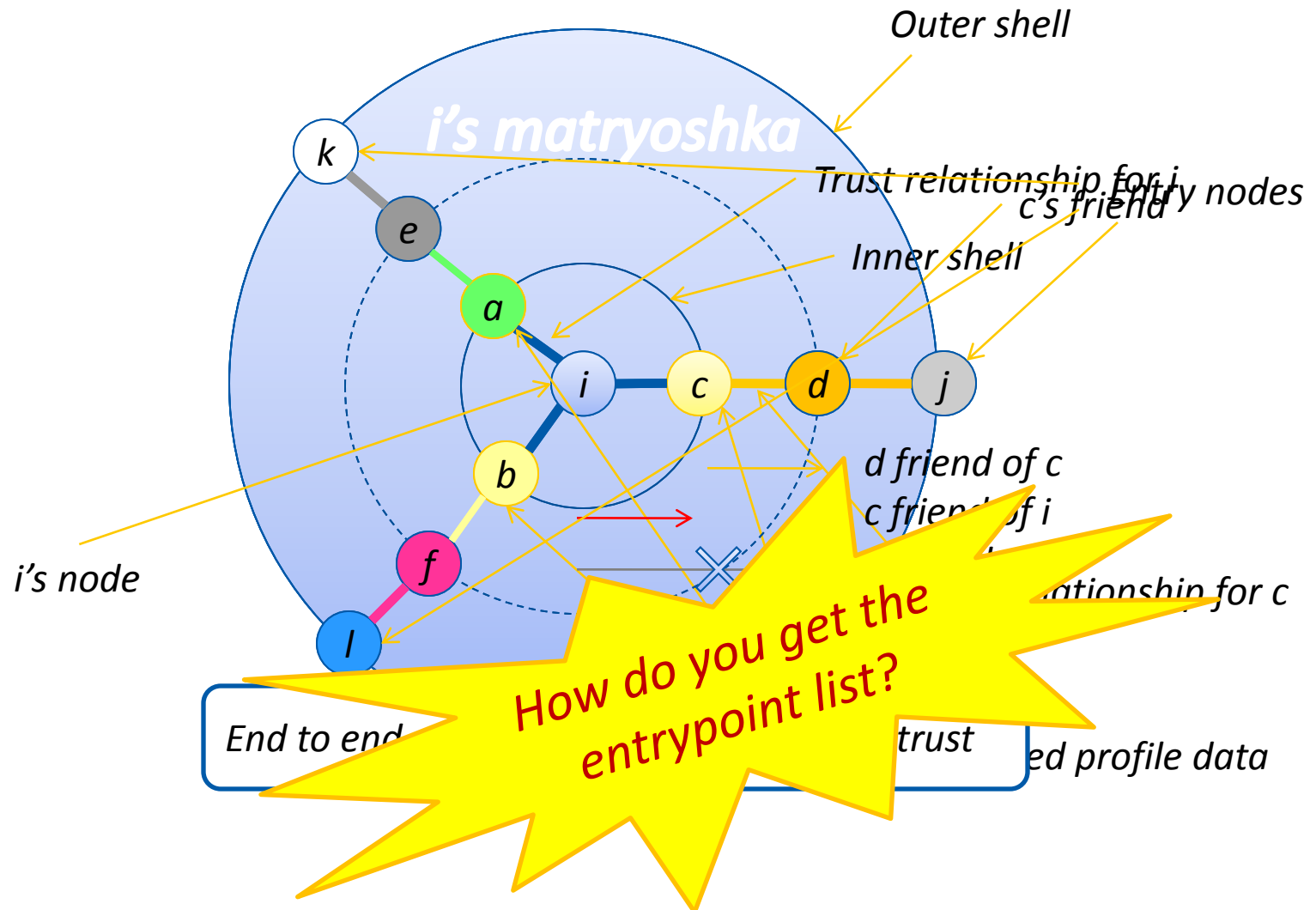




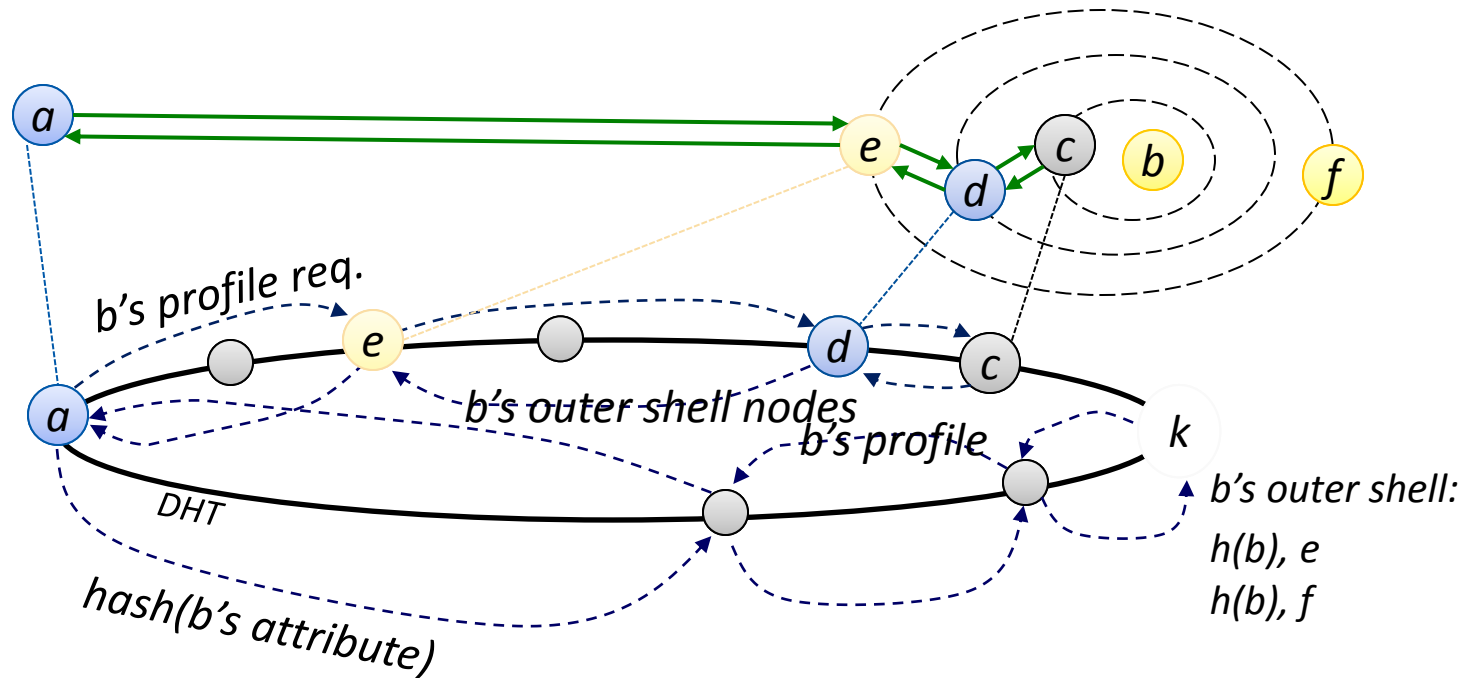
# Network view



# 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

# 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*



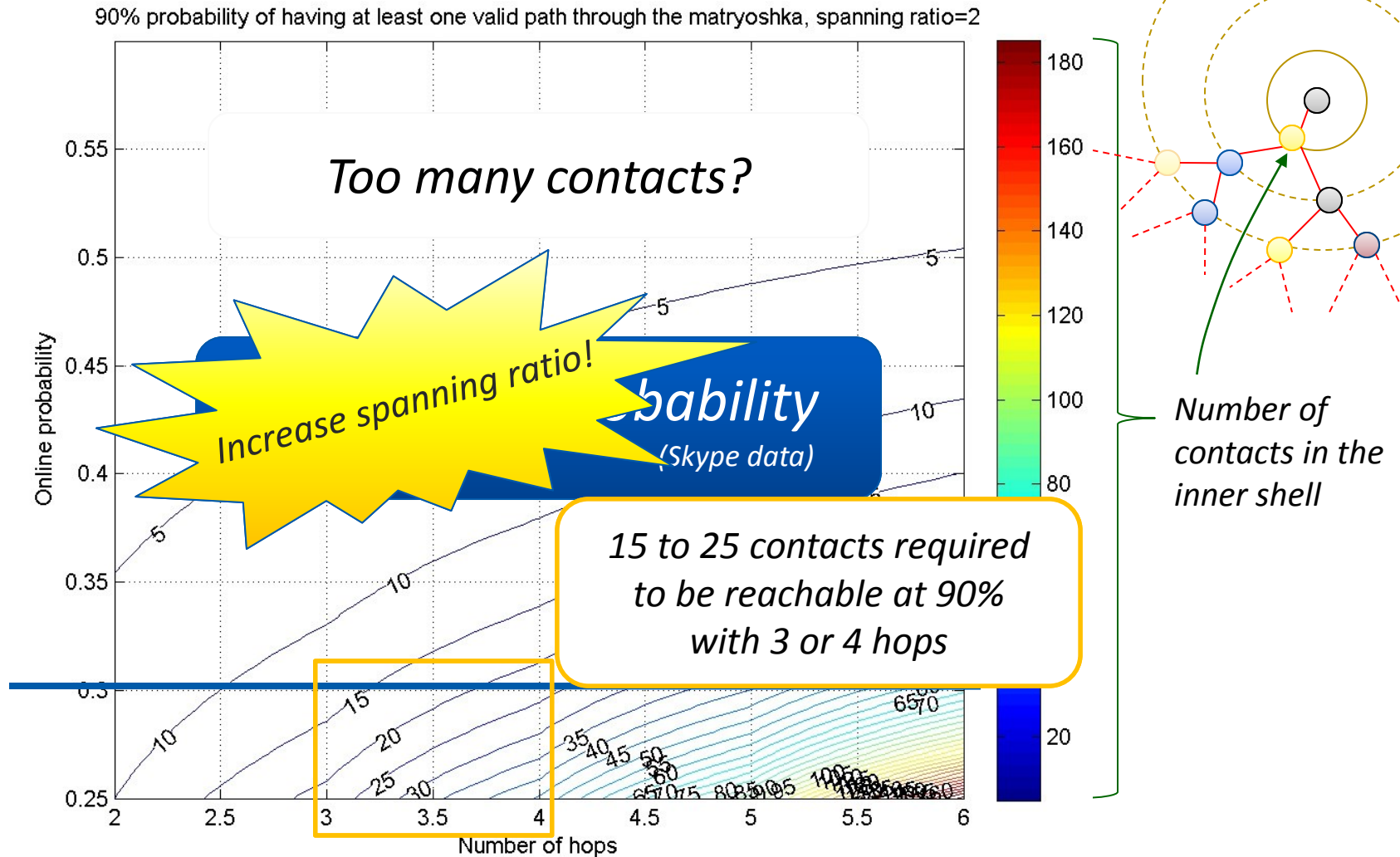
*Key management*

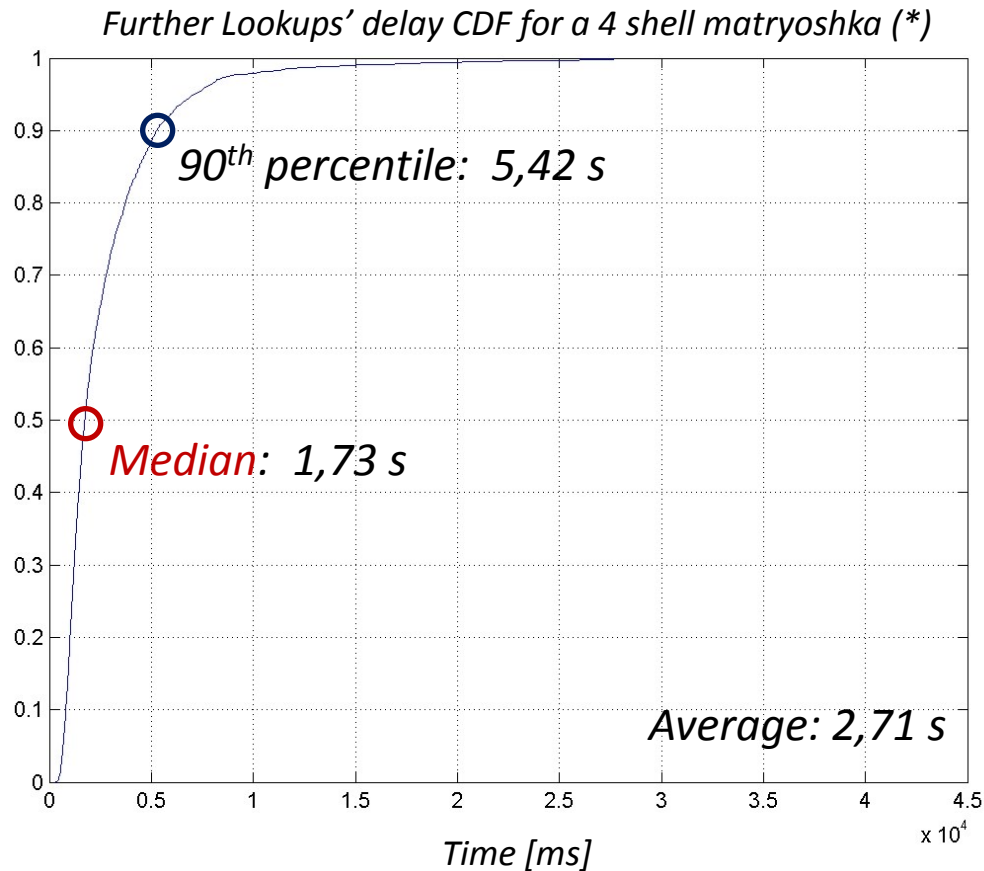
*Availability*



***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..)



# Conclusions for Safebook



- Safebook
  - Decentralized OSN, based on mutual trust
  - Modelled, analysed, simulated and prototype
- Open Challenges
  - **Performance** is insufficient
  - **Availability** questionable (correlated churn)
  - **Concealed participation** impossible

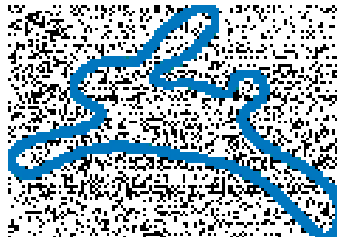
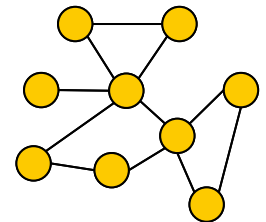
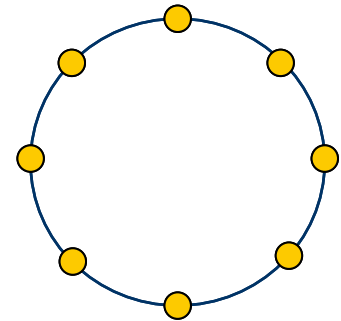


- |                       |                      |
|-----------------------|----------------------|
| [1] WONS 2008         | [5] Eurosys/SNS 2010 |
| [2] WWW 2009          | [6] FC/RLCPS 2011    |
| [3] WoWMoM 2009       | [7] WoWMoM 2011      |
| [4] IEEE CommMag 2010 |                      |

# Bullet Proof Privacy – Social Overlays



- Threat: Observability of behavior and identity
- Aim: Conceal participation and prevent (untrusted) observation
- Communication substrate from trusted links:
  - Social graph defines overlay
  - Restricted connectivity
  - ***Adapt addresses/identifiers*** (without data loss)
  - Estimate routing structure
  - Route on imprecise embedding with (Poly)log path length

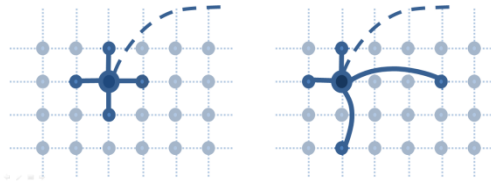


# Enhancing Routing and Embeddings

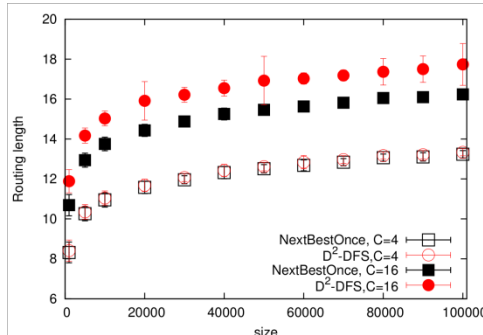


## Darknet Routing [1]

- Darknet model  $K(n, d, C, L)$



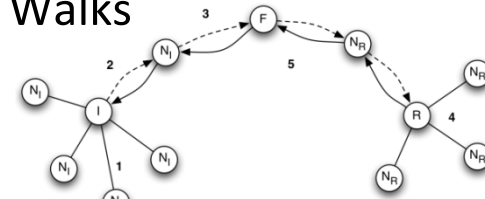
- Freenet routing **not** polylog
- Design of novel routing **NBO**
- Proof of polylog routing length
- Simulate and Implementation



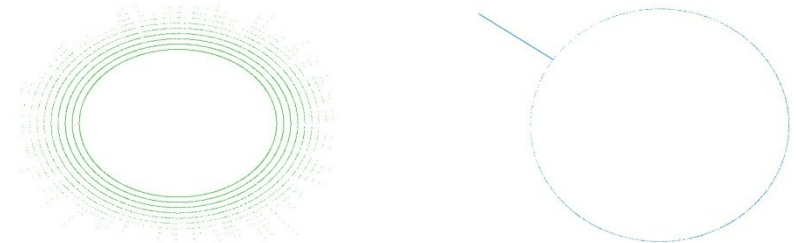
## Attacking Freenet [2]

- Cooperative ID-Adaptation:

- Random Walks
- Swap ID



- Single Adversary Attacks



- Resistant LMC embedding
- Analysis and simulation show superior resistance, performance

- WiP: Novel Tree-based Embeddings[3]

[1] INFOCOM mini '13

[2] SRDS/WNR '11

[3] NetSys '13



- Online Social Networks exhibit sustained, rapid growth
- Decreasing **expertise**, increased **abuse, malfunctions, leaks**
- **Protecting privacy** must be at the focus of CS research
- Important to know user behavior
- Decentralization one possibility
- Large number of projects that leverage the “social” term
- Good (**challenging!**) example for an application that gains from the properties of P2P
- A few real approaches to decentralize
  - Privacy, availability, cost (load balancing)
- Case study: Safebook and its properties