UNIVERSITÄT SIEGEN

Security of Online Social Networks

Privacy

Lehrstuhl IT-Sicherheitsmanagment

Universität Siegen

June 14, 2012

Overview Lesson 08

- Definitions
- Anonymisation
  - Models
  - Communication
  - Database
- OSN and Privacy
  - Network Anonymisation
  - Profile Obfuscation

Motivation

- Privacy in Online Social Networks?
- What is “me” in the Social Network?
  - Pseudonyms/Faces/Nicks/Accounts/Identities
- Who can know what about me?
  - Linkability of Pseudonyms
  - Linkability to “real” person (Anonymity)

Privacy, socially

- Right to be left alone [Warren/Brandeis 1890]
- Informational Self-Determination (Informationelle Selbsbestimmung)

> *"die Befugnis des Einzelnen, grundsätzlich selbst über die Preisgabe und Verwendung seiner persönlichen Daten zu bestimmen"*
>
> *— BVerfGE 65, 1 - Volkszählung (1983)*

> *"Everyone has the right to respect for his private and family life, his home and his correspondence."*
>
> *— ECHR Art. 8 (1)*

Privacy, scientific

- Privacy — undefined
- Anonymity

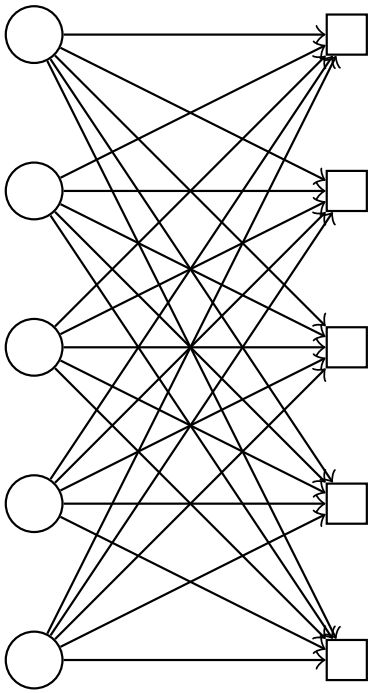
  > *"Anonymity of a subject means that the subject is not identifiable within a set of subjects, the anonymity set."* — *[Pfitzmann and Hansen, 2010]*

- Unlinkability

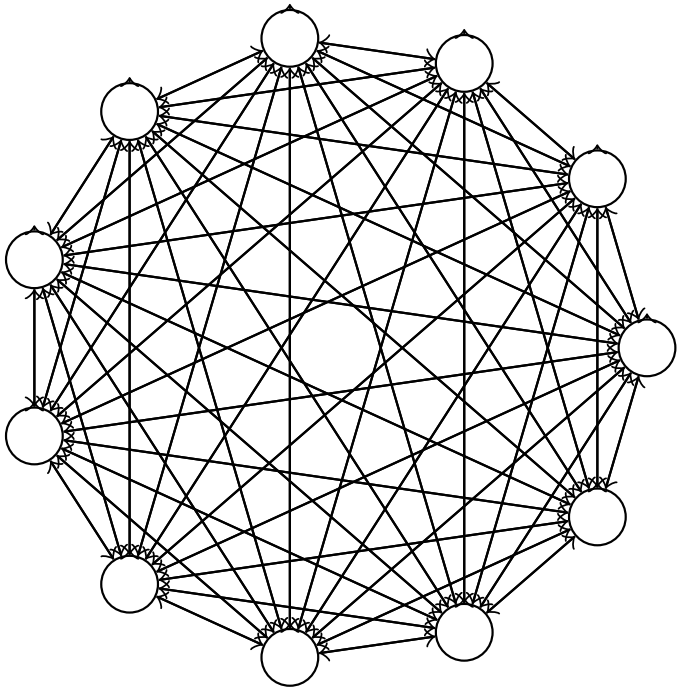
  > *"Unlinkability of two or more IOI from an attacker's perspective means that within the system [. . .], the attacker cannot sufficiently distinguish whether these IOIs are related or not."* — *ibid.*

Anonymity and Unlinkability

Anonymity Problem:

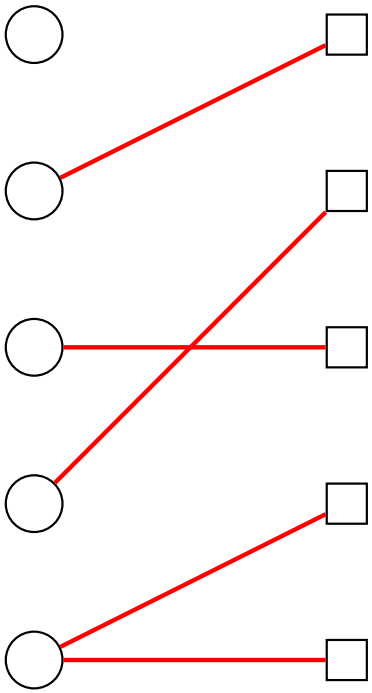


Linkability Problem:

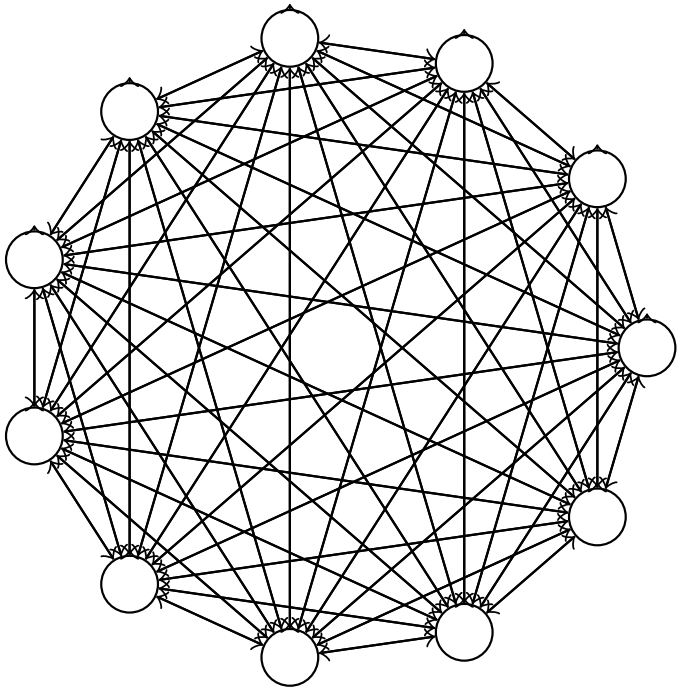


Anonymity and Unlinkability

Anonymity Problem:

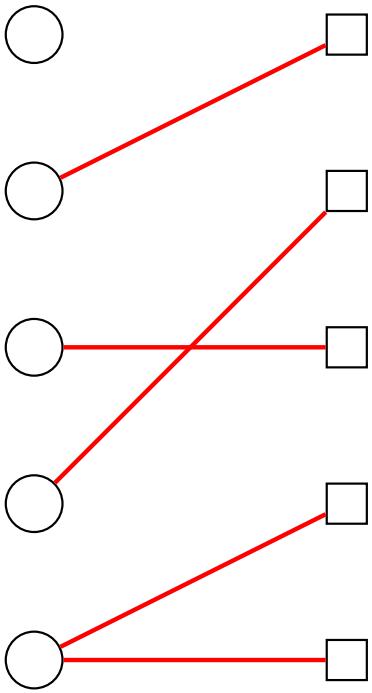


Linkability Problem:

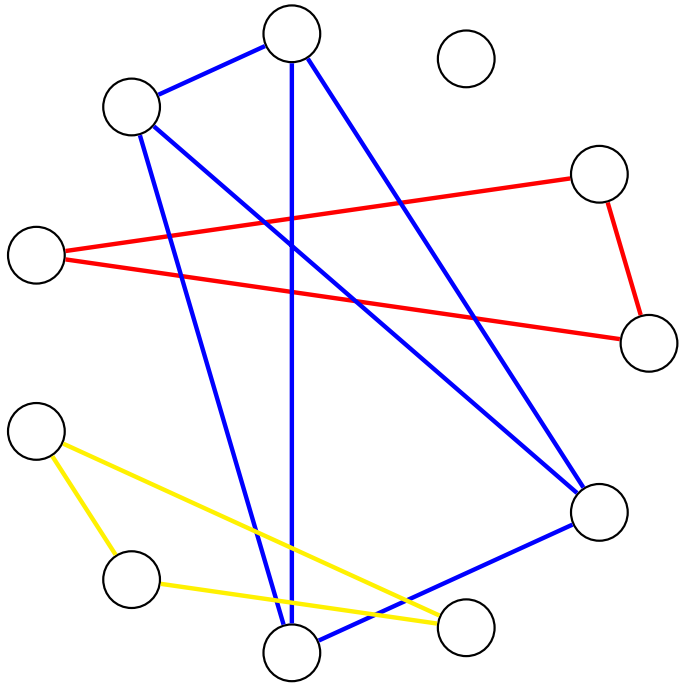


Anonymity and Unlinkability

Anonymity Problem:



Linkability Problem:

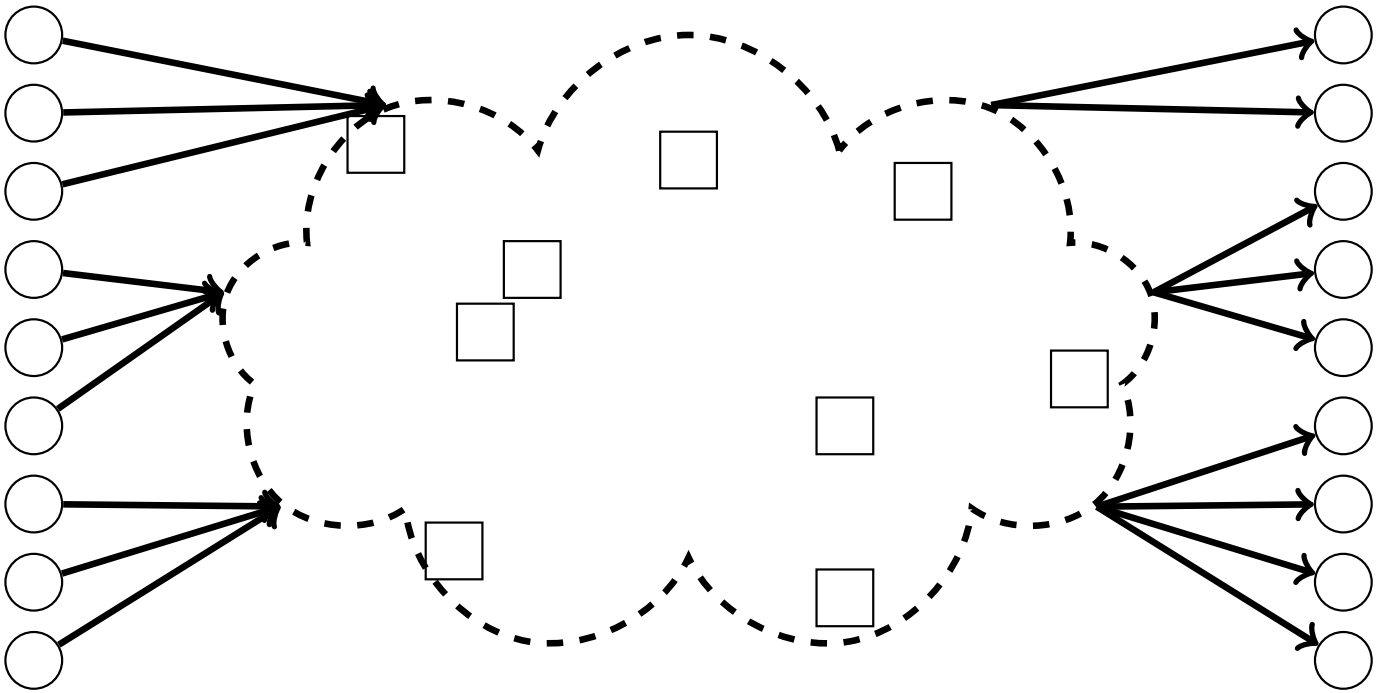


Anonymity Models

- MIX-Communication Model
- Anonymity Set
- PROB-Channel
- Individual Anonymity Degree (IAD)

Communication Model

MIX-Model



(Siehe [Pfitzmann and Hansen, 2010])

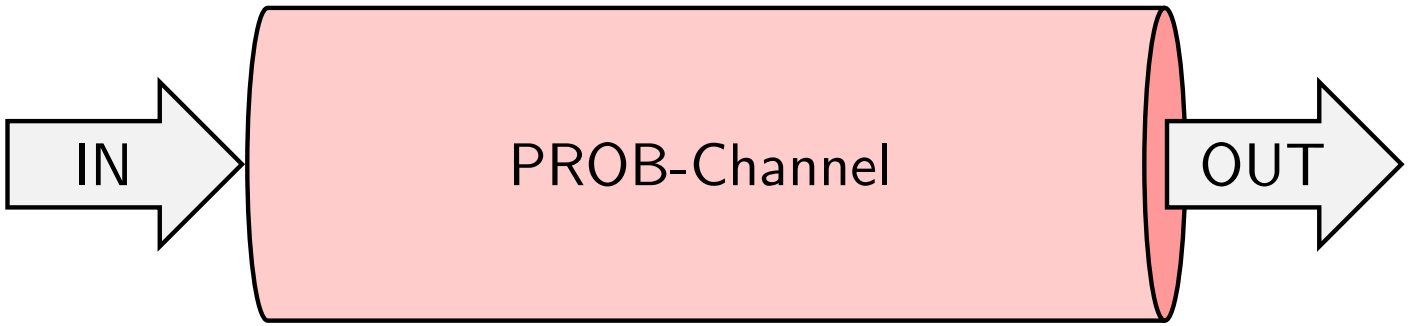
Anonymity Set

Cardinality of Set of possible Senders

> *An anonymity set seen by a set of keys is the set of vertices in a connected component of the graph formed from the original graph by removing the edges concerned.*

(The Dining Cryptographers Problem [Chaum 1988])

PROB-Channel



Residence Time δ

Density of Residence Time: $f(\delta)$: $\sum_{0}^{\infty} f(\delta)d\delta = 1$

Measurement:

IN/OUT	in_0	in_1	...
out_0	$\delta_{0,0}$	$\delta_{0,1}$	...
out_1	$\delta_{nein1,0}$	$\delta_{1,1}$	...
⋮	⋮	⋮	⋱

Attacker's Solution: Minimum Assignment with Respect to

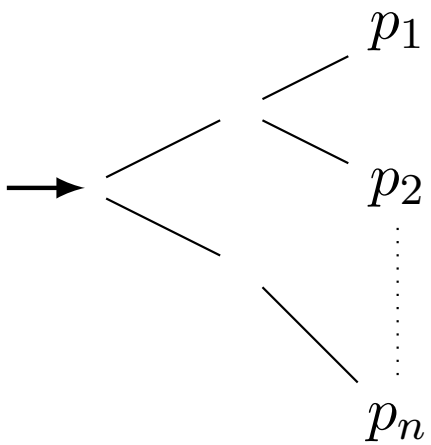
Anonymity of Communication

Objectives

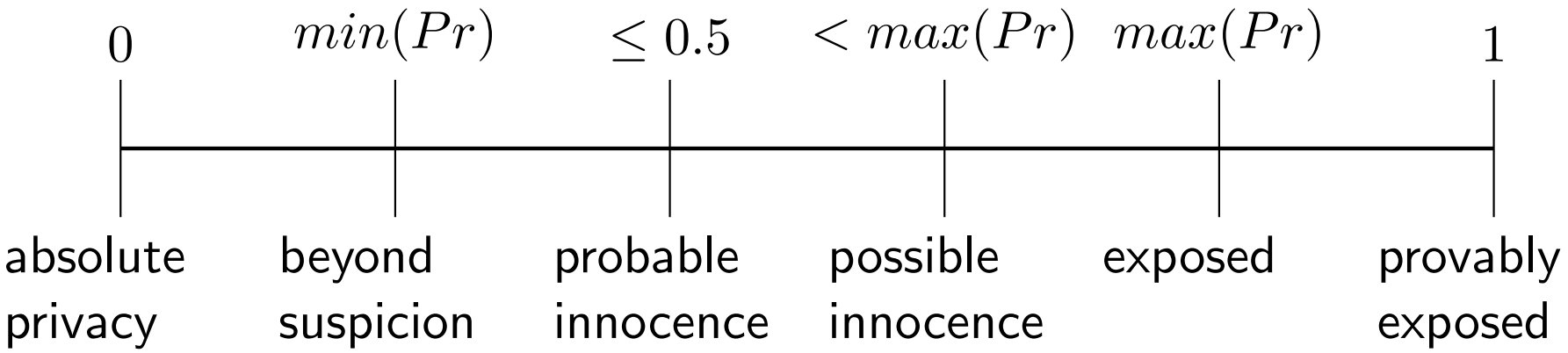
- Sender/Receiver Anonymity
- Message Unlinkability

Anonymity of Communication (1)

- Anonymity Set
- Probability Anonymity Set

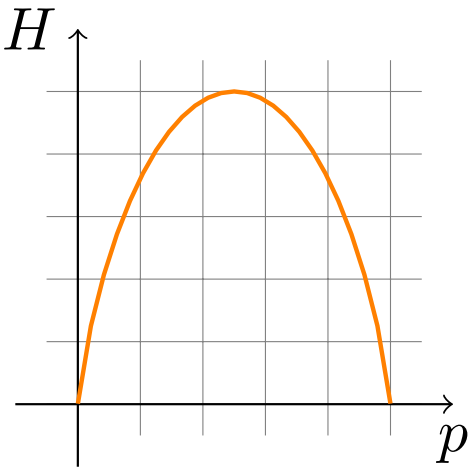


Individual Anonymity Degree (IAD)



(Vgl. [Reiter and Rubin, 1998])

Anonymity of Communication (2)



- Entropy of two Probabilities: $p, (1-p)$

$$H(X) = \sum_{i=1}^{N} p_i \log_2 (1/p_i)$$

(Vgl. [Shannon, 1948], S. 1-12)

- Degree of Anonymity $D(P) = \frac{H(X)}{\log_2(N)}$

(Vgl. [Díaz et al., 2002])

Anonymity of Communication (3)

- Combinatorial Anonymity

  e.g.: How many samples are needed?
  Disclosure Attack
  - Distinct communication partner of Alice
  - potential Recipient/Message
  - Exclusion of Hypotheses by Observation

(Vgl. [Agrawal and Kesdogan, 2003])

- Unlinkability
  - Degree of Unlinkability
  - Expected Distance Unlinkability [Fischer et al., 2008]

Database Privacy

Objectives

- Controlled Information Disclosure
  - Anonymised Database
  - Database Queries

Privacy in Databases

- k-anonymity [Sweeney, 2002]
- l-diversity
- t-closeness [Li et al., 2007]
- ϵ-indistinguisability (Differential Privacy) [Dwork, 2006]
- Procedural Privacy

(See: [Kelly et al., 2008])

Example

1	47677	29	Heart Disease
2	47602	22	Heart Disease
3	47678	27	Heart Disease
4	47905	43	Flu
5	47909	52	Heart Disease
6	47906	47	Cancer
7	47605	30	Heart Disease
8	47673	36	Cancer
9	47607	32	Cancer

(Source: [Li et al., 2007])

Example, anonymised

	ZIP	Age	Diagnosis
1	476**	2*	Heart Disease
2	476**	2*	Heart Disease
3	476**	2*	Heart Disease
4	4790*	≥ 40	Flu
5	4790*	≥ 40	Heart Disease
6	4790*	≥ 40	Cancer
7	476**	3*	Heart Disease
8	476**	3*	Cancer
9	476**	3*	Cancer

k-anonymity $=$ 3

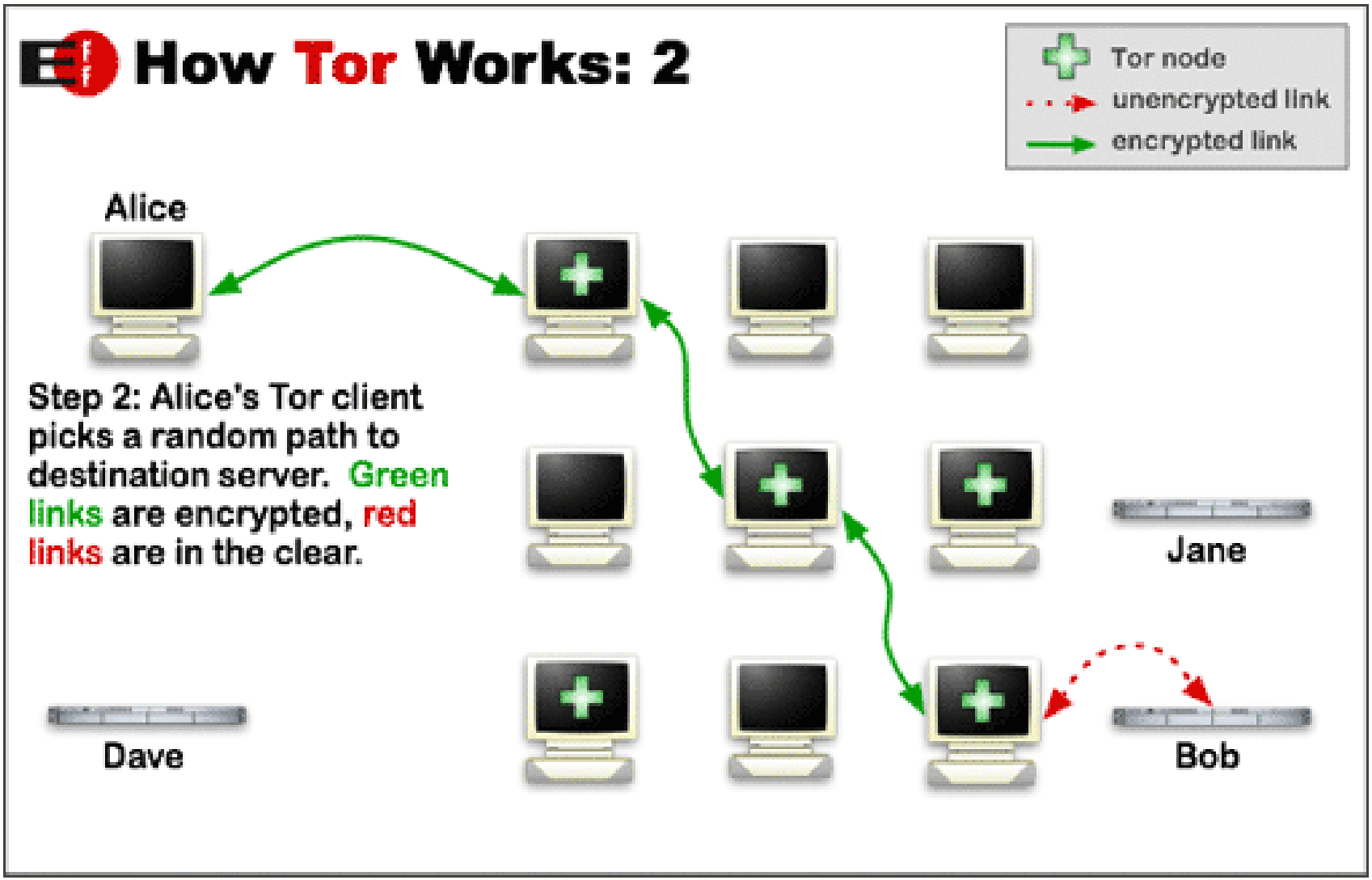
l-diversity $=$ 1

OSN Privacy

- Attacker Model
  - OSN Provider/Database Access
  - Network Observer
  - Web Scraper
  - ...
- Attacker Objectives
  - Link Identities/Uncover Multi-Accounts
  - Reveal Domain-crossing Links
  - Copy/Crawl Database
- But: User want to share data
- Data-Protection vs. Usability

MIX Proxy Networks

- Sender Anonymity
- Redirection of Communication
- Breaking correlation of incoming and outgoing


How Tor Works: 2
Tor node
unencrypted link
encrypted link
Alice
Step 2: Alice's Tor client picks a random path to destination server. Green links are encrypted, red links are in the clear.
Jane
Dave
Bob

MIX-Nodes

- Store-and-forward proxy
- Input: encrypted packets
- Delay packets
- Reorder packets
- Recode (e.g. unify lenght)
-

Profile Obfuscation

- Nondescriptive Identifiers
  - Common Names (not Random Strings)
    - e.g. James Johnson, not cl34all2 unless...
    - But name is rather uninteresting
  - Misleading Fiendship Connections
    - e.g. accept all incoming friends
    - e.g. randomly send friend requests
    - But original graph is subgraph.
  - Fake Attributes
  - e.g. wrong age, hobbies,...
  - But our friends

Profile as Masquerade

- Fool observer/data-miner/clustering
- Creation of “harmless” profile
- Streamlined Profile:
  - Hiding within the masses.
  - Looks like everyone, writes like everyone

Privacy vs. Publication

Question: Why use OSN? And for what?

Literatur I


Agrawal, D. and Kesdogan, D. (2003).
Measuring anonymity: the disclosure attack.
Security & Privacy, IEEE, 1(6):27– 34.


Chaum, D. (1988).
The dining cryptographers problem: Unconditional sender and recipient untraceability.
Journal of Cryptology, 1(1):65–75.


Díaz, C., Seys, S., Claessens, J., and Preneel, B. (2002).
Towards measuring anonymity.
In *Designing Privacy Enhancing Technologies*, LNCS 2482, pages 54–68.

Literatur II


Dwork, C. (2006).
Differential privacy.
In *in ICALP*, pages 1–12. Springer.


Fischer, L., Katzenbeisser, S., and Eckert, C. (2008).
Measuring unlinkability revisited.
In *Seventh ACM Workshop on Privacy in the Electronic Society (WPES '08)*, Alexandria, VA.

Literatur III

Kelly, D. J., Raines, R. A., Grimaila, M. R., and Baldwin, R. O. (2008).
A survey of state-of-the-art in anonymity metrics.
In *Proceedings of the 1st ACM workshop on Network Data Anonymization*, pages 31–40.

Li, N., Li, T., and Venkatasubramanian, S. (2007).
t-closeness: Privacy beyond k-anonymity and -diversity.
In *IEEE 23rd International Conference on Data Engineering*.

Literatur IV

Pfitzmann, A. and Hansen, M. (2010).
A terminology for talking about privacy by data minimization: Anonymity, unlinkability, undetectability, unobservability, pseudonymity, and identity management.
Technical report, TU-Dresden.
v0.34.

Reiter, M. K. and Rubin, A. D. (1998).
Crowds: Anonymity for web transactions.
ACM Transactions on Information and System Security, 1(1):66–92.

Literatur V


Shannon, C. E. (1948).
A mathematical theory of communication.
The Bell System Technical Journal, 27:379–423.


Sweeney, L. (2002).
k-anonymity: a model for protecting privacy.
Journal on Uncertainty, 10(5):557–570.


Warren, S. D. and Brandeis, L. D. (1890).
The right to privacy.
Harvard Law Review, 4(5):193–220.