UNIVERSITY OF AGDER

Privacy-Enhanced Network Monitoring

Nils Ulltveit-Moe
PhD student
University of Agder

Winter School, Finse, 5 May 2009

Funded by Telenor R&I (contract DR-2009-1)

Contents

- Introduction to privacy
- Privacy-preserving network monitoring solutions
- Building blocks for future research

Emphasis of this talk.

- Aim: present a set of techniques that *may* be useful for *privacy-enhanced network monitoring*.
- How can we improve privacy without loosing too much efficiency and usability?
- The main focus is on intrusion detection systems (IDS).
- Some examples on traffic analysis systems.
- The focus is *not* on how to reduce the false positive rate.
  - (Although smaller FP rate would make privacy handling easier.)

How do we perceive privacy?

- Bounded rationality limits our ability to exhaustively search for the best alternative:
  - Framing of a question influences our reaction to it;
  - Heuristics often replace rational searches for the best possible alternative.
- Biases and other anomalities affect the way we compare alternatives, perceive risks or discount values over time:
  - Valence effect - overestimate the likelihood of favourable events;
  - Overconfidence;
  - Rational ignorance;
  - Status quo-bias;
  - Reciprocity and fairness;
  - Inequity aversion.

What is required for a useful Privacy-Enhanced Technology (PET) system?

- **Usability:** Users must be able and want to use it.
- **Deployability:** Easy to install, works on your platform.
- **Effectiveness:** PET must work properly to be useful.
- **Robustness:** Privacy protected even if system is compromised.

PET categorisation

- **Protection level:** type of anonymity - sender, recipient or relationship.
- **Security level:** information theoretic/unconditional or cryptographic/computational security.
- **Attacker model:** Protects against outsiders, participants, network insiders?
- **Trust model:** Who does the user trust? network providers, participants?

Privacy-preserving cryptographic protocols

- **Definition:** A protocol is *privacy-preserving* if it reveals only the result of the collaboration and what can be deduced from this result when given a group of participant's inputs.
- The term privacy is here used for both individuals and other entities, although confidentiality is more suited for the latter.

Adversary models

1) **Honest-but-curious:** an adversary follows the prescribed protocol exactly, but after the protocol is finished, the adversary will try to learn additional information by using its local transcript of the protocol's execution.
2) **Malicious adversary:** the participants deviate arbitrarily from the prescribed protocol to gain advantage.
3) **Rational-selfish:** participants will maximise their expected utility whether it means following the protocol or deviating from it.

2) and 3) may be considered Byzantine adversaries.

Contents

- Introduction to privacy
- Privacy-preserving network monitoring solutions
- Building blocks for future research

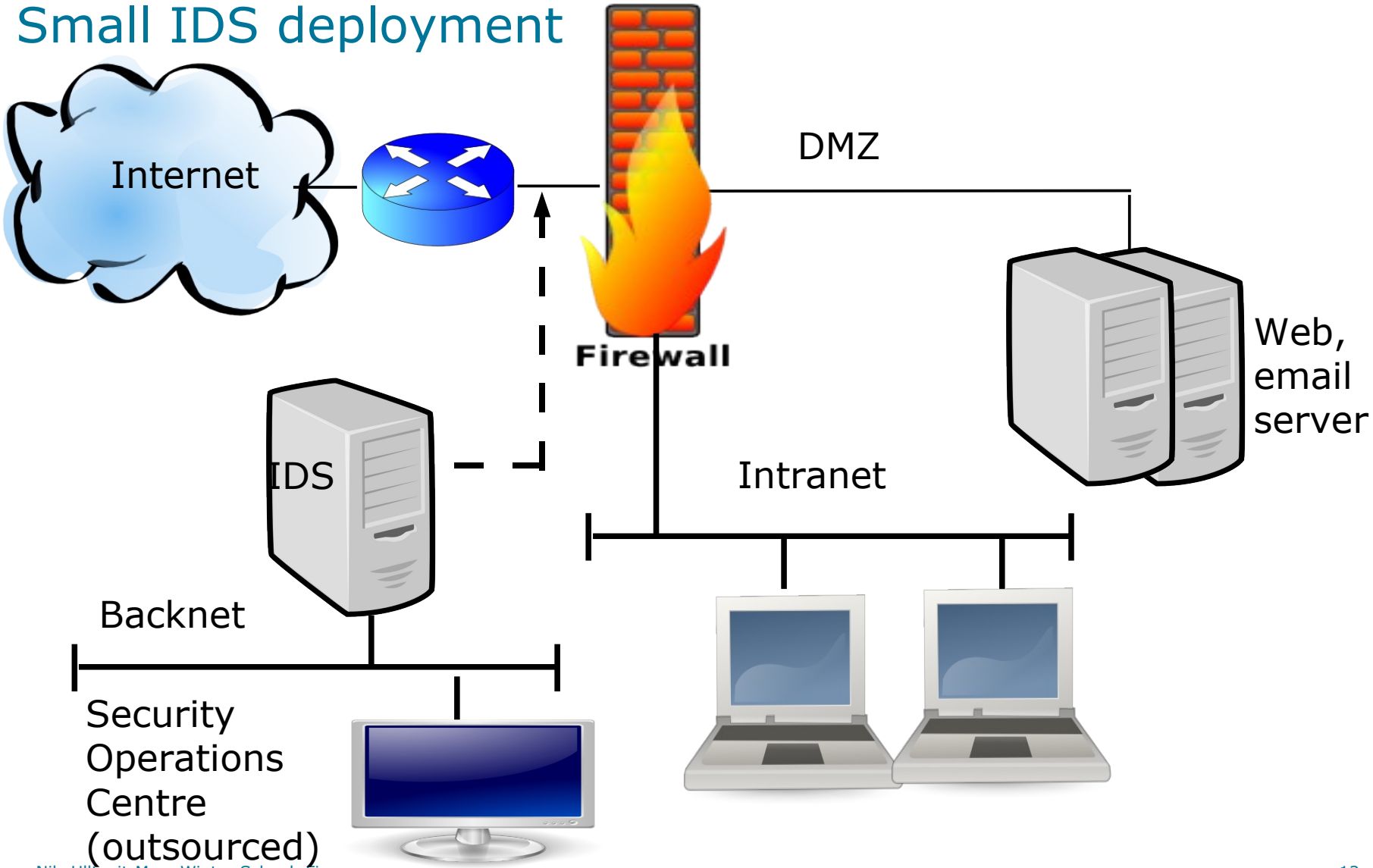
Network monitoring trends

- Examples:
  - Traffic accounting;
  - Intrusion Dection Systems (IDS).
- Monitoring proliferates, gets more efficient and invasive:
  - Cleartext monitoring of SSL traffic possible under given circumstances;
  - Time machines allow for retrospective IDS analysis - data mining of network traffic back in time.
- Some monitoring is even required by law...
- What about *privacy!*?

How is intrusion detection being performed?

- Intrusion Detection Systems (IDSs) - the Internet equivalent of a burglar alarm. Monitoring is performed using deep packet inspection, which means that the following data can be investigated:
  - Packet header information;
  - Payload in each data packet;
  - Reassembled streams of data spanning several data packets;
  - Entire communication sessions between a client machine and a server.
- If a presumed malicious event is detected, an alert is sent.

Small IDS deployment



Privacy-Preserving Intrusion Detection Systems

- A privacy-respecting intrusion detection system has been proposed by Ulrich Flegel:
  - Supports multilateral security (both privacy and security);
  - Separate privacy and security controls;
  - Privacy-sensitive events pseudonymised in separate protocol layer;
  - Pseudonymisation based on threshold cryptography (Both Shamir and Karnin, Greene and Hellman's schemes used);
  - Uses rule language based on Coloured Petri Nets (CPN), called Serial Signature Nets;
  - Pseudonymisation directly derived from IDS rules;
  - Overhead, since both IDS and pseudonymiser evaluates IDS rules;
  - Only partially implemented as far as I know.

Privacy-Preserving Intrusion Detection System (cont)

- An early solution based on Mixes was proposed by Büschkes and Kesdogan.
- BRO:
  - Full-fledged IDS system;
  - Compatible with Snort rules;
  - Time Machine interface (Maier et al.);
  - Unilateral hash-based audit trace anonymisation;

Traffic accounting

- Large-scale Monitoring of Broadband Internet Infrastructure (LOBSTER) and the former SCAMPI projects:
  - Comprehensive Anonymisation API (AAPI);
  - Based on virtual organisation groups;
  - Can do transport layer stream reassembly;
  - Policy for tapping all payload intended for anomaly-based IDS;
  - Many partners, including Uninett.
- Other tools:
  - Transaction-specific pseudonymisation for IP traffic accounting applications (Øverlier et al);
  - tcpdpriv, ip2anonip only work on network layer;
  - PRISM project has proposed a privacy-preserving IPFIX mediator.

Contents

- Introduction to privacy
- Privacy-preserving network monitoring solutions
- Building blocks for future research

Homomorphic encryption (e.g. Pailier or ElGamal)

- Enables addition/multiplication of plaintext values using encrypted values. For example:

$$E(a)E(b)=E(a+b)$$

$$E(a)^c=E(ac)\ for\ c\in N$$

- Enables oblivious polynomial evaluation (OPE).
- Useful in design of provably secure private protocols.
- Possible to re-encrypt a ciphertext to generate another ciphertext with the same plaintext value.
- However, it can be resource demanding...

Computing with encrypted values

- Example: Private scalar product
  - Alice vector $\underline{a}$, Bob vector $\underline{b}$.
  - Alice chooses a homomorphic encryption scheme, gives Bob: $E_{pk}(a_1),\ldots,E_{pk}(a_n)$
  - Bob computes: $\prod_{i=1}^{n} E_{pk}(a_i)_i^b$
  - which is equivalent to: $E_{pk}(\sum_{i=1}^{n}(a_i * b_i))$
  - Alice decrypts value from Bob and learns the scalar product.

Input quality problem

- Problem with privacy-preserving protocols: temptation to gain advantage by lying.
- Solutions:

  1) Design interaction *incentive compatible* so that B cannot gain anything through a lie.
     - Example: *First-price auction* is not incentive compatible, but *Vickrey auction* (as eBay uses) is. This means that no participant can *decrease* what they pay by bidding a value that differs from what the item truly is worth to them.

  2) Use third-party CA for access control, trust negotiations, credit checking, where being untruthful may secure the desired access.

Traffic analysis

- Diffie and Landau: "*Traffic analysis, not cryptoanalysis is the backbone of communications intelligence.*"
- Traffic data records:
  - time and duration of a communication;
  - traffic analysis examins the data to determine shape of the communication streams, the identities of the communicating parties and locations.
- What is data from the EC data retention directive for?

Intrusion detection as Privacy-Preserving Profiling

- Real life case: Airport Security Monitoring for adversaries using Computer Assisted Passenger Prescreening System (CAPPS II).
- Network monitoring analogy: Attacker profiling using Intrusion Detection Systems:
  - profile matching - a classification task;
  - profiles define rules, when matched, describe a behaviour class (target);
  - privacy-preserving profiling possible via cryptographic means;
  - efficient solution requires some trust.

Intrusion detection as Privacy-Preserving Profiling (cont)

- Key tool used: Commutative Encryption:
  - order of encryption and decryption does not matter.
- Two persons can create their own key, encrypt their own secret value and then encrypt the encrypted value of the other person. If the double encrypted values are the same, the input is the same.
- Commutative schemes: Pohlig and Hellman, ElGamal

Privacy-preserving techniques in data mining

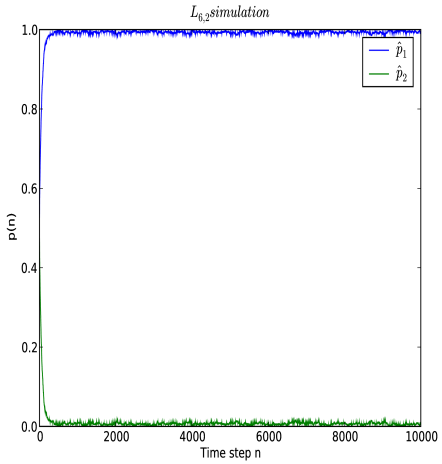
- Technical solutions to avoid misuse of information released to others:
  1) Protect personal information - do not sacrifice privacy of individuals if doing so would not improve security;
  2) Protect sensitive information;
  3) Enable collaboration among different competing units.
- Privacy-preserving techniques can allow the units to determine security relevant outcomes based on joint data without revealing its data to any other party.
- Applicable for database driven network monitoring systems.

Privacy-preserving techniques in data mining (cont)

- Made possible via cryptographic methods for secure computations.
- More secure, but less efficient than other privacy-enhancing data mining methods like e.g. data perturbation.
- Characterised by formal proofs of security.
- For example secure decision tree classification over horizontally partitioned data.

Privacy-preserving inductive learning and IDS

- Supervised learning:
  - Similar to discriminate analysis in statistics.
- Unsupervised learning:
  - Similar to cluster analysis in statistics.
- Machine learning:
  - Learning process automated;
  - Reinforcement learning, learning with teacher etc.
- Model can be used to predict outcome of future unknown situations.
- Applicable for example to anomaly-based IDS or decision support systems built on top of IDSs.



Private K-means or document clustering useful for IDS services?

- Data mining technique: partition *n* observations into *k* clusters.
- Each observation belongs to the cluster with the nearest mean.
- Security requirement: *values* assiociated with an entity is private, but *existence* of an entity may be revealed.
- Privacy-preserving k-means and document clustering solved for two parties.
- Perhaps useful in an Anomaly-based Intrusion Detection service where sensors cooperate on detecting?

k-anonymity, l-diversity

- Is may be useful from a privacy perspective to consider the anonymity set, if network monitoring is outsourced to third-party organisations.
- *Indistinguishability:* Individual hidden in a crowd of individuals with similar/same private values. (for example k-anonymity)
- *Uncertainty:* If all individuals have the same salary value, they are still identified:
  - → uncertainty/diversity measure needed (for example l-diversity).
- Research challenge: Practical methods for doing this...

Privacy Policies for network monitoring systems

- Purpose: define by whom, for which purposes and in which way collected data can be accessed:
  - Can impose obligations on the organisation using the data;
  - Formalises privacy statements;
  - Allows customer preferences;
  - Closely resemble traditional access control policies augmented with privacy-specific characteristics such as purposes, conditions and obligations.
- Examples:
  - OASIS eXtensible Access Control Markup Language (XACML)
  - W3C Platform for Privacy Preferences (P3P)
  - IBM's Enterprise Privacy Authorization Language (EPAL)

Cryptographic Obfuscation to enforce privacy policies?

- Obfuscation: hiding information from plain sight inside computer code or digital data.
  - Diffie and Hellmans-76 - first paper to describe software obfuscation.
  - DH suggested that making the encryption program incomprehensible may be a good way of converting a symmetric cryptosystem into a public key one.
  - White-box cryptography: remains secure even if the program were executed on a computer completely controlled by the adversary.
  - Hard for an adversary to invert encryption function or extract symmetric keys from it.
  - DRM is the most common application for white-box cryptography.

Cryptographic Obfuscation (cont)

- Barak et al's seminal paper (2001) *"On the (Im)possibility of Obfuscating Programs"*
- For a program *P* that we want to obfuscate, an ideal function I_p is defined as a black box that has the same input/output behaviour as *P*.
- I_p is perfectly secure, because it hides everything about the internals of *P*.
- This is an abstraction - how *P* would behave if implemented in tamper-proof hardware.
- Example: hash functions can be regarded as obfuscators.

Obfuscation for access control and data privacy on IDS?

- **Scenario:** Data owner wants to outsource network monitoring to third party.
- Instead of hiding individual data entries - obfuscate the event database, so that only certain queries can be evaluated on it.
- **Goal:** only possible to access the database in ways permitted by the privacy policy.
- **Challenge:** tradeoff between privacy and utility.

Risk model for privacy insurance useful for IDS?

- What risks does a business experience from privacy incidents caused by outsourced network monitoring?
- Can be modeled using an actuarial random utility model.
- User satisfaction in state 0 (status quo) or 1 (privacy lost):

$$u_{1,j}(y,z)+\epsilon_{1,j} \qquad u_{0,j}(y,z)+\epsilon_{0,j}$$

- where *y* is income/wealth and *z* is individual characteristics
- Probability can be calculated if distribution ε is known.
- Possible measure, if the aim is to reduce the cost of privacy violations to a minimum.
- For example privacy impact assessments of company.

Conclusion

- There exists some former work in the area of privacy-preserving network monitoring systems.
- However area does not appear very mature.
- Cryptographic methods useful building blocks.
- Need anonymisation or pseudonymisation of sensitive data
- The main challenge is that practical systems need to strike a good balance between security, privacy and efficiency.
- This is a hard problem to solve.

Thank you!

Questions?
Comments?
Good ideas?

Support slides

Terms

- **Sender anonymity:** sender anonymous within a set of potential senders.
- **Recipient anonymity:** recipient anonymous within a set of potential recipients.
- **Relationship anonymity:** unlinkable who communicates with who.

General principles to achieve anonymity

- **Sender anonymity:** Requests not sent directly from sender to recipient, transferred via other nodes.
- **Recipient anonymity:** A large number or all gets the messages sent. Recipient is not addressed explicitly. Participant decides if the message is intended for him.
- **Hide between others:** Conceal meaningful messages in dummy traffic. Anonymity set can hide the activities of a single user within the activities of many other users.

Privacy-preserving problem solution using commutative encryption scheme:

- Customer profile site D, monitoring site G, untrusted noncolluding site S.
- D and G can send synchronised streams of encrypted data and rule clauses to site S.
- Order of attributes is scrambled, s does not learn about attributes.
- Each attribute has two values: true, "don't care".
- Clause has two values for each attribute: X/invalid (masking the real value) and desired result (actual value or "don't care").
- S compares to see if either the first or second value match. If so, it is either an attribute match or a don't care.
- If there is a match for every clause, the rule is true.
- Problem: if all encryptions are the same, S could correlate across rules and instances, however the "don't care", true and invalid values are encrypted directly for each data/rule pair in the stream in a way shared by D and G but unknown to S.