Institutionen för datavetenskap
Department of Computer and Information Science

Examensarbete

A Metric for Anonymity based on Subjective Logic

av

Asmae Bni

LiTH-IDA/ERASMUS-A--14/001--SE

2014-02-07

Linköpings universitet
SE-581 83 Linköping, Sweden

Linköpings universitet
581 83 Linköping

Linköpings universitet
Institutionen för datavetenskap

Examensarbete

A Metric for Anonymity based on Subjective Logic

av

Asmae Bni

LiTH-IDA/ERASMUS-A--14/001--SE

2014-02-07

Handledare: Dr. Klara Stokes and Dr. Leonardo A. Martucci

Examinator: Prof. Nahid Shahmehri

Final thesis

A Metric for Anonymity based on Subjective Logic

by

Asmae Bni

LITH-IDA-EX-2014/LiTH-IDA/ERASMUS-A–14/001–SE

2014-02-07

Final thesis

A Metric for Anonymity based on Subjective Logic

by

Asmae Bni

LITH-IDA-EX-2014/LiTH-IDA/ERASMUS-A–14/001–SE

2014-02-07

Supervisor: Dr. Klara Stokes and Dr. Leonardo A. Martucci

Examiner: Prof. Nahid Shahmehri

Abstract

§ Anonymity metrics have been proposed to evaluate anonymity preserving systems by estimating the amount of information displayed by these systems due to vulnerabilities. A general metric for anonymity that assess the latter systems according to the mass and quality of information learned by an attacker or a collaboration of attackers is proposed here.

The proposed metric is based on subjective logic, a generalization of evidence and probability theory. As a consequence, we proved based on defined scenarios that our metric provide a better interpretation of uncertainty in the measure and it is extended to combine various sources of information using subjective logic operators. Also, we demonstrate that two factors: trust between collaborating attackers and time can influence significantly the metric result when taking them into consideration.

Acknowledgements

I offer my sincere appreciation to my examiner Prof. Nahid Shahmehri for the learning opportunity and her continued support and guidance.
I would like to thank my supervisor Dr. Leonardo A. Martucci, who proposed this project for the helpful discussions and directions.
My completion of this project could not have been accomplished without the help of my supervisor Dr. Klara Stokes, so thank you for the useful suggestions.
I want to thank Prof. Janerik Lundquist coordinator of the Erasmus-Mendus Program for his support.
Partial support by the Programme Averroés-Erasmus Mundus is acknowledged.
To my parents, I offer my deepest gratitude. Your encouragement and unconditional support when the times get rough are much appreciated.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

Modern Internet usage involves economical and social life of its users, hence new privacy threats are introduced. Interactions between users and their personal information can be easily recorded from the Internet, therefore users can be held accounted for every action, also there is a high risk to expose their personal information without their will.

The privacy of users should be protected in networked life. For this purpose, different solutions named privacy enhancing mechanisms were developed. These mechanisms rely on hiding the identity of the users while browsing the web, sending emails, making online transactions, posting comments to social groups or uploading files to servers.

The effectiveness of privacy enhancing technologies is bounded by time and robustness against various attacks. Also the implementation of these technologies is subject to trade off's privacy in return of high performance. Consequently, in addition to the development phase of privacy preserving techniques, the evaluation process is crucial because it demonstrates the level of efficiency of this techniques. Likewise, metrics are needed to prove the reliability of these techniques to the users.

1.2 Problem Description

One of the purposes of an anonymity metric is to evaluate anonymous communication systems. Since anonymous communication systems use different techniques to conceal the identity of its users such as path randomization [20] or MIX stations [10], robust anonymity depends on some required properties that an anonymous communication system should fulfil. These properties are robustness against attacks e.g. timing attacks or traffic analysis attacks or strong recovery from DOS attacks. Hence, measuring anonymity requires formalizing these properties in order to evaluate them.

Measuring anonymity in an anonymous communication system is a challenging and complex task. Anonymity can be measured relatively to a system designer or possible attackers against the system. However, the attacks against anonymous communication systems are diverse because they are scaled to various attackers profiles such as global attacker or malicious user. Also, it is not sufficient to rely on evaluating the techniques used to hide the users identities from a system designer point of view only. Therefore, it is difficult to create a general metric that evaluates all the various aspects of an anonymous communication system at once.

Time is another important aspect in anonymous communication system. Robustness of an anonymous communication system against attacks is relative to time. So it is relevant to evaluate how long the identity of a user will be kept concealed under attacks such as eavesdropping on the communication channel. Thus, a metric should study the behaviour of the system over time and determine the duration of the usability of an anonymous communication system.

Several metrics were proposed to measure anonymity. But the proposed metrics do not cover all the possible aspects of the system. Also the metrics are reliable only under some constraints. For example, in order to apply an anonymity metric, the anonymity set must be clearly defined and it must include the possible sender or receiver of a message. These constraints cannot always be fulfilled in the case of misleading information or errors which can occur in the case of collaboration between attackers or exchange of information.

Information theoretic metrics [18, 61] rely on a clearly predefined probability assignments, which is not easy to achieve in real situations. Because these probabilities are defined by an attacker, they represent the amount of information learned by the attacker about a certain user. Therefore, evidence based metrics [29, 52] are practical because they use evidence to assign belief masses to a user or over a subset of users. Since anonymity is scaled to the attackers abilities, it is relevant to model the collaboration of independent attackers, and trust plays an important role in this case making it necessary to model it along with the collaboration.

Anonymity metrics must consider the uncertainty mass in the measure. However, the entropy measure does not take uncertainty into account, instead all probability assignments for the subjects must add up to one. Evidence based metrics consider uncertainty mass but we notice that all the subjects in the anonymity set share the same uncertainty mass which is not always correct.

1.3 Objective and Scope

The objective is to develop an anonymity metric for anonymous communication systems. The focus of this project is to measure the anonymity of the communicators accordingly to their identities but not to their location.

The anonymity measure is computed by evaluating only the belief mass for each singleton in the anonymity set excluding the case when the belief mass is computed for subsets of the anonymity set. The scope is to measure anonymity considering the case that collaborating attackers fully trust each other. We, nevertheless, define the trust operators in the metric framework for cases in which the attackers do not fully trust each other.

1.4 Method

The approach is divided into two main steps. First, we identify the problem and formalized it mathematically. Second, we propose a metric and evaluate it according to some metric properties and application scenarios. In the formalization phase, we proposed to build a model based on subjective logic that represent the information or evidence learned by an attacker or colluding attackers about the subjects in the anonymity set in the form of opinions and to use the standard deviation of the expectation probabilities as a metric for anonymity. In the evaluation phase, we redefine some metric properties and we compare the proposed metric with some existing metrics, Also we study some attack scenarios in an anonymity preserving system then we visualize the anonymity level from the metric results.

1.5 Contribution

We choose to develop a metric model based on subjective logic, because the subjective logic is a practical and general version of the evidence and probability theory. Therefore, our solution extend the range of the classical metrics and it can be deployed according to various scenarios.

The metric include two measures; anonymity degree and error estimation measure. The degree value is based on the standard deviation of the expectation probabilities values. The error estimation value detects the amount of invalid information in the opinions. In order to compute belief, disbelief and uncertainty masses for every singleton in the anonymity set, we used binomial opinions as a representation of these values for each singleton instead of multinomial opinions. By deploying binomial opinion for each subject, we can measure belief and disbelief masses independently. Thus, our metric can evaluate anonymous communication systems based on evidence that approve or eliminate a suspect in the anonymity set. Also, errors or misleading information can be detected by monitoring the information flow over the anonymity set. Subjective logic operators are also included in our model, fusion operators are deployed to combine evidence over time and trust operators allow to combine opinions based on trust constraints from the collaboration between different adversaries.

1.6 Structure

This report is structured into 10 Chapters, where the Chapters 2, 3, 4 and 5 represent the background. Chapters 6, 7, 8 and 9 describes our contribution and Chapter 10 concludes the thesis.

- Chapter 2: Anonymity and Privacy.
  This chapter defines the terminology for privacy and anonymity and describes how these two notions are related. Also other privacy enhancing techniques are concisely defined.

- Chapter 3: Trust Mechanisms.
  In this chapter, we recall trust concepts and models which represent an important framework to understand the suggested metric model based on subjective logic.

- Chapter 4: Subjective Logic.
  This chapter represents a brief overview of the theory of subjective logic. We recall binomial and multinomial opinion classes and some subjective logic operators.

- Chapter 5: Related Work.
  This chapter is about the state of art of anonymity metrics, where we describe some anonymity metrics and application examples.

- Chapter 6: Anonymity Metric.
  The anonymity metric model which is based on subjective logic is described in this chapter. The model is composed of the opinion's application frame, evidence collection and their mapping to opinions, expectation probabilities, anonymity degree and error estimation measure.

- Chapter 7: Evaluation and Validation.
  Assessment process of the metric is explained in this chapter. We study and compare the proposed anonymity metric against other anonymity metrics which were described in Chapter 3.

- Chapter 8: Applications.
  This chapter describes the application of the metric on Crowds which is an anonymity preserving system. As a result Crowds is evaluated according to the collected evidence and the anonymity level that can be provided to the users in this system.

- Chapter 9: Anonymity Visualization.
  The visualization chapter is about observations of the suspected senders or receivers clusters in the opinion space and how this observations are conform to the metric results.

- Chapter 10: Conclusions and Future Work.
  In this chapter, conclusions are made about the proposed metric model and some observations and directions for future work are mentioned.

Chapter 2

Anonymity and Privacy

2.1 Privacy

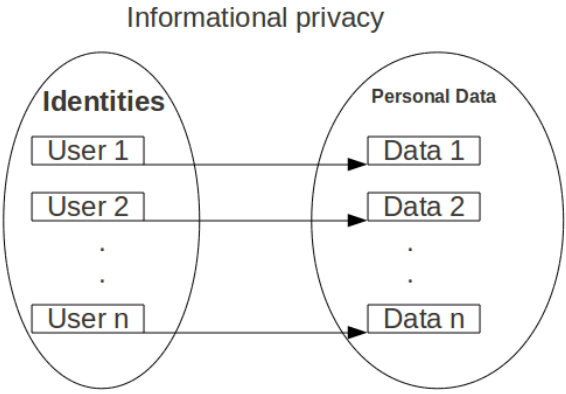
One of the well-accepted definitions of privacy is given by Alan Westin as:

> *"Privacy is the claim of individuals, groups and institutions to determine for themselves, when, how and to what extent information about them is communicated to others[69]."*

This definition states that privacy is a legitimate requirement for individuals or groups. It needs to be protected and enforced whether it concerns personal data, environment or a person himself. Personal information is a valuable asset that needs to be protected in communication systems, hence the focus of this project is informational privacy.

Informational privacy is formally a bijection from the set of user identities to the set of personal data, where every user is one-to-one mapped to his personal data. Which means that only the user should have the right to collect, store or process his own personal data as in the figure 2.1. Also only the individual concerned can grant permission to other parties in order to have access to his personal data. Anonymity is the practice of avoiding or minimizing the use of personal data to identify the related user. Providing anonymity to a system will ensure the highest level of privacy, hence privacy combined with anonymity can be formalized as a one way function such that other parties excluding the legitimate user cannot link the personal data to the legitimate owner.

Figure 2.1: Formal model of informational privacy



2.2 Privacy Enhancing Mechanisms

Privacy enhancing techniques are tools that help to maintain privacy in a communication systems, databases and access control mechanisms. Since privacy cannot be protected by legislation regulations alone, privacy enhancing mechanisms ensure privacy practice by reducing the leakage of personal data [22].

2.2.1 Anonymity

Anonymity is defined by Pfitzmann as:

> *"Anonymity is the state of being not identifiable within a set of subjects, the anonymity set [57]."*

Anonymity reflects the ability of a subject to use a service or have access to a data resource without disclosing his identity. This faculty is achieved by blending the subject into the anonymity set to hide his identity. As a result, other parties may find it difficult to link a data to its owner who is now a subject among the other subjects in the anonymity set. In the context of this thesis we consider anonymity in communication systems. The subjects in the anonymity set may be senders, receivers or the pair sender and receiver, and the linkable item to one of the subjects is a message.

Anonymity ranges from perfect anonymity state to exposure state. The perfect anonymity state is achieved if another party such as one or a group of eavesdroppers on the communication cannot distinguish the potential sender among the senders in the anonymity set. The exposure state is interpreted as a proved identification of the sender or receiver of the item.

2.2.2 Unlinkability

Unlinkability is the inability to link multiple events together that are related to a certain user. A given definition of Unlinkability by Pfitzmann and Hansen is stated as the following:

> *"Unlinkability of two or more items of interest (IOIs, e.g. subjects, messages, events, actions, ...) means that within the system (comprising these and possibly other items), from the attackers perspective, these items of interest are no more and no less related after his observation than they are related concerning his a-priori knowledge [57]."*

Message unlinkability is a form of anonymity since multiple messages cannot be linked to a single subject within the anonymity set.

2.2.3 Pseudonymity

Pseudonymity is using a substitute to a user's identity in order to protect the user's privacy. Consequently, the user's identity cannot be disclosed while accessing resources or using services but he still can be accounted for that use [22].

2.2.4 Unobservability

Unobservability is achieved when other parties who are not concerned by an event cannot prove that this event has taken place. Unlinkability of a sender and a receiver is a form of unobservability since they appear as if they are not communicating with each other [22].

Chapter 3

Trust Mechanisms

The internet is a common ground where users can exchange information or provide services to each others. However, since internet was deployed largely for commercial and marketing purposes also malicious behaviour was noticed among the entities using it. Accordingly, this behaviour triggers issues of reliability and quality of shared information or services which demonstrate the need of effective trust mechanisms approaches to assess the shared information and manage trust.

3.1 Trust

Trust is considered as a psychological concept rather than a computational one, as a result the idea of assessing it was rejected earlier. However, trust is considered an important component in IT security. Hence, trust management is considered valuable because it helps reduce the malicious behaviour by discovering it and enforcing ethical norms in the system by recognizing or rewarding the adopters of this norms. Trust was defined according to Jøsang as:

> *"Trust is a directional relationship between a trustor and a trustee [39]."*

This definition is valid within a scope that includes two types of trust: reliability trust and decision trust that are defined by Jøsang as the following:

> *"**Reliability trust** is the subjective probability[1] by which an individual A, expects that another individual B, to perform a given action on which its welfare depends [39]."*

[1]The subjective probability is a probability based derived from a personal judgement that is not based on formal calculations.

> *"**Decision trust** is the extent to which a given party is willing to depend on something or somebody in a given situation with a feeling of relative security, even though negative consequences are possible [39]."*

3.2 Trust Models

Trust systems describe the relationship between the relying entity and the trusted entity using a score which is quantified using one of the models explained in this section. The score can be a quantitative value or a qualificative expression assigned by the trustor to the trustee which demonstrate the trust level or the reputation of the trustee based on previous encounters.

3.2.1 Discrete Trust

Discrete trust measures uses discrete verbal statement to evaluate trust levels. The discrete verbal statement can range from usually to poorly. These expressions are easily interpreted by humans rather than probabilistic values which may be confusing sometimes, however the verbal statements are not accurately compared to the numerical values.

3.2.2 Probabilistic Trust Models

Probabilistic trust models are based on probability assignments as a numerical representation of the reputation of an agent or a trustee. The probability assignment represents the percentage by which an agent can be reliable from the perspective of a trustor. In order to deduce the trust rate for each agent within a group of agents, the probability assignments can be normalized so that agents can be compared based on their trust rates. Also an agent's trust rate can be combined from different trustors using multiplication operations.

3.2.3 Belief Models

A belief model is a generalization of the probabilistic model. In a belief model, the belief masses are quantified based on evidence and the sum of belief masses over all possible states of the universal set does not necessarily add up to one, therefore the uncertainty mass can be quantified as the remaining mass. There are two belief models: Dempster-Shafer theory [62] and subjective logic [42].

According to Dempster-Shafer theory [62] also known as evidence theory, the belief degree denoted Bel is computed over the power set of a universal set and it represents the extent to which evidence supports that a given claim is true, while the plausibility degree denoted Pl is equal to $1 - Bel$ and it represents the extent to which evidence supports the opposite of the given claim. Also evidence theory provide a mean to combine different sources

of evidence using fusion operators and Dempster's rule of combination to combine belief degrees from independent sources.

Jøsang [33, 42] proposed that the belief of a statement x can be expressed as a subjective opinion created by an individual A. The opinion is noted $w_x^A = (b, d, u, a)$ where b, d, u and a represent respectively belief, disbelief, uncertainty and base rate values. The belief and disbelief masses are computed based on the collected evidence and the uncertainty u is the remaining mass such as $u = 1 - (b + d)$. Also opinions can be combined according to trust constraints from different individuals using discounting and consensus operators.

Chapter 4

Subjective Logic

Subjective logic is a generalization of probability and evidence theory. According to classic probability theory, the sum of probabilities of all possible states in the universal set must add up to one. As a result, this approach does not consider uncertainty measure in the probability distribution over the states. Evidence theory [62] is based on second order probabilities by applying belief mass functions over a power set of the states in order to quantify ignorance about some states however subjective logic provide an intuitive interpretation of belief functions [33].

4.1 Opinions

An opinion is a representation of belief, disbelief and uncertainty masses over a frame. The frame is the set of all possible statements to which an opinions is applied. We denote an opinion by w_X^A where A is the owner of the opinion and X is the frame. The opinions are classified as binomial, multinomial or hyper opinions according to the type of the application frame [42].

4.1.1 Binomial Opinion

The binomial opinion is applied on a binary frame X. The frame includes only two opposite states x and $\bar{x}$. A binomial opinion is equivalent to a β probability distribution function denoted as $\beta(p|r,s,a)$ such that:

$$\beta(p|r,s,a) = \frac{\Gamma(r+s+\omega)}{\Gamma(r+\omega a)\Gamma(s+\omega(1-a))} p^{(r+\omega(1-a)-1)}(1-p)^{(s+\omega(1-a)-1)}$$

where:

- Γ is the Gamma function;
- p is a probability variable such that $0 \leq p \leq 1$;

- ω is the non-informative weight which ensure that β is uniform when $r = s = 0$;
- a is the base rate distribution over X, it represents the a-priori probability distribution in the absence of evidence and by default $a = \frac{1}{2}$;
- $r \geq 0$ is an integer that represent the number of observations to support claim x;
- $s \geq 0$ is an integer that represent the number of observations to support claim $\bar{x}$.

4.1.1.1 Mapping Binomial Opinion to Beta

A binomial opinion $w_X = (b, d, u, a)$, such that:

- b denote the belief mass,
- d denote the disbelief mass,
- u denote the amount of uncommitted belief mass,

is equivalent to $\beta(p|r, s, a)$ by the following mapping:

$$\begin{cases} b = \frac{r}{\omega+r+s} \\ d = \frac{s}{\omega+r+s} \\ u = \frac{\omega}{\omega+r+s} \end{cases} \iff \left(\begin{array}{cc} \text{For } u \neq 0 & \text{For } u = 0 \\ \begin{cases} r = \frac{\omega b}{u} \\ s = \frac{\omega d}{u} \\ 1 = b + d + u \end{cases} & \begin{cases} r = \infty \\ s = \infty \\ 1 = b + d \end{cases} \end{array} \right)$$

4.1.1.2 Expectation Probability

A probability expectation is an average estimation of the possible outcomes in the application frame. The expected value is the average value of the Beta probability density function that is equivalent to a binomial opinion [42].

Definition 4.1. Let $w_X = (b, d, u, a)$ be an opinion. The expectation probability is defined as:

$$E = b + au$$

The uncertainty value is interpreted in the expectation probability by multiplying it to the base rate value and adding it to the belief mass, therefore the base rate determines the extent to which the uncertainty should be included in the expectation value.

4.1.2 Multinomial Opinion

A multinomial opinion is applied on a frame X of cardinality $n > 2$ such that:

$$X = \{x_1, x_2, x_3, \ldots, x_n\}$$

4.1.2.1 Opinion Representation

Definition 4.2. Let $w_X = (\overrightarrow{b}, u, \overrightarrow{a})$ be a multinomial opinion, where

- $\overrightarrow{b}$ denotes the belief mass vector such that $\sum_{i=1}^{n} \overrightarrow{b}(x_i) \leq 1$;
- u denotes the uncertainty scalar;
- $\overrightarrow{a}$ denotes the base rate vector such that $\sum_{i=1}^{n} \overrightarrow{a}(x_i) = 1$.

4.1.2.2 Expectation Probabilities Vector

Definition 4.3. The expectation probabilities vector of a multinomial opinion is quantified as follow:

$$\overrightarrow{E}_X(x_i) = \overrightarrow{b}_X(x_i) + \overrightarrow{a}_X(x_i)u_X, \forall x_i \in X.$$

4.2 Operators

Evidence or opinions can be collected from different sources and during different time periods. The variety of resources generates questions about trust and how we can assess the agents opinions and extract the most valuable or correct information. Also, time is an important aspect since each evidence is related to an interval of time. The model of subjective logic offers operators to combine evidence.

4.2.1 Fusion Operators

Fusion operators are used to combine evidence collected from a trusted source according to time constraints.

- Let $V^A(r^A, s^A, t^A)$ and $V^B(r^B, s^B, t^B)$ two evidence collected by the agents A and B;
- Let $\Delta t = |t^A - t^B|$ be the time difference of these two evidence such that $t^A \geq t^B$.

4.2.1.1 Averaging Fusion

The averaging operator $\bigoplus$ is used when two trusted evidence are collected in the same time span.

If $\Delta t = 0$ then $V^A(r^A, s^A, t^A) \bigoplus V^B(r^B, s^B, t^B) = V(\frac{r^A + r^B}{2}, \frac{s^A + s^B}{2}, t^A)$.

4.2.1.2 Cumulative Fusion

When two trusted evidence are collected in different time intervals, then the cumulative operator $\underline{\bigoplus}$ is used to combine correctly the evidence.

If $\Delta t \neq 0$ then $V^A(r^A, s^A, t^A) \underline{\bigoplus} V^B(r^B, s^B, t^B) = V(r^A + r^B, s^A + s^B, t^A)$.

4.2.2 Consensus

The consensus operator $\diamond$ is used when two or more agents have different opinions about an agent who recommends an opinion about a statement x as in figure 4.1. In this case, the consensus operator combines correctly the opinions by reducing uncertainty, and a more accurate opinion about the recommending agent is created. For more details and examples, see [44, 37, 36].

Definition 4.4. Let w_C^A and w_C^B opinions of agent A and agent B about agent C. Then the opinion held by an imaginary agent $[A \diamond B]$ about C is:

$$w_C^{A\diamond B} = (b_C^{A\diamond B}, d_C^{A\diamond B}, u_C^{A\diamond B}, a_C^{A\diamond B}),$$

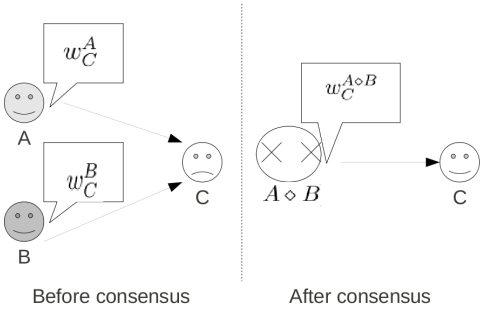
such that:

1. Case: $u_C^A + u_C^B - u_C^A u_C^B \neq 0$
   - $b_C^{A\diamond B} = \frac{b_C^A u_C^B + b_C^B u_C^A}{u_C^A + u_C^B - u_C^A u_C^B}$;
   - $d_C^{A\diamond B} = \frac{d_C^A u_C^B + d_C^B u_C^A}{u_C^A + u_C^B - u_C^A u_C^B}$;
   - $u_C^{A\diamond B} = \frac{u_C^A + u_C^B}{u_C^A + u_C^B - u_C^A u_C^B}$;
   - $a_C^{A\diamond B} = \frac{a_C^A u_C^B + a_C^B u_C^A - (a_C^A + a_C^B) u_C^A u_C^B}{u_C^A + u_C^B - 2u_C^A u_C^B}$.
2. Case: $u_C^A + u_C^B - u_C^A u_C^B = 0$
   - $b_C^{A\diamond B} = (\gamma^{A/B} b_C^A + b_C^B)/(\gamma^{A/B} + 1)$;
   - $d_C^{A\diamond B} = (\gamma^{A/B} d_C^A + d_C^B)/(\gamma^{A/B} + 1)$;
   - $u_C^{A\diamond B} = 0$;
   - $a_C^{A\diamond B} = \frac{\gamma^{A/B} a_C^A + a_C^B}{\gamma^{A/B} + 1}$.

Where the relative weight is $\gamma^{A/B} = \lim u_C^B / u_C^A$ as $u_C^B, u_C^A \to 0$ such that:

- If w_C^A and w_C^B are harmonious opinions, then $\gamma^{A/B}$ is a finite and non-zero value;
- If w_C^A and w_C^B are highly conflicting opinions, then:
  - $w_C^{A\diamond B} = w_C^A$ when $\gamma^{A/B} = \infty$;
  - $w_C^{A\diamond B} = w_C^B$ when $\gamma^{A/B} = \epsilon$.

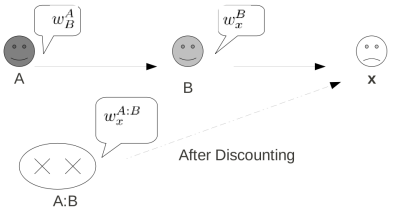
Figure 4.1: Consensus operator



4.2.3 Discounting

The discounting operator takes into account the opinions collected from other agents to formulate the final opinion about a subject x as in figure 4.2. Since the adversary does not trust the agents, the discounting operator can be classified into three categories according to trust constraints.

Figure 4.2: Discounting operator



4.2.3.1 Uncertainty Favouring Discounting

This discounting operator is used when the adversary believes that the opinion recommended by the other adversary is not true, therefore he ignores the belief and disbelief masses and takes into account only the uncertainty values.

Definition 4.5. Let w_B^A an opinion of agent A about agent B and w_x^B an opinion of agent B about subject x in the anonymity set N recommended to A. Then the discounted opinion about x as:

$$w_x^{A:B} = (b_x^{A:B}, d_x^{A:B}, u_x^{A:B}, a_x^{A:B}),$$

where:

- $b_x^{A:B} = b_B^A b_x^B$;

- $d_x^{A:B} = b_B^A d_x^B$;
- $u_x^{A:B} = d_B^A + u_B^A + b_B^A u_x^B$;
- $a_x^{A:B} = a_x^B$.

4.2.3.2 Opposite Belief Favouring Discounting

This operator is used when the recommending agent have a history of misleading others by giving incorrect information. So, instead of accounting the belief value, the adversary considers only the disbelief measure in his opinion.

Definition 4.6. Let w_B^A an opinion of agent A about agent B and w_x^B an opinion of agent B about x one of the subjects in the anonymity set N recommended to A. Then, the final opinion about x is:

$$w_x^{A:B} = (b_x^{A:B}, d_x^{A:B}, u_x^{A:B}, a_x^{A:B}),$$

where:

- $b_x^{A:B} = b_B^A b_x^B + d_B^A d_x^B$;
- $d_x^{A:B} = b_B^A d_x^B + d_B^A b_x^B$;
- $u_x^{A:B} = u_B^A + (b_B^A + d_B^A) u_x^B$;
- $a_x^{A:B} = a_x^B$.

4.2.3.3 Base Rate Sensitive Discounting

Discounting based on the base rate is performed by an adversary if he cannot judge the collected opinion. Therefore, he considers the base rate value to formulate a judgement about the opinion resource.

Definition 4.7. Let w_B^A an opinion of agent A about agent B and w_x^B an opinion of agent B about the subject x collected from A. Then, the final opinion about x is:

$$w_x^{A:B} = (b_x^{A:B}, d_x^{A:B}, u_x^{A:B}, a_x^{A:B}),$$

where:

- $b_x^{A:B} = E(w_B^A) b_x^B$;
- $d_x^{A:B} = E(w_B^A) d_x^B$;
- $u_x^{A:B} = 1 - E(w_B^A)(b_x^B + d_x^B)$;
- $a_x^{A:B} = a_x^B$.

The probability expectation is quantified as $E(w_B^A) = b_B^A + a_B^A u_B^A$.

Chapter 5

Related Work

5.1 Anonymity Set Size

David Chaum was the first to propose quantifying anonymity in the dining cryptographers problem [12]. He used a simple example to illustrate the dinning cryptographers problem. Three cryptographers are sitting in the dining room and the waiter confirm that the dinner was paid anonymously. The three cryptographers wants to know if the dinner was paid by one of them or by someone else but they still want to maintain the anonymity of the payer. So, they developed a small experiment, where each cryptographer flip a fair coin behind his menu and compares his outcome with the outcome of his partner on the left side, then announces the result as same if the outcomes match, otherwise different. However, there is a twist since only the cryptographer who has paid should announce the opposite of the result. Therefore, the anonymity of the payer is preserved, since the two possibilities that the hidden outcome was the same or different than the one declared are equally likely for all the cryptographers.

The DC channels concept is a generalization of the dining cryptographers experiment that is intended to preserve the identity of the senders. Each participant in the channel generate a secret key and share it with his right neighbour. Since each pair of participants has a shared key, the participant in each pair combine the key bits and message bits using the Exclusive OR operator $\oplus$.

Chaum argued that the anonymity set size is an indicator of the anonymity level of the dining cryptographers protocol. The anonymity set is the total number of the communicators that uses DC channels. This assumption is based on the fact that an observer can only link a message with an equal probability to each one of the communicators in the anonymity set since the protocol theoretically offers a perfect anonymity state to the communicators. So, the system is anonymous as long as two or more communicators use the DC channel. However, there are drawbacks in the implementation of this

complex protocol, e.g. only one user can use the protocol at a time and a unique random key must be generated only once for every message, also the DC channel does not provide protection against malicious communicators.

5.2 Individual Anonymity Metric

In order to evaluate the anonymity in Crowds, the designers of this system proposed the Individual anonymity metric [60]. Therefore, Each possible sender in the anonymity set is evaluated individually by a theoretical attacker which assign a probability p_i for each subject i in an anonymity set of size n, such that $\sum_{i=1}^{n} p_i = 1$. The anonymity degree is quantified as:

$$d = 1 - \max(p_1, \ldots, p_n).$$

5.3 Information Theoretic Metrics

Information theoretic metrics are based on Shannon's information theory [9]. The key feature in this theory is the entropy measure which quantifies the amount of uncertainty in a probability distribution. As a consequence, the entropy measure is higher when the probability distribution is uniform.

5.3.1 Entropy based Metric

Serjantov and Danezis [61] used the concept of entropy to measure anonymity and introduced the effective anonymity set size $\mathcal{S}$ concept as an anonymity measure. The effective anonymity set size reflects the randomness of the probability distribution among the subjects in the anonymity set. The effective anonymity set size $\mathcal{S}$ is defined as follow:

$$\mathcal{S} = -\sum_{i=1}^{n} p_i \log_2(p_i),$$

such that:

- n is the cardinal of the anonymity set;
- p_i is the probability that a subject i send or receive the message M.

5.3.2 Normalized Entropy based Metric

The normalized metric as described in [18] is also based on entropy as a measure except that the degree of anonymity is quantified relatively to the state of perfect anonymity. The anonymity measure indicate the amount of information leaked in the system in term of bits and it is described as:

$$d = \frac{-\sum_{i=1}^{n} p_i \log_2(p_i)}{\log_2(n)}$$

5.3.3 Combinatorial Metric

The combinatorial anonymity metric measures the anonymity of the complete communication pattern from the perspective of an adversary who tries to distinguish the users of the communication system according to their roles as senders/receivers and their identities [21]. So the measure includes every possible combination of sender and recipient in the system instead of considering only the anonymity of a sender or receiver that is related to a single message M.

This measure approach is represented using the permanent of a matrix A. The matrix A is a doubly stochastic matrix computed by an attacker such that $\forall(i,j) \in [1 \dots n]^2, A(i,j)$ is the likelihood that a sender i send a message to a receiver j.

The combinatorial anonymity degree is computed from the following formula:

$$d(A) = \begin{cases} 0 & \text{if } n = 1 \\ \frac{\log(per(A))}{\log(\frac{n!}{n^n})} & \text{if } n > 1 \end{cases},$$

such that $per(A) = \sum_{\pi} \prod_{i=1}^{n} A(i, \pi(i))$ and π is the permutation of $[1, \dots, n]$.

The element $\log(\frac{n!}{n^n})$ is the measure in the case of perfect anonymity state and $\log(per(A))$ is the anonymity measure based on the probabilities assigned by the attacker. It is considered that the anonymity is absent when only one pair of sender and receiver exists.

5.4 Evidence based Metric

The evidence based metrics [52] are metrics based on evidence theory [62] and the mentioned evidence metric is applied for wireless mobile ad-hoc networks [29]. An evidence is an intercepted packet since it supports the claim that a communication link is set between two entities in the network, and the collected evidence is represented by the number of intercepted packets within a time span. The probability assignment for each entity in the anonymity set is computed according to the collected evidence and the unit of the anonymity degree is in term of bits.

The anonymity level is quantified from the perspective of adversaries, therefore it is scalable to their traffic analysis skills. In order to locate the source of a packet, the adversary computes the values $w(V), m(V), Bel(V)$ and $Pl(V)$ where:

- $w(V)$ is the quantity of evidence that supports a claim;
- $m(V)$ is the assigned probability to the claim;
- $Bel(V) = \sum_{U|U \subseteq V} m(U)$ is the belief mass;
- $Pl(V) = \sum_{U|U \cap V \neq 0} m(U)$ is the plausibility mass.

U and V represent two ordered sets of communicating nodes since they represent the direction of the detected packets, and F is the power set of the anonymity set such that $U, V \subseteq F$.

The anonymity measure is an entropy-like measure, however belief and plausibility masses over the ordered sets are used instead of probability distribution over the anonymity set. The evidence based measure $D(m)$ is an average measure of plausibility measure $E(m)$ and belief measure $C(m)$, where :

- $E(m) = -\sum_{V \in F} m(V) \log_2(Pl(V))$;
- $C(m) = \sum_{V \in F} m(V) \log_2(Bel(V))$;
- $D(m) = -\sum_{V \in F} m(V) \log_2(\sum_{U \in F} m(U) \frac{|U \cap V|}{|U|})$.

The role of the element $\frac{|U \cap V|}{|U|}$ is to eliminate insignificant evidence from the belief measure. The evidence measure $D(m)$ is bounded by $E(m)$ and $C(m)$ such that $E(m) \leq D(m) \leq C(m)$.

Chapter 6

A Metric for Anonymity

In this chapter, we describe a detailed model to measure anonymity based on subjective logic. The metric is a computational process designed to illustrate the counting process of evidence and their deployment to compute the anonymity degree in a communication system.

6.1 Application Frame

Let Ω be the sample space of all the users in a communication system, the users are the subjects in the anonymity set N related to a message M, where N represents the set of either all possible senders or possible receivers of M. In the beginning, the anonymity set includes all the users in the communication system and we assume that the cardinality of this set is $n = |N|$. The set size may only decrease in time given the amount of information learned by an adversary, as a consequence the anonymity set notion is relative to time [57].

Let $X = \{x_1, x_2, \ldots, x_n\}$ be the multinomial frame as in figure 6.1, such that x_i represent the claim that a subject i in the anonymity set is the sender of M. A multinomial opinion can be substituted into a set of binomial opinions. However the belief mass b_i of the state x_i is evaluated by the positive evidence that supports the claim x_i and the disbelief mass d_i of the state x_i is evaluated implicitly by the positive evidence that supports the claims $x_j, j \in [1, \ldots, n] \setminus \{i\}$, such that:

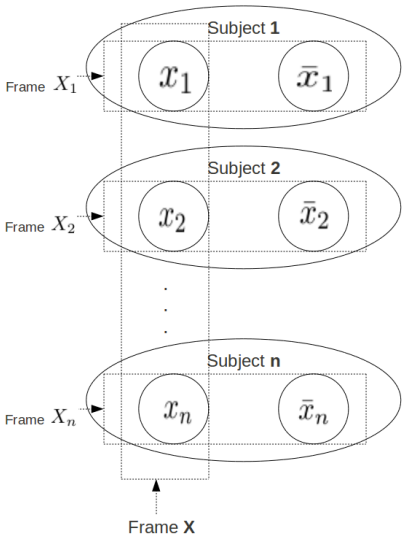
$$d_i = \frac{\sum_{\substack{j \in [1 \ldots n] \\ j \neq i}} r_j}{\omega + r_i + \sum_{\substack{j \in [1 \ldots n] \\ j \neq i}} r_j}, \forall i \in [1, \ldots, n],$$

where r_i denote the positive evidence that supports state x_i and ω denote the non-informative weight.

Applying opinions to the multinomial frame results in a controlled flow of information since $\sum_{i=1}^{n} \overrightarrow{b}_X(x_i) \leq 1$ and $\sum_{i=1}^{n} \overrightarrow{d}_X(x_i) \leq n - 1$. Therefore,

the multinomial frame is conform to the anonymity set concept because only one of the states $x_i, i \in [1, \ldots, n]$ can be true.

Figure 6.1: Multinomial and binomial frames



We consider another approach by applying separately each opinion to its own binary frame X_i, consequently there are no bounds on the sum of belief or disbelief masses. This approach is useful for the case of combining opinions from different adversaries, since it is a flexible representation of opinions and it is possible to detect misleading information or insignificant evidence by monitoring the flow of information.

Let $X_i = \{x_i, \bar{x}_i\}$ be a binomial frame related to a subject i in figure 6.1, such that $i \in [1, \ldots, n]$. This frame is the local frame to which evidence applies, where:

- x_i represents the belief state;
- $\bar{x}_i$ represents the disbelief state.

Since each subject is mapped to its own binary frame, the values of belief, disbelief and uncertainty are computed for each subject independently and the base rate values are used to compute the expectation probabilities[1].

After collecting evidence and computing opinions, four outcomes are possible according to our model:

1. The adversary succeeds to identify either the sender or receiver from the rest in the anonymity set N.

[1]In Chapter 4, there is a detailed description of the stated values.

2. The adversary believes that any subject can be either the possible sender or receiver of the message M.

3. The adversary believes that none of the subjects sent or received the message M.

4. The evidence are not sufficient to detect either the possible sender or the possible receiver in the anonymity set N.

6.2 Evidence Collection

An evidence is a concrete measurement of the real world that validates the truth of a statement and it is used to evaluate the belief degree of a statement. In the context of this project, we consider two types of evidence. The positive evidence that proves a statement and the negative evidence that disproves it. We abstract the values of evidence as positive integers. These values can represent, for example, the number of observed packets or the transmission signal strength in an anonymous communication channel. Therefore, evidence represent the amount of information leaked in the channel or the information exchanged by adversaries.

Definition 6.1. Let $X_i = \{x_i, \bar{x}_i\}$ be the state space for a subject i in the anonymity set N.

The vector $V_i(r_i, s_i, t_i) \in \mathbb{N}^3$ is the evidence vector associated to each subject i, where:

- r_i denotes the number of observations that support the state x_i;
- s_i denotes the number of observations that support the state $\bar{x}_i$;
- t_i denotes the time period when the evidence was collected.

Evidence can be combined according to time constraints using fusion operators [2].

6.3 Subjective Opinion Matrix

Each subject i in the anonymity set N is a potential sender of the message M, therefore each subject i is represented by a binomial opinion W_{X_i}.

Notation 6.2. In order to simplify the model, the opinions about the subjects in the anonymity set are combined into a matrix that we denote the opinion matrix W. The matrix $W \in [0, 1]^{n \times 4}$ is defined as:

[2]The fusion operators are defined in Chapter 4 Section 4.2.1

$$W_{n,4} = \begin{pmatrix} b_1 & d_1 & u_1 & a_1 \\ b_2 & d_2 & u_2 & a_2 \\ \vdots & \vdots & \vdots & \vdots \\ b_n & d_n & u_n & a_n \end{pmatrix}$$

Properties 6.3. The properties of the subjective opinion matrix are:

- $b_i + d_i + u_i = 1, \forall i \in [1, \ldots, n]$;
- $\sum_{i \in [1 \ldots n]} b_i \leq n$;
- $\sum_{i \in [1 \ldots n]} d_i \leq n$;
- $\sum_{i \in [1 \ldots n]} u_i \leq n$;
- $\sum_{i \in [1 \ldots n]} b_i + d_i + u_i = n$.

Hypothesis 6.4. By default, in the absence of evidence the subjective opinion matrix is expressed as follows:

$$W_{n,4} = \begin{pmatrix} 0 & 0 & 1 & \frac{1}{n} \\ 0 & 0 & 1 & \frac{1}{n} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 1 & \frac{1}{n} \end{pmatrix}$$

6.4 Expectation Probabilities Vector

The expectation probability is a probability estimated according to the amount of evidence.

Notation 6.5. We denote E the expectation probabilities vector that includes the expectation probabilities for all the subjects in the anonymity set N, where $\sum_{i=1}^{n} E_i \leq n$.

$$E = \begin{bmatrix} b_1 + a_1 u_1 \\ b_2 + a_2 u_2 \\ \vdots \\ b_n + a_n u_n \end{bmatrix}$$

Example 6.6. Consider the case where we represent an opinion about each of three users in the anonymity set $N = \{x_1, x_2, x_3\}$. Let the default non-informative prior weight be $\omega = 3$ and the base rate for each user $a = \frac{1}{3}$. Also, let the opinions be w_{x_1}, w_{x_2} and w_{x_3} such as they are respectively equivalent to the following β probability distribution functions, where the arguments are arbitrarily chosen;

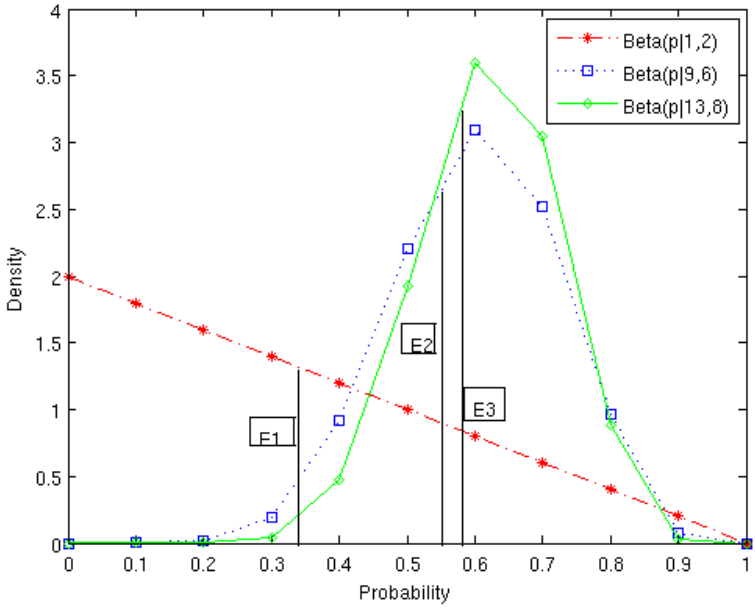
- $\beta(p|1, 2) \equiv w_{x_1} = (0, 0, 1, \frac{1}{3})$;

- $\beta(p|9,6) \equiv w_{x_2} = (0.5, 0.334, 0.16, \frac{1}{3})$;
- $\beta(p|13,8) \equiv w_{x_3} = (0.54, 0.334, 0.125, \frac{1}{3})$.

When computing the expectation probabilities for each user, it can be observed that they are different as the following:

- $E_1 = 0.33$;
- $E_2 = 0.553$;
- $E_3 = 0.581$.

Figure 6.2: Probability density function of the β distributions



In figure 6.2, we notice that the length of the vertical lines confirms the empirical order of the corresponding expectation values for each subject, which means the higher the positive evidence masses for a subject, the higher is the expectation probability related to this subject.

6.5 Our Proposed Measure for Anonymity

We propose a degree formula based on the standard deviation of the expectation probabilities vector for measuring anonymity. The proposed degree

describes the anonymity level provided by the anonymity preserving communication system and it represents the amount of evidence an adversary needs in order to reveal the identity of a subject in the anonymity set. Therefore the unit of measurement depends on the evidence unit.

Definition 6.7. Let E be the expectation probabilities vector. The normalized expectation probabilities vector E' is defined as $E' = \frac{1}{n\bar{E}} \times (E_1, E_2, \ldots, E_n)$, such that $\bar{E} = \frac{1}{n}\sum_{i=1}^{n} E_i$ is the mean of the expectation probabilities.

Definition 6.8. Let E' be the normalized expectation probabilities vector. The anonymity degree $d(E')$ is defined as:

$$d(E') = \begin{cases} 0 & \text{if } n = 1 \\ 1 - n\sqrt{\frac{1}{n(n-1)}\sum_{i=1}^{n}(E'_i - \bar{E}')^2} & \text{if } n > 1 \end{cases}$$

Lemma 6.9. Let $d(E')$ be the anonymity degree, then $0 \leq d(E') \leq 1$

Proof. Let $E' \in [0,1]^n$ be the normalized expectation probabilities vector and let σ be the standard deviation of the expectation probabilities, such that:

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(E'_i - \bar{E}')^2}.$$

Let $i, j \in [1, \ldots, n]$, the standard deviation σ of E' is maximum if $\exists j$ such that $E'_j = 1$ and $\forall i \in [1, \ldots, n] \setminus \{j\}$ $E'_i = 0$ in which case $\sigma = \frac{\sqrt{n-1}}{n}$ implying that:

$$\sigma \leq \frac{\sqrt{n-1}}{n}$$

The standard deviation σ of E' is minimum if $E'_1 = E'_2 = \ldots = E'_n$ which implies that:

$$\sigma \geq 0$$

Therefore we conclude that:

$$0 \leq \sqrt{\frac{1}{n}\sum_{i=1}^{n}(E'_i - \bar{E}')^2} \leq \frac{\sqrt{n-1}}{n}$$

$$\Rightarrow -\frac{\sqrt{n-1}}{n} \leq -\sqrt{\frac{1}{n}\sum_{i=1}^{n}(E'_i - \bar{E}')^2} \leq 0$$

$$\Rightarrow -1 \leq -\frac{n}{\sqrt{n-1}}\sqrt{\frac{1}{n}\sum_{i=1}^{n}(E'_i - \bar{E}')^2} \leq 0$$

$$\Rightarrow 0 \leq 1 - \frac{n}{\sqrt{n-1}}\sqrt{\frac{1}{n}\sum_{i=1}^{n}(E'_i - \bar{E}')^2} \leq 1$$

As a consequence:

$$0 \leq d(E') \leq 1$$

□

6.6 Error Estimation Rate

Since the anonymity degree is computed from the normalized expectation probabilities vector, we propose the error estimation rate as a complement to the degree. The error estimation rate is a measure that detects the absence or the overflow of information or evidence in the expectation probabilities vector.

Definition 6.10. Let E be the expectation probabilities vector, we define the error estimation rate $err(E)$ as:

$$err(E) = \bar{E} - 1/n,$$

where $\bar{E}$ is the mean of the expectation probabilities vector.

The error estimation measure can be interpreted as an indicator of the lack or the overflow of the information used to compute the expectation probabilities such that:

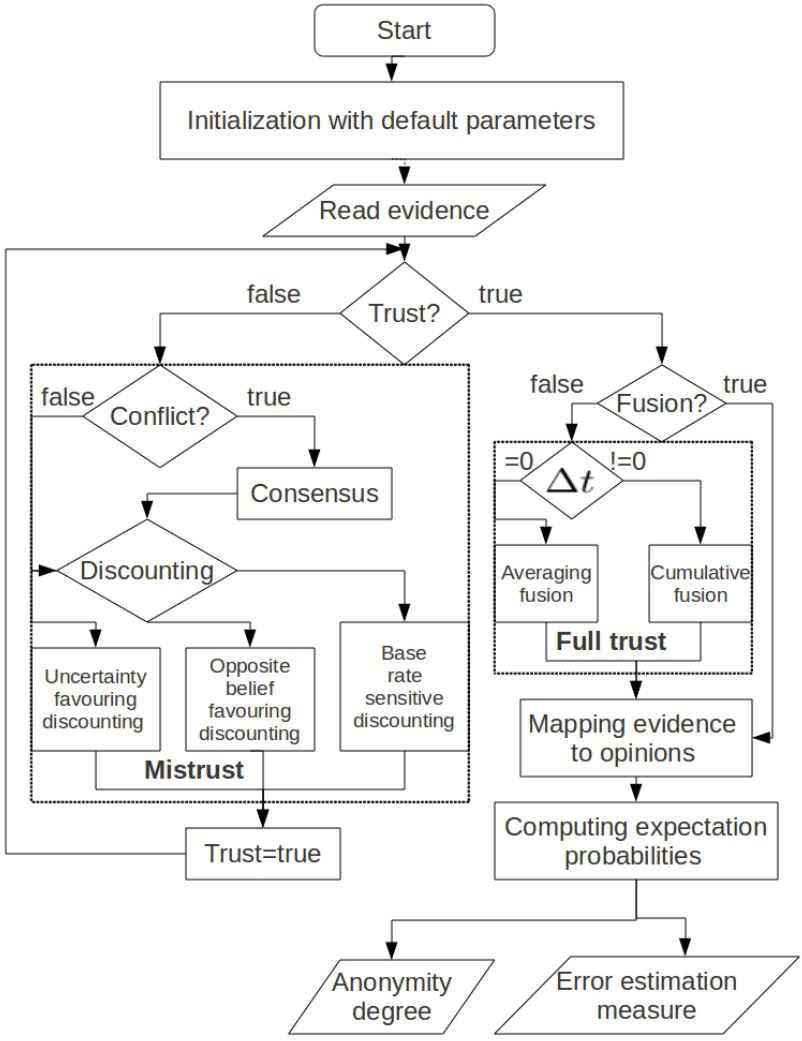
- If $err(E) < 0$, then $\sum_{i=1}^{n} E_i < 1$. Therefore the error estimation value can be interpreted as:
  1. an overflow of negative evidence when $\sum_{i=1}^{n} d_i > n-1$, since the sum of the disbeliefs is supposed to be inferior or equal to $n-1$ according to the multinomial opinion model;
  2. otherwise, as a lack of evidence expressed by high uncertainty.
- If $err(E) > 0$, then $\sum_{i=1}^{n} E_i > 1$. Therefore the error estimation value can be interpreted as:
  1. an overflow of positive evidence when $\sum_{i=1}^{n} b_i > 1$, since the sum of the beliefs is supposed to be inferior or equal to 1 according to the multinomial opinion model;
  2. otherwise, as a lack of evidence expressed by high uncertainty.

The error estimation measure bases its analysis on comparing between the multinomial opinion model and the binomial opinion model. Within the multinomial model, the sum of the expectation probabilities is always equal to one, however in our model the sum of the expectation probabilities can be equal to or different from one. This makes it possible to detect misleading information in our model in some cases, that is, when the sum of the expectation probabilities is different from one. If the sum of the expectation probabilities equals one, then our error estimation measure assumes the non-existence of misleading information.

6.7 Metric Model

The chart 6.3 below summarize the metric model and represent the computational steps derived from Chapter 4 in order to compute the proposed anonymity degree and error estimation rate. The metric model supports both collaborating attackers and single adversary profiles by applying trust operators such as the consensus and discounting operators in order to combine opinions from different sources.

Figure 6.3: Metric flow chart



Chapter 7

Evaluation

In this chapter, we validate the proposed metric model by comparing it to several anonymity metrics defined in Chapter 5.

7.1 Comparison Model

In order to compare our metric with entropy and evidence based metrics[1], we cite in table 7.1 some criteria that an anonymity metric should fulfil. The features noted by F2 and F3 were proposed by Andersson and Lundin [1] and those features are extended to features F1, F4 and F5.

Table 7.1: Desired features of an anonymity metric

Feature	Description
F1	When combining multiple opinions from different adversaries, the anonymity degree is lower when the adversaries agrees on a subject as the potential sender and higher if they disagree.
F2	An anonymity metric must have a well defined range between two end points noted by a maximum and a minimum value.
F3	The anonymity degree is maximum when the distribution of belief masses or probabilities over the subjects in the anonymity set is uniform.
F4	Anonymity metric results should strive for being objective even if the analysis is based on misleading information.
F5	A realistic anonymity metric should base its analysis on evidence.

[1] See Chapter 5 Sections 5.3 and 5.4

7.2 Data Representation

We represent Data by a matrix of dimension $n \times 4$, the subjective opinion matrix, including belief, disbelief, uncertainty and base rate measures according to Notation 6.4 in Chapter 6. In order to compare the subjective metric to evidence based metric and entropy based metric, we represent belief mass and probability measures in the same matrix format as the subjective opinion matrix.

7.2.1 The Evidential Opinion Matrix

The evidential opinion matrix describe the belief mass over each singleton s_i in the anonymity set where a singleton represents a subject [2].

$$W'_{n,4} = \begin{array}{c} \\ s_1 \\ s_2 \\ \vdots \\ s_n \end{array} \begin{array}{c} \begin{array}{cccc} b_i & d_i & u_i & a_i \end{array} \\ \begin{pmatrix} b_1 & \sum_{\substack{i\in[1\ldots n]\\ i\neq 1}} b_i & u_1 & 0 \\ b_2 & \sum_{\substack{i\in[1\ldots n]\\ i\neq 2}} b_i & u_2 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ b_n & \sum_{\substack{i\in[1\ldots n]\\ i\neq n}} b_i & u_n & 0 \end{pmatrix} \end{array}$$

Properties 7.1. The properties of the evidential opinion matrix are:

- $u = u_1 = u_2 = \ldots = u_n$;
- $\sum_{i\in[1\ldots n]} b_i + u = 1$;
- $\sum_{i\in[1\ldots n]} b_i \leq 1$;
- $\sum_{i\in[1\ldots n]} u_i \leq n$;
- $\sum_{i\in[1\ldots n]} d_i \leq n - 1$;
- $b_i + d_i + u_i = 1, \forall i \in [1, \ldots, n]$.

The anonymity degree from Chapter 5 Section 5.4 associated with the evidential opinion matrix is redefined here by considering the special case when the belief mass is computed for each singleton as:

$$D = -\sum_{i=1}^{n} b_i \log_2(1 - d_i).$$

[2]The evidence based metric is defined in Chapter 5, Section 5.4

7.2.2 The Probabilistic Opinion Matrix

The probabilistic opinion matrix is interpreted as follow:

$$W''_{n,4} = \begin{array}{c} \\ s_1 \\ s_2 \\ \vdots \\ s_n \end{array} \begin{array}{c} \begin{array}{cccc} b_i & d_i & u_i & a_i \end{array} \\ \begin{pmatrix} p_1 & 1-p_1 & 0 & 0 \\ p_2 & 1-p_2 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ p_n & 1-p_n & 0 & 0 \end{pmatrix} \end{array}$$

Properties 7.2. The properties of a probabilistic opinion matrix are:

- p_i denote the probability that the subject s_i send the message;
- $\sum_{i\in[1,\ldots,n]} b_i = 1$;
- $\sum_{i\in[1,\ldots,n]} d_i = n - 1$.

The anonymity degree is computed from the probabilities assignments using the formula for the normalized entropy based metric in Section 5.3.2.

7.2.3 Analysis

The proposed matrices are included in the subjective opinion matrix as special cases such that:

- The subjective opinion matrix is equal to an evidential opinion matrix if

  $$\forall i \in [1,\ldots,n], s_i = \sum_{\substack{j=1 \\ j\neq i}}^{n} r_j,$$

  where r_i and s_i are respectively the positive and negative evidence that supports the claim that the subject i is the potential sender of the message M;
- The subjective opinion matrix is equal to a probabilistic opinion matrix if both these conditions are satisfied:
  1. the subjective opinion matrix is an evidential opinion matrix;
  2. $u = 0$.

7.3 Feature Evaluation

In this section, we propose examples according to each feature in **Table 7.1** and we check if these metrics; Normalized entropy based metric, Evidence based metric and Subjective logic based metric supports the stated features. We score each case by 1 if the metric fulfil the feature or by 0 otherwise then we compare the metrics.

Suppose we have an anonymity set of cardinality n and a group of adversaries that evaluate individually which subject is the sender of a message M and each adversary propose an opinion about each subject. Then we combine the opinions together into a matrix, the anonymity degree must decrease when the adversaries agrees together on one subject as the potential sender or if the adversaries information complement each other (which means an adversary detect a potential sender and other adversaries eliminates the rest of the suspected subjects). However, when each adversary have an opinion about each subject that states that the later is the sender of M, the anonymity degree will increase.

Feature F1: "the anonymity degree is lower when the adversaries agrees and higher otherwise."

- **The subjective logic based metric** is associated with an error estimation measure and the anonymity degree increases when the error estimation rate is non-zero value implying that feature F1 is fulfilled;
- **The normalized entropy based metric** is applied from the perspective of one adversary only implying that feature F1 is not fulfilled;
- **The evidence based metric** is applied from the perspective of one adversary implying that feature F1 is not fulfilled.

Feature F2: "An anonymity metric must have a well defined range between two end points."

- **The subjective logic based metric** have a defined range that varies between 0(no anonymity) and 1(perfect anonymity) implying that feature F2 is fulfilled;
- **The normalized entropy based metric** have also a defined endpoints 0 and 1 implying that feature F2 is fulfilled;
- **The evidence based metric** is bounded by the plausibility measure $E(m)$ and the belief measure $C(m)$ [3] therefore it does not have defined endpoints implying that feature F2 is not fulfilled.

Feature F3: "The anonymity degree is maximum when the distribution is uniform."

[3]See definition in Chapter 5 Section 5.4

- **The subjective logic based metric** quantifies the anonymity degree using the standard deviation of the expectation probabilities therefore the anonymity degree is maximum when the distribution of the expectation probabilities is uniform implying that feature F3 is fulfilled;
- **The normalized entropy based metric** uses the entropy of a probability distribution as a measure therefore the anonymity degree is maximum when this distribution is uniform implying that feature F3 is fulfilled;
- **The evidence based metric** uses the entropy of the belief masses distribution to quantify anonymity level hence the anonymity level is maximum when this distribution is uniform implying that feature F3 is fulfilled;

Feature F4: "Anonymity metric results should strive for being objective."

- **The subjective logic based metric** uses the standard deviation that is applied for a normalized expectation probabilities distribution to quantify the anonymity degree implying that feature F4 is fulfilled;
- **The normalized entropy based metric** is applied for a probability distribution such that all the probabilities must sum up to 1 implying that feature F4 is not fulfilled;
- **The evidence based metric** include an element that eliminates insignificant evidence in its measure implying that feature F4 is fulfilled.

Feature F5: "A realistic anonymity metric base its analysis on evidence."

- **The subjective logic based metric** is equipped with a complete model based on evidence implying that feature F5 is fulfilled;
- **The normalized entropy based metric** bases its measure on a predefined probability distribution implying that feature F5 is not fulfilled;
- **The evidence based metric** uses the body of evidence as a basis to compute the belief masses implying that feature F5 is fulfilled.

7.3.1 Results

We present the score chart in table 7.2 and we observe that our metric support all the stated features.

Table 7.2: Score chart

Metric	F1	F2	F3	F4	F5
Subjective logic based metric	✓	✓	✓	✓	✓
Normalized entropy based metric	✗	✓	✓	✗	✗
Evidence based metric	✗	✗	✓	✓	✓

7.4 Summary

We provided a realistic metric model that can be applied for a variety of situations, e.g. collaboration of adversaries, trust and exchange of information. The proposed metric is a generalization of entropy based metric and evidence based metric, and we proposed the error estimation measure in addition to the anonymity degree in order to assess the collected information.

Chapter 8

Application

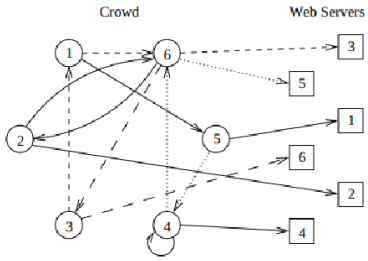
In this chapter, we simulate attack scenarios on the anonymous communication system Crowds. Crowds is selected because it is a well established anonymous communication system, simple and easy to implement. Then, we use the subjective logic based metric to evaluate anonymity against these attack scenarios. Also, we state some evidence that can be collected using some attack models and we apply the metric in the context of these attacks.

8.1 Crowds in a Nutshell

Crowds is a peer-to-peer communication system proposed by Reiter and Rubin [60] that preserves the anonymity of its users while browsing a web server. The anonymity of the communicators is protected by blending them into a larger group. The communicators in this system are named jondos and the original sender of a message is the initiator of the communication. In order to hide the identity of the initiator, the message is forwarded depending on a known probability, either to a randomly selected jondo, or directly to the destination as shown in the figure 8.1.

Figure 8.1: Crowds

Source: Michael K. Reiter and Aviel D. Rubin [60]



8.2 Methodology

There are two available attacker models for Crowds protocol; the global and the local adversary. It is known that Crowds is totally vulnerable to the global adversary since there is no encryption or mixing stations for the transferred messages [60], therefore we will focus here on the local adversary model. The local adversary can be classified as:

1. single attacker;
2. colluding attackers;
3. curious but passive;
4. active.

In order to identify the initiator of the message M, the attacker needs to collect evidence. The evidence are information that prove to a certain degree which jondo can be the initiator of the message. Therefore, evidence are retrieved by a single attacker or a group of attackers and according to each situation evidence are combined by subjective logic operators [1].

Evidence can be an information leaked from the system which is accumulated over a time span or an information exchanged by attackers. Thus, an evidence can be derived from a vulnerability in the system design or from the collaboration of attackers. We summarize some of the evidence we use to measure the anonymity level of Crowds system.

- Predecessor: the jondo who forwards the message to the attacker.
- Packets: the collected items that an attacker links to the jondos.

[1] Subjective logic operators are mentioned in Chapter 4 Section 2, and the metric flow chart 6.3 shows the suitable situations where subjective logic operators are applied in the metric context

- Linkability: the ability to link messages together according to the contents since they are not encrypted.
- Time stamps: the time when a packet is received from a jondo.
- Collaboration and trust: exchanging information between different jondos by considering their loyalty to each other.

8.3 Single Attacker

We model the single adversary as a curious jondo that uses Crowds. We assume that this jondo receives a message M or a set of linkable messages with similar content. The jondo tries to link the message to one of the remaining jondos. In this section, we suppose that the jondo works alone and does not collaborate with another jondo. We establish this assumption to compare the single status with the collaborative status of attackers and to demonstrate that the transition from the single status to the collaborative status leads to more effective and powerful attacks.

8.3.1 Attack Models

The attack analysis described here is based on the Predecessor attack by Panchenko and Pimenidis [55] and the probabilistic model checking of anonymous systems by Vitaly Shmatikov [64]. We summarize in table 8.1 the values used to study the system.

Table 8.1: Variables used in Crowds analysis

Variable	Description
p_f	Probability to forward a message to another jondo
n	Number of honest jondos
c	Number of corrupted jondos
t	Time interval
λ_i	The estimated arrival rate per honest jondo $i \in [1, \ldots, n]$

8.3.1.1 Path Length

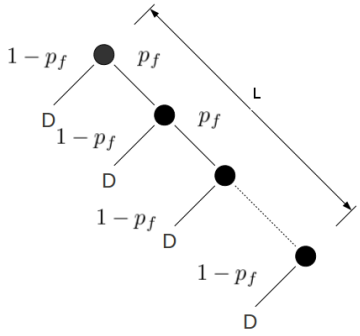
An attacker can learn some information about the length of the path from the probability of forwarding p_f which is a known system parameter. The path length is the number of nodes to forward the message M before it reaches the final destination. However, a message can be forwarded multiple times by the same jondo, so the node represent the number of occurrence of M in each jondo on the forwarding path. Thus the attacker can place a reasonable bet on the path length. We define a discrete random variable L such as:

$$L = \text{the number of nodes to forward } M.$$

Let $\{l_1, l_2, \ldots, l_r\} \subset \mathbb{N}$ be the possible values that L can take as the length of the path and let $P(L = l_j)$ for $1 \leq j \leq r$ be the probability distribution of L.

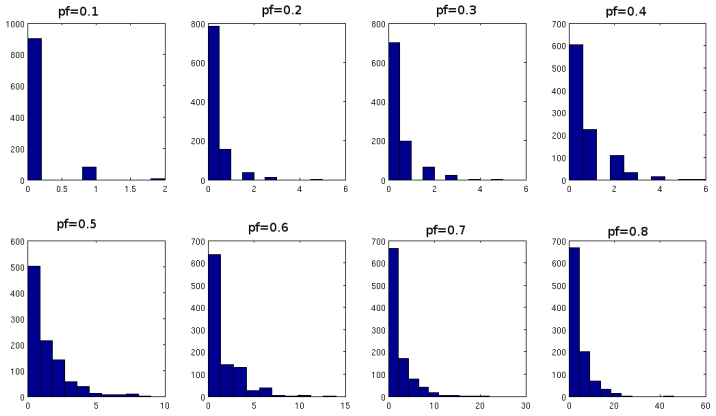
We set up a small experiment to learn about which value of l_j is most likely to occur according to the probability of forwarding p_f. If the size of the crowd is n and the length $l_j > n$ then it is most certainly that some jondos forwarded the same message M multiple times, and a node represent each time M is forwarded to another jondo. So, we model the path as a repetitive Bernoulli process with $p = p_f$ as in figure 8.2.

Figure 8.2: Path formulation



So we ran a small simulation based on the described model and we set some seeds to generate random probabilities then we compute the length of the forwarding path for some probabilities of forwarding independently from the size of the crowds. We use the histogram as a representation of the number of occurrence of a path length as in figure 8.3.

Figure 8.3: Path length histogram



The path length represented by 0 means that the message M is forwarded directly to the end server and we notice from the histogram representation that the number of occurrence differ for each length variable. Therefore, the attacker can place a reasonable bet whether the predecessor of M is the initiator according to the probability of forwarding value. Therefore, if the probability of forwarding is low then the predecessor of M is most likely to be the initiator.

8.3.1.2 Predecessor

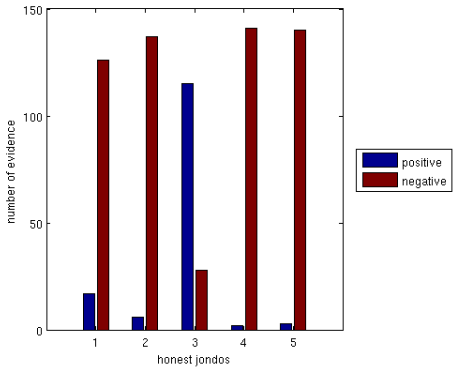
It has been proved that Crowds does not provide perfect anonymity when $c > 0$ [60, 55], therefore the predecessor of M is most likely to be the initiator of M even in a large crowd. The purpose of the experiment is to accumulate a series of linkable packets and since some information is leaked in the system, this information can be collected over a time span using fusion operators. Thus, the longer period the attacker observe the communication, the more chances he have to identify the initiator. Also the period of observation of the traffic can be linked to the active state of the initiator so if he/she is an active user of Crowds then the attacker will deduce the identity of the user in a shorter span of time.

Example 8.1. We assume that the evidence are collected by a single jondo controlled by the attacker Bob. In the context of this example, we set some arrival rates adjusted to the observations made by Bob. Also, we consider that the path is dynamic, therefore the path changes every time the initiator send a linkable message. Let $n = 5$ be the number of honest jondos and let λ_{Bob} be the vector of arrival rates such that:

$$\lambda_{Bob} = \begin{matrix} s1 & s2 & s3 & s4 & s5 \\ \big(0.11 & 0.04 & 0.8 & 0.01 & 0.02 \big). \end{matrix}$$

The arrival rates are set for each honest jondo and they describe the packets arriving from an honest jondo i to the attacker Bob, the honest jondo i is considered as the predecessor. According to the arrival rates, Bob accumulates packets received from each honest jondo within a time span using the cumulative fusion operator [2]. Furthermore, we consider that the packets collected by Bob have similar content, so that there is linkability between the packets and the same origin. The number of collected packets from a predecessor i are considered as positive evidence that supports the claim that i is the possible initiator of the message M. In opposition, the observation that M was not forwarded by i is marked as a negative evidence.

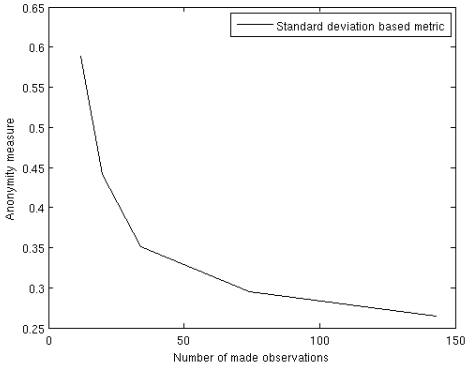
Figure 8.4: Number of accumulated evidence at the corrupted jondo from each honest jondo



The corrupted jondo observes the communication and computes the rate of linkable packets arriving from each honest jondo. As a result, Bob accumulates the linkable packets and he notices in Figure 8.4 that the jondo 3 have a higher rate of forwarding the linkable packets to him.

[2]Chapter 4 Section 4.2.1.1

Figure 8.5: Anonymity vs number of observations



Based on the observations made by Bob, we compute the belief and disbelief masses for each jondo and we measure the anonymity degree. We conclude from figure 8.5 that the anonymity degree decreases as the number of observations becomes higher.

8.4 Collaboration and Trust

In this section, we assume that the attacker can create more accounts related to him in Crowds or that he can collaborate with other attackers. If the corrupted jondos are controlled by different attackers, then two possibilities may occur in the collaboration:

- The collaborators trust each other, therefore fusion operators are deployed to combine evidence;
- The collaborators mistrust each other, in this case the opinions are combined using:
  1. The consensus operator if conflicting opinions were suggested;
  2. The discounting operator according to one of this cases:
     (a) Uncertainty favouring: If the adversary knows that the opinion is suggested from an adversary that lacks information about the opinion state;
     (b) Opposite belief favouring: If one or some of the adversaries provide the opposite of the correct information;
     (c) Base rate sensitive: If there is no judgement about the adversary.

We consider in the rest of this section that the attackers trust each others and we leave the mistrust case as a proposition for future work.

8.4.1 Collaboration Mass

We evaluate the anonymity level of the system according to the number of colluding jondos, providing that the chances to reveal the initiator of the message M increases when the number of colluding jondos increases. Let n be the number of the jondos in Crowds and let $c < n$ be the number of colluding jondos, we interpret data in the following matrices; the subjective opinion matrix W and the probabilistic opinion matrix W' such that:

$$W_{n,4} = \begin{pmatrix} 0 & 0 & 1 & \frac{1}{n-c} \\ 0 & 0 & 1 & \frac{1}{n-c} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 1 & \frac{1}{n-c} \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 1 & 0 & 0 \end{pmatrix} \quad W'_{n,4} = \begin{pmatrix} \frac{1}{n-c} & \frac{n-c-1}{n-c} & 0 & 0 \\ \frac{1}{n-c} & \frac{n-c-1}{n-c} & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ \frac{1}{n-c} & \frac{n-c-1}{n-c} & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

The anonymity degree is computed as follow:

1. $d_1 = 1 - \sqrt{\frac{c}{(n-1)(n-c)}}$, according to the subjective logic based anonymity metric defined in Chapter 6 Section 6.5;
2. $d_2 = \frac{\log_2(n-c)}{\log_2 n}$, according to the normalized entropy based anonymity metric defined in Chapter 5 Section 5.3.2.

We measure the anonymity degree using two approaches variance which represent our metric and entropy based metric. We have considered the number of colluding jondos as evidence, and we notice that when the number of collaborating jondos increases until $c = n - 1$ then we reach a global adversary profile, so in this case the initiator is provably exposed.

Figure 8.6: Anonymity vs number of collaborating jondos

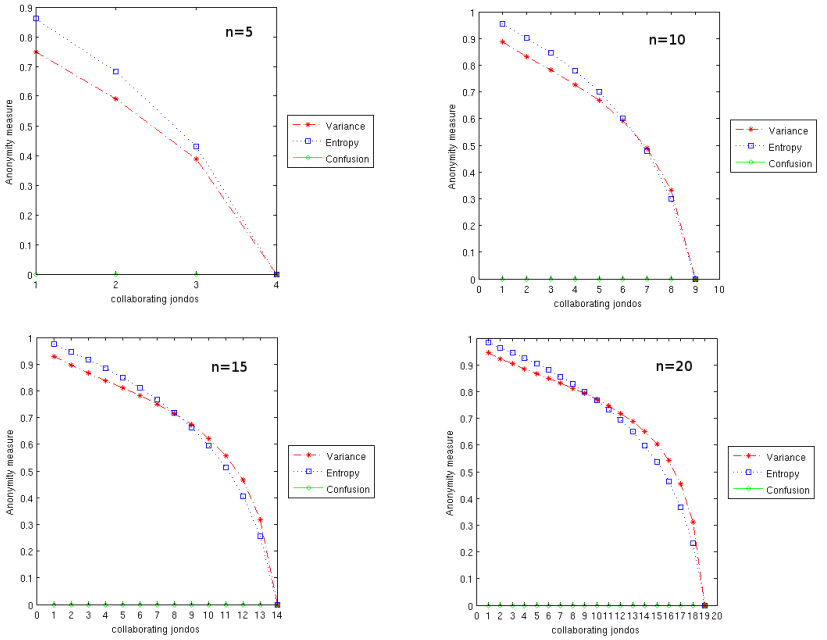


Figure 8.6 shows that the anonymity degree decreases when the number of the corrupted jondos in the system increases.Therefore, the anonymity degree should be higher when the number of honest jondos is higher.

8.4.2 Exchange of Information

In this section, we illustrate a model of collaboration between attackers by exchanging time stamps of the observations of the message M. We assume that attackers synchronize their clocks and each attacker records the time when he receives M from a predecessor. Therefore, attackers can track the source of M and the predecessor detected earlier is the probable initiator of M.

In addition to time tags, the anonymity of the initiator is affected by the mass of collaborating adversaries and the probability of forwarding p_f. A lower probability of forwarding respectively a higher probability of forwarding will require a lower mass respectively a higher mass of collaborating attackers in order to reveal the initiator of M.

8.5 Summary

We analysed the vulnerabilities of Crowds system and we presented a set of evidence that could be used to evaluate the anonymity provided by Crowds.

Further, Crowds can be evaluated according to a local adversary or a group of adversaries collaborating together in order to expose the identity of the initiator. The attacks based on the collaboration of attackers are more efficient than a single adversary profile.

Chapter 9

Anonymity Visualization

A metric is an abstraction of the empirical relations in the real world. Therefore, the measure describes the anonymity state of an anonymous communication system and it should be reflected when the opinions about the subjects in the anonymity set are projected in the opinion space. For this purpose, we used visualization in the opinion space to confirm some of the metric results.

9.1 Cluster Structure

We use Euclidean distance to examine the structure of the clusters formed by the subjects in the anonymity set. A subject's anonymity can be rated according to 4 states: Exposed, possible innocence, probable innocence and beyond suspicion according to the scale of individual anonymity metric[60]. We redefine the anonymity states in table 9.1, and we set some boundaries for each state and to confirm the state for each user we use pair-wise Euclidian distances between the subjects which are represented each by an opinion vector in a $3D$ space.

Definition 9.1. If the subjects in the anonymity set form a cluster in the opinion space such that all the pairwise Euclidean distances are equal to zero then the system is perfectly anonymous .

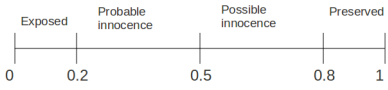
Table 9.1: Anonymity states

State	Description
Exposed	The attacker can demonstrate with a high certainty that a specific subject is the sender of the intercepted message M.
Probable innocence	The attacker detect a suspect among the subjects in the anonymity set but there exists same amount of belief and disbelief masses that this suspect can be the sender of M.
Possible innocence	If for a given suspect there is a non trivial belief mass that another subject in the anonymity set can be the sender of M.
Preserved	All the subjects in the anonymity set are perceived by the attacker as equally likely to send the intercepted message M.

9.2 Projection Example

Let the cardinal of the anonymity set be $n = 5$, we set in this example four opinion matrices with 5 rows where each row represents a subject in the anonymity set and 3 columns where the first column is composed of the belief mass for each user, the second represents the disbelief masses and the third column is composed of the uncertainty masses. Also we define a scale in figure 9.1 for each of the stated anonymity states accordingly to the anonymity degree $d \in [0, 1]$ that describes the level of anonymity provided to the users of a communication system from the point of view of an attacker or a collaboration of attackers.

Figure 9.1: Anonymity degree scale



- If $d = 0$ then the sender is provably exposed to the attacker
- If $d = 1$ then the attacker failed to single out the sender from the

group of users.

We set four matrices where each matrix express a state as the following:

1. $WPv \Rightarrow$ Preserved state;
2. $WP \Rightarrow$ Possible innocence state;
3. $WPr \Rightarrow$ Probable innocence state;
4. $WE \Rightarrow$ Exposed state.

$$WPv = \begin{pmatrix} 0 & 0 & 1.0000 \\ 0 & 0.1000 & 0.9000 \\ 0.1000 & 0.0500 & 0.8500 \\ 0.1000 & 0.1000 & 0.8000 \\ 0 & 0.0500 & 0.9500 \end{pmatrix} \quad WP = \begin{pmatrix} 0.3000 & 0.6000 & 0.1000 \\ 0.3000 & 0.6000 & 0.1000 \\ 0.2500 & 0.6000 & 0.1500 \\ 0.1000 & 0.7000 & 0.2000 \\ 0.1000 & 0.8000 & 0.1000 \end{pmatrix}$$

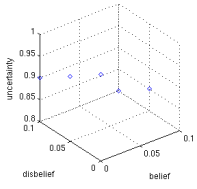
$$WPr = \begin{pmatrix} 0.5000 & 0.3000 & 0.2000 \\ 0.5000 & 0.3000 & 0.2000 \\ 0.1000 & 0.8000 & 0.1000 \\ 0.1000 & 0.8500 & 0.0500 \\ 0.0500 & 0.8000 & 0.1000 \end{pmatrix} \quad WE = \begin{pmatrix} 0.9000 & 0 & 0.1000 \\ 0 & 0.8000 & 0.2000 \\ 0 & 0.7000 & 0.3000 \\ 0 & 0.6000 & 0.4000 \\ 0 & 0.7500 & 0.2500 \end{pmatrix}$$

The empirical order of the anonymity degrees related to each state that is expressed by each matrix in this example is confirmed by the following order of the anonymity degrees computed using the suggested matrices, such that:

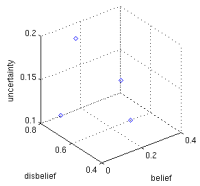
$$0 \leq d(E) < d(Pr) < d(P) < d(Pv) \leq 1,$$

where E, P, Pr and Pv are the expectation vectors computed consecutively from the matrices WE, WP, WPr and WPv.

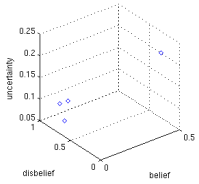
Figure 9.2: Projection in the opinion space



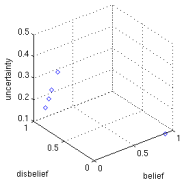
(a) Preserved



(b) Possible innocence



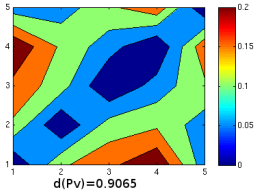
(c) Probable innocence



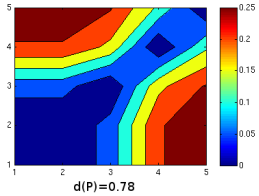
(d) Exposed

Figure 9.2 represent the projection of each subject opinion, which is represented by each row in the opinion matrix, in a three dimensional space according to the belief, disbelief and uncertainty axes.

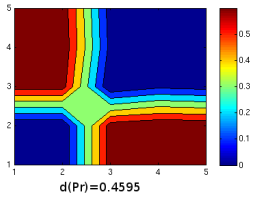
Figure 9.3: Pair-wise Euclidean distance



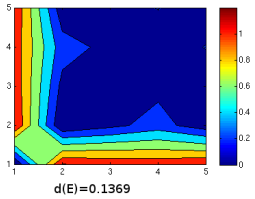
(a) Preserved



(b) Possible innocence



(c) Probable innocence



(d) Exposed

Figure 9.3 is a graphical representation of the pair-wise Euclidean distance between the five subjects in the anonymity set. In the first graph, we cannot single out any subject from the rest of the four subjects and this representation is confirmed with the associated anonymity degree. However, in the last graph, we can clearly single out the first subject 1 from the rest so the anonymity state is exposed in this case, which is again confirmed by the measure result. The second and third graphs represent the degradation of the anonymity from the preserved state to the exposed state. Also we can notice this degradation in the anonymity degree and the graphical representation, as in the possible innocence state three suspects among the five subjects have higher chance that one of them is the actual sender, however in the probable innocence state the suspects are delimited to subject 1 and subject 2 that are isolated from the other subjects.

Chapter 10

Conclusions and Future Work

The proposed metric for anonymity is based on subjective logic where the belief, disbelief and uncertainty masses about each subject in the anonymity set are represented by a binomial opinion assigned by an adversary. The evidence are a basis for the creation of the binomial opinions and the body of subjective logic provides a set of operators that can be deployed to combine evidence according to time constraints or opinions according to trust constraints. Since each opinion is mapped to its own expectation probability, we compute the anonymity degree using the standard deviation of the normalized expectation probabilities. The latter anonymity degree is complemented by an error estimation measure that detects the overflow or lack of evidence in the opinions.
Our contribution relies on the fact that:

1. Our metric outperforms evidence based metric and information theoretic metric, because it supports different adversaries profiles. Our approach is suitable for modelling collaboration between the adversaries and we have suggested using subjective logic operators to handle trust between collaborating attackers. Therefore, Our metric can be applied for various attack scenarios and we believe that by deploying correctly trust operators the quality of shared information can be improved;

2. The metric provides a realistic and accurate assessment of the anonymity of users, since it is based on evidence. Therefore, the metric results can be provable to the users;

3. By using binomial opinions, our approach allows a better interpretation of uncertainty. Further, we notice that the subjects in the anonymity set does not share the same amount of uncertainty as in the evidence based metric.

4. The metric model is complemented by the error estimation rate that controls the flow of information over the anonymity set, Thus the error estimation measure detects the level of confusion and errors when attackers collaborate;

5. Anonymity states can be visualized accordingly to how users are disperse in the opinion space. The dispersion can be analysed using pairwise Euclidean distance between users. Therefore user's anonymity is preserved if the users form a firm cluster in the opinion space, otherwise if one user is singled out from the cluster then he is exposed to the adversary.

For future work, we propose the following:

1. Using the error estimation rate associated with the anonymity measure as an indicator of the trust level of the source of information. Since the error estimation measure can be used to estimate the quality of shared information by monitoring the information flow, it can be accumulated and used as an important parameter to evaluate the trust between the collaborative attackers;

2. Eliminating the erroneous information from the expectation probabilities by detecting the source of this erroneous information instead of using normalization operation for example. This can be achieved by assessing the collected information according to the trust level associated to the source. Therefore, we have a better assessment of anonymity when modelling adversaries collaboration;

3. Modelling and assessing evidence. The proposed metric is a function of two arguments positive and negative evidence. Thus, the metric needs to be deployed correctly by using the two arguments according to examples or applications. So, it would be interesting to build a model that evaluates in term of integers the evidence collected from network or application layers;

4. Evaluating anonymity preserving systems and comparing them. Therefore, it is important to build a basis from which systems can be compared and to demonstrate from the application of the metric the reliability of an anonymity preserving system;

5. Improving the visualization techniques discussed in Chapter 9 with user studies. We believe that visualizing the anonymity states of users and matching these states with the metric results will ease the use and understandability of privacy preserving systems by users.

Bibliography

[1] C. Andersson and R. Lundin. On the fundamentals of anonymity metrics. In S. Fischer-Hübner, P. Duquenoy, A. Zuccato, and L. Martucci, editors, *The Future of Identity in the Information Society*, volume 262, pages 325–341. IFIP International Federation for Information Processing, Springer, June 2008.

[2] L. O. Anyanwu, J. Keengwe, and G. A. Arome. Anonymity leakage reduction in network latency. In T. M. Sobh, editor, *SCSS (1)*, pages 561–565. Springer, 2008.

[3] K. B. Athreya and S. N. Lahiri. *Measure Theory and Probability Theory*. Springer Texts in Statistics, 2010.

[4] A. Back, U. Moller, and A. Stiglic. Traffic analysis attacks and trade-offs in anonymity providing systems. In I. S. Moskowitz, editor, *Information Hiding*, volume 2137 of *Lecture Notes in Computer Science*, pages 245–257. Springer, April 2001.

[5] R. Bagai. A graphical framework for representing anonymity metrics. *DIMACS Working Group on Measuring Anonymity*, May 2013.

[6] R. Bagai and N. Jiang. Measuring anonymity by profiling probability distributions. In G. Min, Y. Wu, L. C. Liu, X. Jin, S. A. Jarvis, and A. Y. Al-Dubai, editors, *TrustCom*, pages 366–374, Los Alamitos, CA, USA, 2012. IEEE Computer Society.

[7] S. L. Blond, P. Manils, A. Chaabane, M. A. Kaafar, C. Castelluccia, A. Legout, and W. Dabbous. One bad apple spoils the bunch: Exploiting p2p applications to trace and profile tor users. *CoRR*, abs/1103.1518, 2011.

[8] N. Borisov, G. Danezis, P. Mittal, and P. Tabriz. Denial of service or denial of security? how attacks on reliability can compromise anonymity. In *Computer and Communications Security*, pages 92–102, New York, USA, October 29 - November 2 2007. ACM.

[9] C.E.Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423, October 1948.

[10] D. Chaum. Untraceable electronic mail, return addresses and digital pseudonyms. *Communications of the Association for Computing Machinery*, 24(2):84–88, February 1981.

[11] D. Chaum. Security without identification: Transaction systems to make big brother obsolete. *Communications of the ACM*, 28(10):1030–1044, October 1985.

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

[13] S. Clauß and S. Schiffner. Structuring anonymity metrics. In A. Juels, M. Winslett, and A. Goto, editors, *Digital Identity Management*, pages 55–62. ACM, 2006.

[14] G. Danezis and I. Goldberg. Sphinx: A compact and provably secure mix format. In *IEEE Symposium on Security and Privacy*, pages 269–282. IEEE Computer Society, May 2009.

[15] G. Danezis and L. Sassaman. How to bypass two anonymity revocation schemes. In N. Borisov and I. Goldberg, editors, *Privacy Enhancing Technologies*, volume 5134 of *Lecture Notes in Computer Science*, pages 187–201. Springer, 2008.

[16] G. Danezis and P. Syverson. Bridging and fingerprinting: Epistemic attacks on route selection. In N. Borisov and I. Goldberg, editors, *Privacy Enhancing Technologies*, volume 5134 of *Lecture Notes in Computer Science*, pages 151–166. Springer, 2008.

[17] C. Díaz. Anonymity metrics revisited. In S. Dolev, R. Ostrovsky, and A. Pfitzmann, editors, *Anonymous Communication and its Applications*, number 05411 in Dagstuhl Seminar Proceedings, Dagstuhl, Germany, 2006. Internationales Begegnungs- und Forschungszentrum für Informatik (IBFI), Schloss Dagstuhl, Germany.

[18] C. Díaz, S. Seys, J. Claessens, and B. Preneel. Towards measuring anonymity. In R. Dingledine and P. F. Syverson, editors, *Proceedings of the Workshop on Privacy Enhancing Technologies*, volume 2482 of *Lecture Notes in Computer Science*, pages 54–68. Springer, April 2002.

[19] C. Díaz, C. Troncoso, and G. Danezis. Does additional information always reduce anonymity? In P. Ning and T. Yu, editors, *WPES*, pages 72–75. ACM, October 2007.

[20] R. Dingledine, N. Mathewson, and P. Syverson. Tor: The second-generation onion router. In *USENIX Security Symposium*, pages 303–320. USENIX, August 2004.

[21] M. Edman, F. Sivrikaya, and B. Yener. A combinatorial approach to measuring anonymity. In *Intelligence and Security Informatics*, pages 356–363. IEEE, 2007.

[22] S. Fischer-Hübner. *IT-Security and Privacy - Design and Use of Privacy-Enhancing Security Mechanisms*, volume 1958 of *Lecture Notes in Computer Science*. Springer, 2001.

[23] B. Gierlichs, C. Troncoso, C. Díaz, B. Preneel, and I. Verbauwhede. Revisiting a combinatorial approach toward measuring anonymity. In V. Atluri and M. Winslett, editors, *WPES*, pages 111–116, New York, USA, 2008. ACM.

[24] I. Goldberg. Privacy-enhancing technologies for the internet, II: Five years later. In *Privacy enhancing technologies*, pages 1–12. Springer, 2002.

[25] I. Goldberg. Privacy enhancing technologies for the internet III: Ten years later. In A. Acquisti, S. Gritzalis, C. Lambrinoudakis, and S. D. C. di Vimercati, editors, *Digital Privacy: Theory, Technologies and Practices*, pages 3–18. Auerbach Publications, 2007.

[26] I. Goldberg, D. Wagner, and E. A. Brewer. Privacy enhancing technologies for the internet. In *Proceedings Compcon'97*, pages 103–109. IEEE, February 1997.

[27] X. Hong, J. Kong, and M. Gerla. Mobility changes anonymity: New passive threats in mobile ad-hoc networks. *Wireless Communications and Mobile Computing*, 6(3):281–293, 2006.

[28] N. Hopper, E. Y. Vasserman, and E. Chan-Tin. How much anonymity does network latency leak? *ACM Transactions on Information and System Security*, 13(2):1–28, February 2010.

[29] D. Huang. On measuring anonymity for wireless mobile ad-hoc networks. In *Local Computer Networks*, pages 779–786. IEEE Computer Society, 2006.

[30] R. Jansen and R. Beverly. Toward anonymity in delay tolerant network: Threshold pivot scheme. In *MILITARY COMMUNICATIONS CONFERENCE*, pages 587–592, 31 October-3 November 2010.

[31] A. Jøsang and S. Knapskog. A metric for trusted systems. In *Proceedings of the 21st National Security Conference NSA*, 1998.

[32] A. Jøsang, F. V. Laenen, S. Knapskog, and J. Vandewalle. How to trust systems. In *Proceedings of the 1997 IFIP-SEC International Information Security Conference*, 1997.

[33] A. Jøsang. Artificial reasoning with subjective logic. In *Proceedings of the 2nd Australian Workshop on Common sense Reasoning*, December 1997.

[34] A. Jøsang. An algebra for assessing trust in certification chains. In *Network and Distributed Systems Security*. The Internet Society, 1999.

[35] A. Jøsang. A logic for uncertain probabilities. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 9(3):279–311, June 2001.

[36] A. Jøsang. The consensus operator for combining beliefs. *Artificial Intelligence Journal*, 141(1/2):157–170, October 2002.

[37] A. Jøsang. Subjective evidential reasoning. In *Information Processing and Management of Uncertainty in Knowledge-Based Systems*, pages 1671–1678, July 2002.

[38] A. Jøsang. Probabilistic logic under uncertainty. In J. Gudmundsson and J. C. Barry, editors, *CATS*, volume 65 of *CRPIT*, pages 101–110. Australian Computer Society, January 2007.

[39] A. Jøsang. Trust and reputation systems. In A. Aldini and R. Gorrieri, editors, *Foundations of Security Analysis and Design*, volume 4677 of *Lecture Notes in Computer Science*, pages 209–245, Bertinoro, Italy, September 2007. Springer.

[40] A. Jøsang. Abductive reasoning with uncertainty. In *Proceedings of the International Conference on Information Processing and Management of Uncertainty*, IPMU '08, June 2008.

[41] A. Jøsang. Conditional reasoning with subjective logic. *Multiple-Valued Logic and Soft Computing*, 15(1):5–38, 2008.

[42] A. Jøsang. *Subjective logic, Draft.* University of Oslo, October 2012.

[43] A. Jøsang, J. Diaz, and M. Rifqi. Cumulative and averaging fusion of beliefs. *Information Fusion*, 11(2):192–200, April 2010.

[44] A. Jøsang, R. Hayward, and S. Pope. Trust network analysis with subjective logic. In *Australasian Computer Science Conference*, volume 48 of *CRPIT*, pages 85–94. Australian Computer Society, Inc., January 2006.

[45] A. Jøsang, S. O'Hara, and K. O'Grady. Base rates for belief functions. In *Workshop on the Theory on Belief Functions*, April 2010.

[46] A. Jøsang, S. Pope, and M. Daniel. Conditional deduction under uncertainty. In L. Godo, editor, *European Conference on Symbolic and Quantitative Approaches*, volume 3571 of *Lecture Notes in Computer Science*, pages 824–835. Springer, 2005.

[47] A. Jøsang, S. Pope, and M. Daniel. Conditional deduction under uncertainty. In L. Godo, editor, *European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, volume 3571 of *Lecture Notes in Computer Science*, pages 824–835. Springer, July 2005.

[48] A. Jøsang, S. Pope, and D. McAnally. Normalising the consensus operator for belief fusion. In *Information Processing and Management of Uncertainty*, 2006.

[49] D. J. Kelly, R. A. Raines, M. R. Grimaila, R. O. Baldwin, and B. E. Mullins. A survey of state-of-the-art in anonymity metrics. In S. Antonatos, M. Bezzi, E. Boschi, B. Trammell, and W. Yurcik, editors, *Network Data Anonymization*, pages 31–40, New York, USA, 2008. ACM.

[50] D. Kesdogan, J. Egner, and R. Büschkes. Stop-and-go MIXes: Providing probabilistic anonymity in an open system. In D. Aucsmith, editor, *Information Hiding*, volume 1525 of *Lecture Notes in Computer Science*, pages 83–98. Springer, 1998.

[51] J. Kong, X. hong, M. Y. Sanadidi, and M. Gerla. Mobility changes anonymity: Mobile ad-hoc need efficient anonymous routing. In *ISCC*, pages 57–62. IEEE Computer Society, 2005.

[52] Z. Ling, J. Luo, and M. Yang. An evidence theory based anonymity metrics. In *First IEEE International Conference on Ubi-Media Computing*, pages 82–87, July 2008.

[53] L. A. Martucci. *Identity and Anonymity in Ad-Hoc Networks*. PhD thesis, Karlstad University, 2009.

[54] S. J. Murdoch. Covert channel vulnerabilities in anonymity systems. Technical report, University of Cambridge, December 2007.

[55] A. Panchenko and L. Pimenidis. Crowds revisited: Practically effective predecessor attack. In *12th Nordic Workshop on Secure IT-Systems*, October 2007.

[56] J.-H. Park, Y.-H. Jung, K.-H. Lee, H. Ko, and M.-S. Jun. A new privacy scheme for providing anonymity technique on sensor network. In *International Conference on Ubiquitous Computing and Multimedia Applications*, volume 0, pages 10–14, Los Alamitos, CA, USA, 2011. IEEE Computer Society.

[57] A. Pfitzmann and M. Hansen. Anonymity, unobservability and pseudonymity a proposal for terminology. In H. Federrath, editor, *Designing Privacy Enhancing Technologies*, volume 2009 of *Lecture Notes in Computer Science*, pages 1–9. Springer, 2001.

[58] J.-F. Raymond. Traffic analysis: Protocols, attacks, design issues and open problems. In H. Federrath, editor, *Workshop on Design Issues in Anonymity and Unobservability*, volume 2009 of *Lecture Notes in Computer Science*, pages 10–29. Springer, 25-26 July 2000.

[59] D. Rebollo-Monedero, J. Parra-Arnau, C. Díaz, and J. Forné. On the measurement of privacy as an attacker's estimation error. *Int. J. Inf. Sec.*, abs/1111.3567(2):129–149, 2013.

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

[61] A. Serjantov and G. Danezis. Towards an information theoretic metric for anonymity. In R. Dingledine and P. F. Syverson, editors, *Privacy Enhancing Technologies*, volume 2482 of *Lecture Notes in Computer Science*, pages 41–53. Springer, April 2002.

[62] G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, 1976.

[63] F. Shirazi, C. Díaz, C. Mullan, J. Wright, and J. Buchmann. Towards measuring resilience in anonymous communication networks. In *6th Workshop on Hot Topics in Privacy Enhancing Technologies*, Bloomington, USA, 12 July 2013.

[64] V. Shmatikov. Probabilistic model checking of an anonymity system. *Journal of Computer Security*, 12(3/4):355–377, 2004.

[65] K. Stokes and V. Torra. N-confusion: A generalization of k-anonymity. In *Proceedings of the 2012 Joint EDBT/ICDT Workshops*, EDBT-ICDT '12, pages 211–215, New York, USA, 2012. ACM.

[66] J. TIAN, C. Li, X. HE, and R. TIAN. A trust model based on the multinomial subjective logic for p2p network. *International Journal of Communications, Network and System Sciences*, 2(6):546–554, 2009.

[67] V. Torra and K. Stokes. A formalization of re-identification in terms of compatible probabilities. *CoRR*, abs/1301.5022, 2013.

[68] G. Tóth and Z. Hornák. Measuring anonymity in a non-adaptive real-time system. In *Privacy Enhancing Technologies*, volume 3424 of *Lecture Notes in Computer Science*, pages 226–241, 2004.

[69] A. F. Westin. *Privacy and Freedom*. Atheneum, New York, 1970.

[70] B. Zhu, Z. Wan, M. S. Kankanhalli, F. Bao, and R. H. Deng. Anonymous secure routing in mobile ad-hoc networks. In *the 29th Annual IEEE International Conference on Local Computer Networks*, pages 102–108. IEEE Computer Society, 2004.

Avdelning, Institution
Division, Department

ADIT,
Dept. of Computer and Information Science
581 83 Linkping

Datum
Date

2014-02-07

Sprk
Language
☐ Svenska/Swedish
☒ Engelska/English
☐ _____

Rapporttyp
Report category
☐ Licentiatavhandling
☒ Examensarbete
☐ C-uppsats
☐ D-uppsats
☐ vrig rapport
☐ _____

ISBN
-

ISRN
LiU-Tek-Lic–2014:14

Serietitel och serienummer ISSN
Title of series, numbering -

Linkping Studies in Science and Technology

Thesis No. LiTH-IDA/ERASMUS-A–14/001–SE

URL fr elektronisk version

http://XXX

Titel
Title

A Metric For Anonymity Based On Subjective Logic

Frfattare
Author

Asmae Bni

Sammanfattning
Abstract

§ Anonymity metrics have been proposed to evaluate anonymity preserving systems by estimating the amount of information displayed by these systems due to vulnerabilities. A general metric for anonymity that assess the latter systems according to the mass and quality of information learned by an attacker or a collaboration of attackers is proposed here.

The proposed metric is based on subjective logic, a generalization of evidence and probability theory. As a consequence, we proved based on defined scenarios that our metric provide a better interpretation of uncertainty in the measure and it is extended to combine various sources of information using subjective logic operators. Also, we demonstrate that two factors: trust between collaborating attackers and time can influence significantly the metric result when taking them into consideration.

Nyckelord
Keywords Anonymity, trust management, privacy and communication systems