Trust in Crowds: probabilistic behaviour in anonymity protocols

Vladimiro Sassone and Ehab ElSalamouny and Sardaouna Hamadou

School of Electronics and Computer Science
University of Southampton, United Kingdom

Abstract. The existing analysis of the Crowds anonymity protocol assumes that a participating member is either 'honest' or 'corrupted.' This paper generalises this analysis so that each member is assumed to maliciously disclose the identity of other nodes with a probability determined by her vulnerability to corruption. Within this model, the trust in a principal is defined to be the probability that she behaves honestly. We investigate the effect of such a probabilistic behaviour on the anonymity of the principals participating in the protocol, and formulate the necessary conditions to achieve 'probable innocence.' Using these conditions, we propose a generalised Crowds-Trust protocol which uses trust information to achieves 'probable innocence' for principals exhibiting probabilistic behaviour.

1 Introduction

Anonymity protocols often use random mechanisms. It is therefore natural to think of anonymity in probabilistic terms. Various notions of such probabilistic anonymity have been proposed and a recent line of work in the literature explores formalising these notions through information-theoretic concepts (e.g. [1, 4–6, 12, 15, 18]). Such approaches usually assume that participants in the protocol can be partitioned in two classes: *honest* members, who always behave correctly, and *attackers*, who try to break the protocol. Although a clear separation between trustworthy members and attackers makes the analysis easier, it is not a realistic assumption for open and dynamic systems in the era of ubiquitous computing. Indeed, traditional approaches to security base on *authentication* and *roles* are not sufficient in open systems. A promising approach is to base security and privacy decisions on attributes linked to some level of *trust* a principal can provide evidence for. The principals participating in a protocol will in general have individual trust judgements; accordingly, interactions between any two of them are governed by their mutual levels of trust. As an illustrating example, consider the social network of Facebook, where members can require some of their activities or information to be accessible only to members who they *explicitly* accepted as friends. This could easily (and does) give misplaced confidence to Facebook users, and encourages them to share sensitive information with 'trusted' friends, without considering that those friends' security system may just be vulnerable to attacks: even though they would not maliciously reveal a user's privata data, friends provide different levels of vulnerability according to the robustness of their security systems, such as the strength of their passwords, the quality of their anti-viruses, and so on. In other words, at each

interaction with user i, there is a probability t_i that she is *not corrupted* and hence acts honestly, and a corresponding probability $1 - t_i$ that instead she is *corrupted*. Moreover, between any given two interactions with a given user, her state may change from honest to corrupted (e.g., as a result of being infected) and vice versa (e.g., as a result of running an antiviral software). In this paper we postulate such probabilistic behavioural model for principals, and investigate its effect on the security of anonymity protocols such as Reiter and Rubin's Crowds protocol [16].

Crowds allows Internet users to perform anonymous web transactions by sending their messages through a random chain of users participating in the protocol. Each user in the '*crowd*' establishes a path between her and a set of servers by selecting randomly other users to act as routers. The random selection process is performed in such a way that when a user in the path relays a message, she does not know whether or not the sender is the initiator (or originator) of the message, or just another forwarder. Each user only has access to messages routed through her. It is well known that Crowds cannot ensure strong anonymity [3, 16] in presence of corrupted participants; yet, when the number of corrupted users is sufficiently small, it provides a weaker notion of anonymity known as *probable innocence*: informally, a sender is probably innocent if to an attacker she is no more likely to be the originator than not to be.

This paper is to the best of our knowledge the first to investigate the impact on the security of Crowds of principals alternating in probabilistically between honest and corrupt behaviours.

Related work. The research on quantitative approaches to information-hiding has recently become very active and fruit-bearing. Several formal definitions and frameworks have been proposed for reasoning about *secure information flow analysis* (e.g., [7, 8, 19]), *side-channel analysis* (e.g., [13]) and *anonymity*. Our work follows a recent trend in the analysis of anonymity protocols directed to the application of information-theoretic notions (e.g., [1,2,4–6,9,12,15,17,18]), whereby the work closer to the present one are those by Reiter and Ruben [16], Halpen and O'Neill [10], Chatzikokolakis and Palamidessi [3], and a recent paper Hamadou et al [12].

In [16] the authors propose a formal definition of probable innocence predicated over the probability of certain observable events induced by the actions of anonymous users participating in the protocol. They require that the probability of an anonymous user producing any observable to be less than one half. In [10] the authors formalise probable innocence in terms of the adversary's confidence that a particular anonymous event happened, after performing an observation. Their definition requires that the probability of an anonymous event should be at most one half, under any observation. In [3] the authors argue that the definition of [16] makes sense only for systems satisfying certain properties, whilst the definition of [10] depends on the probabilities of anonymous events external to the protocol. Thus they propose a definition of probable innocence that combines both by considering both the probability of producing some observable and the adversary's confidence after the observation.

In [12] the authors first generalise the concepts of probable innocence and relate it to Smith's concept of protocol vulnerability [19]. Instead of just comparing the probability of being innocent with the probability of being guilty, they compare such probabilities against a parameter α. Informally, a protocol is α-probable innocent if for any anony-

mous user the probability of being innocent is less than or equal to α. Then, they extend the definition to deal with the adversary's extra knowledge about the correlation between anonymous events and some observables independent of the protocol. The latter is meant to arise from an independent source such as the environment in which the protocol is executed. The paper shows that the presence of extra knowledge makes probable innocence more difficult to achieve, and quantifies such difficulty.

The main difference between these approaches and the one we present in this paper is that we consider the scenario where each participant in the protocol exhibits honest or malicious behaviours according to a fixed probability. In our opinion, such a scenario is a highly likely in ubiquitous computing. This paper is not intended to propose a new definition of probable innocence; rather, we are interested in studying the impact on the protocol's security of its participants' probabilistic behaviour. To this end, we first extend the scenario of attack by associating to each principal a trust level $t \in [0, 1]$ denoting her robustness against corruption. We then modify the protocol accordingly; rather, than selecting a forwarding node uniformly, the forwarding process is governed by a policy where the probability of selecting a node depends on her trust level. We then establish necessary and sufficient criteria for choosing an appropriate policy of forwarding between members in order to achieve probable innocence. It is important to observe that the trust levels t are parameters representing the real world, and not part of the protocol. However, as will be made clear below, the protocol participants will need to have estimates of them. There are well-studied distributed methods for that, based e.g. on Bayesian analysis (cf. [14]), whilst in the current centralised implementation of Crowds, observation leading to the estimation of t can be made by the mechanism which manages crowd membership, the so-called '*blender*.' We do not cover such issues and the related techniques in the current exposition, as we consider them largely orthogonal and scarcely relevant to the focus of this paper.

Structure of the paper. The paper is organised as follows: in §2 we fix some basic notations and recall the fundamental ideas and properties of the Crowds protocol, including the notion of probable innocence. In §3 we present our first main contribution: Crowds protocol extended with trust information of its participating members; §4 delivers our second main contribution by studying the anonymity provided by the extended protocol and establishing necessary and sufficient conditions for achieving probable innocence.

2 Background

This section describes our conceptual framework and revises the Crowds protocol and its notion of probable innocence. We use capital letters A, B to denote discrete random variables, small letters a, b and calligraphic letters $\mathcal{A}$, $\mathcal{B}$ for their values and set of values, respectively. We denote by $P(a)$ the probability of a and by $P(a, b)$ the *joint probability* of a and b. The *conditional probability* of a given b is defined as

$$P(a \mid b) = \frac{P(a, b)}{P(b)}$$

Bayes' theorem relates the conditional probabilities $P(a \mid b)$ and $P(a \mid b)$ as follows

$$P(a \mid b) = \frac{P(b \mid a)\, P(a)}{P(b)} \tag{1}$$

We consider a framework commonly used in probabilistic approaches to anonymity and information flow (e.g. [5, 11, 15, 19]). This focuses on *total* protocols and programs with one *high level* (or *anonymous*) input A, a random variable over a finite set $\mathcal{A}$, and one *low level* output (observable) O, a random variable over a finite set $\mathcal{O}$. We represent a protocol/program by the matrix of the conditional probabilities $P(o_j \mid a_i)$, where $P(o_j \mid a_i)$ is the probability that the low output is o_j given that the high input is a_i. We assume that the high input is generated according to an *a priori* publicly-known probability distribution. An adversary or eavesdropper can see the output of a protocol, but not the input, and she is interested in deriving the value of the input from the observed output.

2.1 The Crowds protocol

Crowds is a protocol proposed by Reiter and Rubin in [16] to allow Internet users to perform anonymous web transactions, i.e., to protect their identities as originators of request messages. The central mechanism is that the originator forwards the message to a randomly-selected user, which in turn forwards the message to another user, and so on until the message reaches its destination (the end server). This routing process ensures that when a user is detected sending a message, there is a substantial probability that she is not acting for herself but simply forwarding it on behalf of somebody else.

More specifically, a crowd is a *fixed* number of users participating in the protocol. Some members (users) in the crowd may be corrupted (the *attackers*), and they can collaborate in order to discover the originator's identity. The purpose of the protocol is to protect the identity of the message originator from the attackers. When an *originator* – also referred to as *initiator* – wants to communicate with a server, she creates a random *path* between herself and the server through the crowd by the following process.

- *Initial step:* the initiator selects randomly a member of the crowd (possibly herself) and forwards the request to her. We refer to the latter user as the *forwarder*.
- *Forwarding steps:* a forwarder, upon receiving a request, flips a *biased* coin. With probability $1 - p_f$ she delivers the request to the end server or, with probability p_f, she selects randomly a new forwarder (possibly herself) and relays the original request to her, to repeat the forwarding process again.

The response from the server to the originator follows the same path in the opposite direction. Users (including corrupted ones) are assumed to have only access to messages routed through them, so that they only know the identities of their immediate predecessors and successors in the path, and of the destination server.

2.2 Probable innocence

In [16] Reiter and Rubin have proposed a hierarchy of anonymity notions in the context of Crowds. These range from '*absolute privacy*,' where the attacker cannot perceive

the presence of communication, to '*provably exposed*,' where the attacker can prove the sender and receiver relationship to third parties. Clearly enough, Crowds cannot ensure absolute privacy in presence of attackers or corrupted users; it can only provide weaker notions of anonymity. In particular, in [16] the authors propose an anonymity notion called *probable innocence* and prove that, under suitable conditions on the parameters of the protocol, Crowds ensures the probable innocence property to the originator. Informally, they define it as follows:

> *A sender is probably innocent if, from the attacker's point of view, the sender appears no more likely to be the originator than to not be the originator.* (2)

In other words, the attacker may have good reasons to consider the sender more likely than any other user to be the originator, yet it still appears at least as likely that she is not.

Let n be the number of users participating in the protocol and let c and m be the number of the corrupted and honest users, respectively, with $n = m+c$. Since anonymity makes only sense for honest users, we define the set of anonymous events as $\mathcal{A} = \{a_1, a_2, \ldots, a_m\}$, where a_i indicates that user i is the initiator of the message.

As it is usually the case in the analysis of Crowds, we assume that attackers will always deliver a request to forward immediately to the end server, since forwarding it any further cannot help them learn anything more about the identity of the originator. Thus in any given path, there is at most one detected user: the first honest member to forward the message to a corrupted member. We therefore define the set of observable events as $O = \{o_1, o_2, \ldots, o_m\}$, where o_j indicates that user j forwarded a message to a corrupted user. In this case we also say that user j *is detected* by the attacker.

Reiter and Rubin formalise their notion of probable innocence via the conditional probability $P(I \mid H)$ that the initiator is detected given that any user is detected at all. Here H denotes the event that there is an attacker in the path (and thus the user before it will be detected), whilst I is the event that precisely the initiator will forward the message to the attacker.[1] Probable innocence holds if $P(I \mid H) \leq 1/2$.

In our setting the probability that user j is detected given that user i is the initiator, can be written simply as $P(o_j \mid a_i)$. As we are only interested in the case in which a user is detected, for simplicity we do not write such condition explicitly. Therefore, the notion of probable innocence proved in [16] translates in our setting as:

$$P(o_i \mid a_i) \leq \frac{1}{2} \quad \text{for all } i = 1, \ldots, m \tag{3}$$

Reiter and Rubin proved in [16] that the following property holds for Crowds.

$$P(o_j \mid a_i) \;=\; \begin{cases} 1 - \dfrac{m-1}{n} p_f & i = j \\ \dfrac{1}{m} p_f & i \neq j \end{cases} \tag{4}$$

[1] Observe that this does not necessarily mean that the attacker is the second user in the path, as the originator could herself be selected as a forwarder in the path she initiated!

Therefore, probable innocence (3) holds if and only if

$$m \geq \frac{c-1}{p_f - 1/2} p_f \,.$$

As previously noticed in several papers (e.g., [3]), there is a mismatch between the idea of probable innocence expressed informally in (2) and property (3) actually proved by Reiter and Rubin. Indeed, the former seems to correspond to the following interpretation given by Halpern and O'Neill [11]:

$$P(a_i \,|\, o_i) \leq \frac{1}{2} \quad \text{for all } i = 1, \ldots, m \tag{5}$$

Properties (3) and (5) however coincide under the standard assumption in Crowds that the *a priori* distribution is uniform, i.e., that each honest user has equal probability of being the initiator.

Finally we recall that the concept of probable innocence was recently generalised in [12]. Instead of just comparing the probability of being innocent with the probability of being guilty, *loc. cit.* considers, so to say, 'degrees' of innocence. Formally, given a real number $\alpha \in [0, 1]$, a protocol satisfies α-probable innocence if and only if

$$P(a_i \,|\, o_i) \leq \alpha \quad \text{for all } i = 1, \ldots, m \tag{6}$$

Clearly, α-probable innocence coincides with the probable innocence for $\alpha = 1/2$.

3 Using trust information

In the previous section, we have revised the fundamental ideas of the Crowds protocol and its properties under the assumption that each user participating in the protocol is either always honest or always an attacker, and all members are treated equally. However, as observed in §1, this is not a realistic assumption for open and dynamic systems in ubiquitous computing. Indeed, open and dynamic systems often use attributes related to some level of *trust* to enhance security and privacy. In this section we reformulate Crowds under the novel scenario where interaction between users is governed by their level of trust. We then study the effect of such probabilistic principals' behaviour on the security of the protocol.

3.1 Crowds protocol extended

We now extend the Crowds protocol to take into account the trust levels of its participating members. We associate a trust level $t_{ij} \in [0, 1]$ to each pair of users i and j to indicate the trust of user i in user j according to evidence provided by j. Here t_{ij} denotes the probability that when the principal i chooses principal j as a forwarder, j behaves honestly and protects i's identity. Accordingly, each user i defines her *policy of forwarding* to other members (including herself) based on her trust of them. A policy of forwarding for a user i is probability distribution $\{q_{i1}, q_{i2}, \cdots, q_{in}\}$, such that for all i,

$\sum_{j=1}^{n} q_{ij} = 1$. Here q_{ij} denotes the probability that j is chosen as a forwarder by i (given that i has decided to forward the message).

Defining trust as an individual judgement as we did above matches the current assumptions in the research on trust (cf. [14]) and is certainly desirable in general. However for some applications – specifically the Crowds protocol – it is more reasonable to consider a simplified notion where trust in a user is common to everybody. In other words $t_{ij} = t_{kj}$ for all i and k. Indeed, in the case of the Crowds protocol, we want a trust in a user to reflect her robustness to becoming *corrupt* (a.k.a. *infected*). Allowing each member to adopt her own level of trust would make the value of trust subjective and could hardly reflect the user's actual robustness against corruption.

We therefore assume that a trust in a user is shared. Its value could be established cooperatively by the members of the crowd, or by a suitable local authority (e.g., the blender in case of Reiter and Rubin's implementation of Crowds) based on evidence provided by the user. Accordingly, in the rest of the paper, we will simply write t_i to denote the trust level of user i. Similarly, we require the policy of forwarding to be common to all members of the crowds. This means that all participants treat any given user in the same way, as all of them have the same trust in her. We therefore write $\{q_1, q_2, \cdots, q_n\}$ to represent the common forwarding policy.

Under these assumptions, we extend the protocol. When an initiator wants to communicate with a server, she creates a random *path* between herself and the server through the crowd by the following process.

- *Initial step:* With probability q_j the initiator selects a member j of the crowd (possibly herself) according to the policy of forwarding $\{q_1, q_2, \cdots, q_n\}$ and forwards the request to her. We refer to the latter user as the *forwarder*.
- *Forwarding steps:* a forwarder, upon receiving a request, flips a *biased* coin. With probability $1 - p_f$ she delivers the request to the end server or, with probability $p_f \cdot q_k$, she selects a new forwarder k (possibly herself) and relays the original request to her, to repeat the forwarding process again.

3.2 Probable innocence revisited

In order to study the anonymity provided by the extended protocol, we first spell out the hypotheses of our analysis. As is the previous section, we assume that corrupted members will always deliver a request immediately to the end server, since forwarding it any further cannot help the attacker learn anything more about the identity of the originator. Consequently, when an infected user initiates a transaction, her message is delivered directly to the end server.[2]

We also assume that server replies are *short*, so that the status of each user in an anonymous paths from users to servers is maintained for the time it takes for the reply to travel back from server to originator. That is, we do not consider the case where users on a given path may switch to become corrupt (or indeed honest) between request and answer, which might happen if the server's replies are very long or very slow. From servers to users so which would normally follow the same paths in reverse direction.

[2] Her anonymity is broken at the start, so there is no need to continue the anonymity protocol.

Under these assumptions, there is always at most one corrupted member on a path, it occupies its last position, and detection always occurs while forwarding a request and not while relaying a reply.[3]

Finally since each user i has probability t_i of being honest when she initiates a request, we extend the set of anonymous events a_i and observable events o_i to the whole set of participating members.

Under these assumption we study the privacy level ensured to each member participating in the protocol, i.e., $P(a_i \mid o_i)$. We remind the reader that by Bayes' theorem (Eq. 1) we have

$$P(a_i \mid o_i) = \frac{P(a_i, o_i)}{P(o_i)} \tag{7}$$

We first evaluate the denominator in the above expression. Let H_k be the event that the first corrupted node in the message path to the server occupies the kth position, where $k \geq 0$. Note that H_0 means that the initiator itself is corrupted.

$$P(o_i, H_k) = \begin{cases} \dfrac{1}{n}(1 - t_i) & k = 0 \\ \dfrac{1}{n} t_i \displaystyle\sum_{j=1}^{n} q_j (1 - t_j) & k = 1 \\ \displaystyle\sum_{j=1}^{n} \frac{1}{n} t_j \left(\sum_{j=1}^{n} q_j t_j \right)^{k-2} \cdot & \\ \qquad q_i t_i \left(\sum_{j=1}^{n} q_j (1 - t_j) \right) \cdot p_f^{k-1} & k \geq 2 \end{cases} \tag{8}$$

The above equation for the case $k \geq 2$ is implied by the fact that the message is initiated by any honest participant, forwarded to $k-2$ honest principals before it is passed to the detected principal i, and finally to a corrupted one. For convenience, we will write T for $\sum_{j=1}^{n} q_j t_j$ and S for $\sum_{j=1}^{n} t_j$. Since the joint events $\{o_i, H_k\}$, for $k \geq 0$ are mutually exclusive, we evaluate $P(o_i)$ as follows.

$$\begin{aligned} P(o_i) &= \sum_{k=0}^{\infty} P(o_i, H_k) \\ &= \frac{1}{n}(1 - t_i) + \frac{1}{n} t_i (1 - T) \\ &\quad + \sum_{k=2}^{\infty} \frac{1}{n} S T^{k-2} \cdot q_i t_i (1 - T) \cdot p_f^{k-1} \\ &= \frac{1}{n} \left(1 - t_i T + S p_f q_i t_i \left(\frac{1 - T}{1 - p_f T} \right) \right) \end{aligned} \tag{9}$$

[3] We are currently working on a refined protocol where this assumption is dropped. This means that there can be users on a path which while not infected in the forward direction, become corrupt by the time they receive the response from the server. Hence they report their predecessor as the detected user.

From Equation (9), it is worth noticing that $P(o_i) = 0$ only if $T = 1$ and $t_i = 1$. Observe that $T = 1$ means that $t_j = 1$ for all participants j where $q_j \neq 0$, i.e., all forwarders are always honest. In this case i is never detected by any forwarder. If moreover $t_i = 1$, the principal i is never detected by herself. Thus in the case where $T = 1$ and $t_i = 1$ the principal i is never detected by any corrupted node.

Now we turn to evaluating the probability $P(a_i, o_i)$ appearing as the numerator in Equation (7). To such purpose, we first formulate the probability $P(a_i, H_k, o_i)$, i.e., the probability that i is the initiator and is also detected by a corrupted node at position k in the message path.

$$P(a_i, H_k, o_i) = \begin{cases} \frac{1}{n}(1 - t_i) & k = 0 \\ \frac{1}{n} t_i \sum_{j=1}^{n} q_j (1 - t_j) & k = 1 \\ \frac{1}{n} t_i \left(\sum_{j=1}^{n} q_j t_j \right)^{k-2} \cdot & \\ \quad q_i t_i \left(\sum_{j=1}^{n} q_j (1 - t_j) \right) \cdot p_f^{k-1} & k \geq 2 \end{cases} \tag{10}$$

Similar to the argument of Equation (8), the formula in the case $k \geq 2$ is implied by the fact that the message is initiated by the principal i, forwarded to $k - 2$ honest principals before it is passed back to i, and finally to a corrupted principal. Since the joint events $\{a_i, H_k, o_i\}$, for $k \geq 0$ are mutually exclusive, we evaluate $P(a_i, o_i)$ as follows.

$$\begin{aligned} P(a_i, o_i) &= \sum_{k=0}^{\infty} P(a_i, H_k, o_i) \\ &= \frac{1}{n}(1 - t_i) + \frac{1}{n} t_i (1 - T) \\ &\quad + \sum_{k=2}^{\infty} \frac{1}{n} t_i T^{k-2} \cdot q_i t_i \, (1 - T) \cdot p_f^{k-1} \\ &= \frac{1}{n} \left(1 - t_i T + p_f q_i t_i^2 \left(\frac{1 - T}{1 - p_f T} \right) \right) \end{aligned} \tag{11}$$

Assuming $P(o_i) \neq 0$, we substitute Equations (9) and (11) in Equation (7), and we therefore get,

$$P(a_i \mid o_i) = \frac{1 - t_i T + p_f q_i t_i^2 \left(\frac{1-T}{1-p_f T} \right)}{1 - t_i T + S \, p_f q_i t_i \left(\frac{1-T}{1-p_f T} \right)} \tag{12}$$

From Equation (12), we observe that for a detectable principal i (i.e., $P(o_i) \neq 0$), it holds that $P(a_i \mid o_i) > 0$. That is, there is always a non zero probability that i is the initiator if she is detected. This confirms that Crowds never achieves the highest degree of anonymity known as *absolute privacy* in [16].

3.3 Provably exposed principals

It would also be interesting to investigate the conditions under which the protocol can only ensure the degree of anonymity known as *provably exposed* to a given principal i. Such a degree, defined in [16], represents the lowest level of anonymity where an attacker can prove the identity of the message initiator. This happens when i is the only possible initiator, given that i is detected, i.e., $P(a_i \mid o_i) = 1$. These conditions are precisely stated by the following proposition.

Proposition 1 (Provably exposed). *For all user i such that $P(o_i) \neq 0$, we have that $P(a_i \mid o_i) = 1$ if and only if one of the following conditions holds:*

- $p_f = 0$;
- $t_i = 0$;
- $q_i = 0$;
- $T = 1$;
- $S = t_i$.

Proof. Solving the following equation $P(a_i \mid o_i) = 1$ using the formula given by Equation (12) yields only the above conditions.

The following paragraphs discuss the meaning of these results. Firstly, we observe that $p_f = 0$ implies that, provided she is not corrupt, the initiator will pick her first forwarder according to the forwarding policy $\{q_1, \cdots, q_n\}$, who then delivers directly the message to the end server, regardless of her being corrupt or not. Thus, in this case a path is always at most of length 2, excluding the end server. Hence, i can only be detected at position 0 (by herself if she is initially corrupted) or at position 1 by her forwarder when the latter is corrupted. Therefore, in both cases, i is the only possible initiator. That is if a principal i is detected, then she must be the initiator.

In the case where $t_i = 0$, i is always corrupted and therefore when she initiates a message, she will detect herself and deliver the message directly to the end server (by assumption). Hence nobody except herself will detect her, and i will be detected if and only if she is the initiator.

Consider the case where $q_i = 0$. This implies that i is never chosen as a forwarder. In this case, i is detected only if she initiates a message and is corrupted at the same time, i.e., she detects herself. Thus, the detection of i implies that i is the initiator.

The case $T = 1$ happens if and only if $t_j = 1$ for all $q_j \neq 0$, which means that only honest members can be chosen as forwarders. In this case too, i is detected only if she originates a message and is corrupted at the same time: she detects herself. Thus, the fact that i is detected, implies that i is the initiator.

Finally, suppose that $S = t_i$. Here $t_j = 0$ for all $j \neq i$, that is all participants other than i are corrupted. In this case if i is detected then it is the only possible initiator because otherwise the initiator would just detect herself at the start of the protocol. Therefore, once again, if i is detected, she must be the initiator.

It is worth noticing that the original Crowds protocol is the protocol obtained by assuming that each principal i is either always honest or always corrupted, i.e., $t_i \in$

{0, 1}, and by choosing a uniform forwarding policy, that is for all j,

$$q_j = \frac{1}{n} .$$

Thus when the number of corrupted principals is c, we have

$$T = \sum_{j=1}^{n} q_j t_j = \frac{n-c}{n} ,$$

and

$$S = \sum_{j=1}^{n} t_j = n - c .$$

By substituting the values of q_j, T and S in Equation (12) for a honest initiator i, i.e., one for which $t_i = 1$, we get

$$P(a_i \mid o_i) = 1 - p_f \left(\frac{n-c-1}{n} \right) .$$

which is the same expression derived in [16] for standard Crowds and given by (4).

4 Achieving probable innocence

For any fixed number of principals n, the extended protocol described in the previous section has three main parameters: the forwarding probability p_f, members' trust values $\{t_1, \cdots, t_n\}$, and the forwarding policy $\{q_1, \cdots, q_n\}$. We study in this section how each of them affect the anonymity of participating members. We begin by the probability of forwarding.

4.1 Probability of forwarding

The following result states that for fixed trust values $\{t_1, \cdots, t_n\}$ and forwarding policy $\{q_1, \cdots, q_n\}$, the probability $P(a_i \mid o_i)$ for any participant i is a monotonically decreasing function with respect to the forwarding probability p_f.

Theorem 1 (Monotonicity). *For all $i = 1, \ldots, n$,*

$$\frac{\partial P(a_i \mid o_i)}{\partial p_f} \leq 0$$

Proof. By differentiating $P(a_i \mid o_i)$ as given by Equation (12) with respect to p_f, we have

$$\frac{\partial P(a_i \mid o_i)}{\partial p_f} = \frac{t_i \, q_i \, (1-T) \, (1 - t_i T) \, (t_i - S)}{\Big((1 - p_f T)(1 - t_i T) + p_f S \, q_i t_i (1-T) \Big)^2} . \tag{13}$$

Given that $0 \leq t_j \leq 1$ for each principal j, and that $T = \sum_{j=1}^{n} q_j t_j$, we have $0 \leq T \leq 1$ and $0 \leq t_i T \leq 1$. We have also $t_i \leq S$, because $S = \sum_{j=1}^{n} t_j$, and therefore

$$\frac{\partial P(a_i \mid o_i)}{\partial p_f} \leq 0 ,$$

i.e., $P(a_i \mid o_i)$ is either fixed or decreasing with respect to p_f.

From Equation (13) above, $P(a_i \mid o_i)$ is fixed irrespectively of p_f if and only if i is always corrupted ($t_i = 0$), i is never used as a forwarder ($q_i = 0$), all forwarders are honest ($T = 1$), or all participants other than i are corrupted ($S = t_i$). It has been shown by Proposition 1 in the previous section that $P(a_i \mid o_i) = 1$ in these cases.

Theorem 1 justifies using a high value of p_f as it decreases the probability of identifying the initiator and therefore enhance her privacy. However, large p_f implies longer message path to the server, and therefore the performance of the protocol is degraded. Thus a trade-off is required for choosing the forwarding probability p_f.

Corollary 1 (Anonymity range). *For all $i = 1, \ldots, n$,*

$$1 \geq P(a_i \mid o_i) \geq 1 - \frac{q_i t_i \sum_{j \neq i}^{n} t_j}{1 - t_i \sum_{j \neq i}^{n} q_j t_j + q_i t_i \sum_{j \neq i}^{n} t_j}$$

Proof. By Theorem 1, and taking into account that $0 \leq p_f \leq 1$, the above range for $P(a_i \mid o_i)$ is obtained by substituting $p_f = 0$ and $p_f = 1$ in Equation (12).

The corollary above describes the range of probabilities that a principal i is the initiator given that i is detected. Observe that with $p_f = 0$ the message is passed directly to the server, and therefore if i is detected, then she must be the initiator and also detected by herself. Taking $p_f = 1$ minimises $P(a_i \mid o_i)$, but in this case the message never reaches the server.

4.2 Trust values

We now turn our focus to the trust values. Observe that the anonymity of a member i, indicated by $P(a_i \mid o_i)$, is affected by the trust values t_j of all participating members. Therefore, the above lower bound can be used as a criterion to decide whether a new member i is accepted to join the network or not based on her trust t_i. For instance, such a criterion can be chosen to achieve the *α-probable innocence* according to the following theorem.

Theorem 2 (α-probable innocence). *Let $\alpha \in [0, 1]$ be a positive value. If for all $i = 1, \ldots, n$*

$$\frac{q_i t_i \sum_{j \neq i}^{n} t_j}{1 - t_i \sum_{j \neq i}^{n} q_j t_j + q_i t_i \sum_{j \neq i}^{n} t_j} \geq 1 - \alpha \ ,$$

then the extended protocol ensures α-probable innocence to all its participating members.

Proof. Results from Corollary 1 and Definition 6.

4.3 Forwarding policy.

We now propose a strategy for choosing a forwarding policy $\{q_1, \cdots, q_n\}$ based on the trust information $\{t_1, \cdots, t_n\}$ in order to achieve α-probable innocence for a given degree

of privacy α. The key idea is that the forwarding probabilities q_j are adjusted depending on the given trust information t_j.

Choosing the forwarding policy q_i for a given user i can then be done by maintaining the lower bounds of $P(a_i \mid o_i)$ below a chosen threshold α, i.e., by achieving α-probable innocence. By Theorem 2 the plausible values of q_i are obtained by solving the following system of linear inequalities.

$$1 - \alpha \;\leq\; \frac{q_i t_i \sum_{j \neq i}^{n} t_j}{1 - t_i \sum_{j \neq i}^{n} q_j t_j + q_i t_i \sum_{j \neq i}^{n} t_j} \qquad 1 \leq i \leq n$$

$$1 \;=\; \sum_{i=1}^{n} q_i$$

Example 1. Consider an instance of Crowds-Trust protocol where three principals are involved. Let the trust values in these principals be:

$$t_1 = 0.70, \qquad t_2 = 0.97, \qquad t_3 = 0.99$$

Solving the above problem for $\alpha = \frac{1}{2}$ yields the two solutions:

$$0.2479 \leq q_2 \leq 0.2620$$
$$1.1411 - 3.4138\, q_2 \leq q_3 \leq 0.5479 - 1.0206\, q_2$$
$$q1 = 1 - q_2 - q_3$$

and

$$0.2620 \leq q_2 \leq 0.3074$$
$$0.3197 - 0.2784\, q_2 \leq q_3 \leq 0.5479 - 1.0206\, q_2$$
$$q_1 = 1 - q_2 - q_3\ .$$

Thus the following forwarding distribution satisfies the $\frac{1}{2}$-probable innocence:

$$q_1 = 0.4575, \qquad q_2 = 0.2620, \qquad q_3 = 0.2805\ .$$

However, if the uniform distribution is used (as in the original Crowds protocol), i.e., $q_1 = q_2 = q_3 = \frac{1}{3}$, probable innocence is not achievable because according to Corollary 1 the minimum value of $P(a_1 \mid o_1)$ is 0.543, which is greater than $\frac{1}{2}$. Note that such sets of constraints are not always solvable, in which case the required level of anonymity cannot be provided to all members.

Observe that the forwarding distribution above increases the frequency at which the less reliable user 1 will be involved in a message path, so as to make it more difficult for an attacker to detect her with a high degree of confidence. The higher security for 1 is of course achieved at the price of a lower overall security for other two, more reliable users, and can therefore considered a 'social' approach to crowds membership. The flexibility of the protocol means that the forwarding policy can be chosen to provide a lower degree of anonymity to a subset of the members to guarantee probable innocence to a larger crowd ('*social strategy*'), or to reject principals having the low trust values who, therefore, exhibit a greater threat to others ('*rational strategy*').

5 Conclusion

In this paper we focused on the Crowds anonymity protocol and asked the question of how its existing analyses are affected by postulating that each principal behaves honestly or becomes corrupt according to a given probability (as opposed to being either honest or malicious once and for all). This amounts to providing each member i of the crowd with a trust level t_i denoting her robustness against corruption, and a preference level of forwarding q_i denoting the probability of choosing her as the next forwarder in the routing process. Given a probability of forwarding p_f, a level of anonymity α, and the trust levels $t_1, t_2, \cdots, t_n$ of the crowd's members, we have identified the conditions on the probability of choosing a forwarder which are necessary to achieve α-probable innocence. Thus, in presence of untrusted members, the protocol users can exploit these results to derive an interaction policy $q_1, q_2, \cdots, q_n$, if any exists, that guarantees a satisfactory level of anonymity; and in doing so, they can act both 'rationally' or 'socially.'

In conclusion, we remark that although the scenario in which members participating in a protocol can exhibit probabilistic behaviours is highly likely in real-world scenarios, this is the first paper to deal with the question in the context of anonymity protocols. In the near future, we expect to tackle even more interesting scenarios, in particular by extending this work to the case where a possibly slow or long response from the server may follow in the reverse direction to the initiator, as the honesty status of the users on the path has changed since the request was forwarded to the server.

References

1. M. Bhargava and C. Palamidessi. Probabilistic anonymity. In M. Abadi and L. de Alfaro, editors, *CONCUR*, volume 3653 of *Lecture Notes in Computer Science*, pages 171–185. Springer, 2005.
2. C. Braun, K. Chatzikokolakis, and C. Palamidessi. Compositional methods for information-hiding. In R. M. Amadio, editor, *FoSSaCS*, volume 4962 of *Lecture Notes in Computer Science*, pages 443–457. Springer, 2008.
3. K. Chatzikokolakis and C. Palamidessi. Probable innocence revisited. *Theor. Comput. Sci.*, 367(1-2):123–138, 2006.
4. K. Chatzikokolakis, C. Palamidessi, and P. Panangaden. Probability of error in information-hiding protocols. In *CSF*, pages 341–354. IEEE Computer Society, 2007.
5. K. Chatzikokolakis, C. Palamidessi, and P. Panangaden. Anonymity protocols as noisy channels. *Inf. Comput.*, 206(2-4):378–401, 2008.
6. K. Chatzikokolakis, C. Palamidessi, and P. Panangaden. On the Bayes risk in information-hiding protocols. *Journal of Computer Security*, 16(5):531–571, 2008.
7. D. Clark, S. Hunt, and P. Malacaria. A static analysis for quantifying information flow in a simple imperative language. *Journal of Computer Security*, 15(3):321–371, 2007.
8. M. R. Clarkson, A. C. Myers, and F. B. Schneider. Belief in information flow. In *CSFW*, pages 31–45. IEEE Computer Society, 2005.
9. Y. Deng, J. Pang, and P. W. 0002. Measuring anonymity with relative entropy. In T. Dimitrakos, F. Martinelli, P. Y. A. Ryan, and S. A. Schneider, editors, *Formal Aspects in Security and Trust*, volume 4691 of *Lecture Notes in Computer Science*, pages 65–79. Springer, 2006.
10. J. Y. Halpern and K. R. O'Neill. Anonymity and information hiding in multiagent systems. *Journal of Computer Security*, 13(3):483–512, 2005.

11. J. Y. Halpern and K. R. O'Neill. Anonymity and information hiding in multiagent systems. *Journal of Computer Security*, 13(3):483–512, 2005.
12. S. Hamadou, C. Palamidessi, V. Sassone, and E. ElSalamouny. Probable Innocence in the presence of independent knowledge. In *To appear in the Proc. of the sixth International Workshop on Formal Aspects in Security and Trust (FAST2009)*, LNCS. Spr.-Ver., 2009.
13. B. Köpf and D. A. Basin. An information-theoretic model for adaptive side-channel attacks. In P. Ning, S. D. C. di Vimercati, and P. F. Syverson, editors, *ACM Conference on Computer and Communications Security*, pages 286–296. ACM, 2007.
14. K. Krukow, M. Nielsen, and V. Sassone. Trust models in ubiquitous computing. *Philosophical Transactions of the Royal Society A*, 366:3781–3793, 2008.
15. P. Malacaria and H. Chen. Lagrange multipliers and maximum information leakage in different observational models. In Ú. Erlingsson and M. Pistoia, editors, *PLAS*, pages 135–146. ACM, 2008.
16. M. K. Reiter and A. D. Rubin. Crowds: Anonymity for web transactions. *ACM Transactions on Information and Systems Security*, 1(1):66–92, 1998.
17. 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, 2002.
18. V. Shmatikov and M.-H. Wang. Measuring relationship anonymity in mix networks. In A. Juels and M. Winslett, editors, *WPES*, pages 59–62. ACM, 2006.
19. G. Smith. On the foundations of quantitative information flow. In L. De Alfaro, editor, *Proceedings of the Twelfth International Conference on Foundations of Software Science and Computation Structures (FOSSACS 2009)*, volume 5504 of *Lecture Notes in Computer Science*, pages 288–302, York, UK, March 2009 2009. Springer.