Traffic Analysis Attacks in Anonymity Networks: Relationship Anonymity-Overhead Trade-off

Ognjen Vuković, György Dán, Gunnar Karlsson
ACCESS Linnaeus Center, School of Electrical Engineering
KTH Royal Institute of Technology, Stockholm, Sweden
Email: {vukovic,gyuri,gk}@ee.kth.se

***Abstract*—Mix networks and anonymity networks provide anonymous communication via relaying, which introduces overhead and increases the end-to-end message delivery delay. In practice overhead and delay must often be low, hence it is important to understand how to optimize anonymity for limited overhead and delay. In this work we address this question under passive traffic analysis attacks, whose goal is to learn the traffic matrix. For our study, we use two anonymity networks: MCrowds, an extension of Crowds, which provides unbounded communication delay and Minstrels, which provides bounded communication delay. We derive exact and approximate analytical expressions for the relationship anonymity for these systems. Using MCrowds and Minstrels we show that, contrary to intuition, increased overhead does not always improve anonymity. We investigate the impact of the system's parameters on anonymity, and the sensitivity of anonymity to the misestimation of the number of attackers.**

***Index Terms*—Relationship anonymity, communication overhead, traffic analysis, Bayesian analysis.**

I. Introduction

Many communication systems, for example modern industrial networks [1], [2], require high availability between a fixed set of nodes on a pairwise basis. The nodes can be the subsidiaries of an enterprise connected by a virtual private network over the public Internet, or they can be sensors, actuators and operation centers in a wide area industrial control system, e.g., in a supervisory control and data acquisition (SCADA) network. Cryptography may provide authentication, confidentiality and data integrity for the communication, but source and destination addresses would still be visible to an outside attacker who is able to observe one or more network links. The outside attacker may identify traffic patterns: who is communicating with whom, when and how often. Using this information the attacker can infer the importance of messages, and may perform targeted attacks on the communication between any two nodes. These targeted attacks might be hard to detect and can lead to incorrect system operation.

Mix networks [3] are a way to mitigate outside attacks by providing relationship anonymity, i.e., by making it untraceable who communicates with whom [4]. Nodes in a mix network relay and delay messages such that an outside attacker cannot trace the route of the individual messages through the mix. While relaying renders outside attacks more difficult, it introduces the possibility of inside attacks. Due to the often long life-cycles of industrial systems, software corruption is a threat and the complexity of the code-base makes it hard to detect. Corrupted nodes that are part of the mix network can perform inside attacks to determine the sender-receiver pair for messages that are relayed through them. Anonymity networks can provide some level of relationship anonymity against inside attackers (e.g., [5], [6]) by hiding the sender or the receiver from the relay nodes. Good sender (or receiver) anonymity in itself does not necessarily lead to good relationship anonymity [8], hence we focus on relationship anonymity in this paper.

The relationship anonymity provided by mix networks and anonymity networks comes at the price of delay and communication overhead. Excessive delays can negatively impact the system performance, while overhead leads to high resource requirements, so that in practice both have to be kept low. At the same time, the relationship anonymity may be a function of the number of nodes in the system and the number of nodes controlled by the attacker. Since the number of attacker nodes is unknown, finding the optimal level of overhead can be a challenging problem in practice.

In this paper we investigate the inherent trade-off between the communication overhead introduced and the level of relationship anonymity provided by anonymity networks. While intuition says that increased overhead should result in better anonymity, our results show that this is not necessarily the case. The results also show that larger anonymity networks provide better relationship anonymity for the same ratio of attacker nodes. Moreover, we show that it is in general better to overestimate the number of attacker nodes when choosing the level of overhead.

We consider an attacker whose goal is to perform a traffic analysis attack in order to determine the communication patterns between a set of communicating nodes, i.e., to learn the traffic matrix. We consider two methods for traffic analysis: the *Bayesian inference method* and the *Maximum posteriori method*. According to the *Bayesian inference method* the attacker considers all pairs of nodes as possible sender-receiver pairs for an intercepted message. According to the *Maximum posteriori method* the attacker only considers the most likely pairs of nodes as possible sender-receiver pairs for an intercepted message.

For our study we use two anonymity networks that provide relationship anonymity. First, MCrowds, a modification of Crowds [6], which provides anonymity by introducing unbounded message delivery delay. MCrowds provides sender anonymity using the same mechanism as Crowds, which

was shown to provide optimal sender anonymity for given average path length [7]. Unlike Crowds, MCrowds hides the receiver among a small subset of anonymity network nodes. This modification possibly leads to lower sender anonymity than in Crowds [7], but this way MCrowds allows us to explore the best combination of sender and receiver anonymity that provides optimal relationship anonymity. Second, Minstrels, which provides relationship anonymity by introducing bounded message delivery delay. Bounding the path length is achieved by limiting the number of visited nodes for each message.

Early works on traffic analysis attacks against anonymity networks by an external global attacker considered long term intersection attacks [8], [9], [10]. These attacks exploit the distribution of message destinations to decrease the relationship anonymity by relying on cases when the sender's anonymity is not *beyond suspicion*, i.e., the sender is distinguishable from other nodes. Disclosure attacks considered in [11] formulate traffic analysis as an optimization problem, under more general assumptions. More recent works have formulated traffic analysis attacks by an external global adversary in the context of Bayesian inference [8], [12], [13]. These attacks consider that the receiver is outside the anonymity network. In our system the sender and the receiver are part of the anonymity network, and message destinations can have an arbitrary distribution. We use Bayesian inference, but we consider an internal adversary instead of an external global observer. The relationship between anonymity and traffic overhead was investigated in [15] for a global adversary. The authors considered an anonymity network in which routes have a fixed length, and padding (i.e., dummy traffic) is sent over links to hide traffic patterns. In our work the overhead is measured in terms of route length and the adversary cannot observe the global traffic, only traffic traversing compromised nodes. Sender anonymity in the presence of compromised nodes was considered for Crowds [7] and for systems inspired by Crowds [15]. In our work, we consider relationship anonymity instead of sender anonymity, and address the trade-off between anonymity and overhead.

The rest of the paper is organized as follows. Section II describes our system model, the anonymity metric, and the traffic analysis methods. Section III describes of the MCrowds and Minstrels anonymity networks. In Section IV we develop analytical models of the relationship anonymity provided by MCrowds and Minstrels, and we show numerical results based on the models in Section V. Section VI concludes the paper.

II. System model and metrics

We consider an anonymity network that consists of a set $\mathcal{N}$ of nodes, $N = ||\mathcal{N}||$. The nodes act as *sources*, *destinations* and as *relay* nodes for each others' messages. The underlying communication network is a complete graph. We consider that encryption and authentication are done end-to-end between the sender and the receiver, but the relay nodes do not perform cryptographic operations on the messages in order to limit their computational burden.

The *inside attacker* is in control of a set $\mathscr{C} \subset \mathcal{N}$ ($C = ||\mathscr{C}||$) of compromised nodes. The attacker can observe the messages traversing the nodes in $\mathscr{C}$ and the protocol specific information contained in the messages. It can make use of the payload of the messages to recognize if the same message visits several compromised nodes. The attacker has an *a-priori* belief of the system traffic matrix in the form of the distribution $P(S(a),R(b))$ for every pair of nodes $(a,b) : a \in \mathcal{N}, b \in \mathcal{N} \setminus \{a\}$ (nodes do not send messages to themselves over the anonymity network.). For every message that the attacker observes, it calculates the *a-posteriori* probability $P(\hat{S}(a),\hat{R}(b))$ for every pair of *trusted* (not compromised) nodes that it is the sender-receiver pair of the message. The attacker maintains a real-valued counter for every pair of nodes (*a*,*b*), and it increases the counters with values that are calculated from the a-posteriori probabilities for every observed message. The attacker uses the counters to estimate the number of exchanged messages between every pair of nodes in a given time interval, that is, to learn the actual traffic matrix. The attacker starts the estimation by initializing the counters to zero.

We consider two metrics: the *overhead* of the anonymity network and the *relationship anonymity*. We define the *overhead* as the average number of nodes $E[K]$ that an arbitrary message visits. We quantify the *relationship anonymity* by the average increase of the counter corresponding to the real sender-receiver pair (s,r) for every message sent by (s,r). Consequently, for the messages that are not observed by the attacker, the counter is not incremented. The lower the relationship anonymity is, the more difficult it is for the attacker to learn the intensity of messages sent from node s to node r. Note that the relationship anonymity may not be the same for (s,r) and for (r,s). In general, the relationship anonymity depends on two factors. First, on the probability of having an attacker node on the path. Second, on the a-posteriori probability assigned to the sender-receiver pair $P(\hat{S}(s),\hat{R}(r))$ by an attacker node on the path. Both factors are functions of the anonymity protocol, the number of nodes N and the number of inside attacker nodes C. Furthermore, the probability $P(\hat{S}(s),\hat{R}(r))$ is used to determine the value by which the corresponding counter is incremented. This value depends on the method that the attacker uses for counting. We consider two counting methods.

A. Bayesian inference method

Using the Bayesian Inference (BI) method, when the attacker intercepts a message, it increments the corresponding counters with the a-posteriori probabilities. Let us denote by $P(H_{1+}|S(s),R(r))$ the probability that an attacker node occurs on the path given that (s,r) is the sender-receiver pair, and by $P(\hat{S}(s),\hat{R}(r)|H_{1+},S(s),R(r))$ the a-posteriori probability that the attacker identifies (s,r) as the sender-receiver pair given its occurrence on the path. Then we can express the relationship anonymity under the BI method as

$$P_{relB}(s,r) = P(\hat{S}(s),\hat{R}(r)|H_{1+},S(s),R(r)) \cdot P(H_{1+}|S(s),R(r)), \quad (1)$$

B. Maximum posteriori method

Using the Maximum Posteriori (MP) method, when the attacker intercepts a message, it populates the set $\mathscr{Q}$ of most

likely sender-receiver pairs with the pairs of nodes that have the highest a-posteriori probability. In the worst case the set $\mathscr{Q}$ is a singleton, $||\mathscr{Q}|| = 1$, and the anonymity is likely to be low. At the other extreme, $\mathscr{Q}$ can contain all possible send-receiver pairs, $||\mathscr{Q}|| = (N-C)\cdot(N-C-1)$, which corresponds to perfect relationship anonymity. In general $(a,b) \in \mathscr{Q}$ does not imply that (a,b) is the actual sender-receiver pair, not even when $||\mathscr{Q}|| = 1$. Nevertheless, intuitively, we can say that $(s,r) \in \mathscr{Q}$ is more likely than $(s,r) \notin \mathscr{Q}$.

Let us denote by $P((s,r) \in \mathscr{Q}|H_{1+}, S(s), R(r))$ the probability that the sender-receiver pair is one of the most likely sender-receiver pairs, i.e., $(s,r) \in \mathscr{Q}$. If $(s,r) \in \mathscr{Q}$, then the attacker identifies (s,r) as the sender-receiver pair given its occurrence on the path and $(s,r) \in \mathscr{Q}$ with probability $1/||\mathscr{Q}||$. Note that if $(s,r) \notin \mathscr{Q}$, the attacker assigns probability 0 to the event that (s,r) is the sender-receiver pair. Using this notation we can express the relationship anonymity under the MP method as

$$P_{relM}(s,r) = \frac{P((s,r) \in \mathscr{Q}|H_{1+}, S(s), R(r))}{||\mathscr{Q}||} \cdot P(H_{1+}|S(s), R(r)). \quad (2)$$

III. Anonymity system descriptions

In the following we describe the considered anonymity networks: MCrowds and Minstrels.

A. MCrowds system description

MCrowds is an anonymity network inspired by Crowds [6], which was proven to provide optimal sender anonymity [7]. In MCrowds the sender specifies a set $\mathscr{M}$ of nodes as receiver for a message. The number $M = ||\mathscr{M}||$ of receiver nodes is a system parameter. Nodes specified in the set $\mathscr{M}$ are not used for relaying. For a message to reach its intended receiver r it must be that $r \in \mathscr{M}$; the other $M-1$ nodes are chosen uniformly at random. The sender then relays the message to one of the $\mathscr{N} \setminus \mathscr{M}$ nodes (including itself) selected uniformly at random. A relay node relays the message with probability p_f to one of the $\mathscr{N} \setminus \mathscr{M}$ nodes chosen uniformly at random. Note that a node can relay the message to itself, in which case the message does not leave the node. Otherwise, the message is sent as a multicast message to all receiver nodes specified in $\mathscr{M}$ (i.e., with probability $1-p_f$). Upon multicasting, the receiver set is removed from the message. Node r recognizes that it is the receiver while the other $\mathscr{M} \setminus \{r\}$ nodes discard the message. For $M = 1$ MCrowds is equivalent to Crowds, except that the receiver node is part of the anonymity network, $r \in \mathscr{N}$. In principle the nodes could use different values of M and p_f, but to ease the analysis we consider that all nodes use the same parameter values.

B. Minstrels system description

Minstrels uses nodes as message relays in the same way as Crowds with the difference that the number of nodes visited by a message is bounded.

When a node s wants to send a message to a node r it picks a node uniformly at random among the other $N-1$ nodes (excluding s) and forwards the message. The next node forwards the message to one of the other $N-2$ nodes (excluding itself and the sender node s) chosen uniformly at random. Every subsequent forwarder picks one of the non-visited nodes to forward the message. When node r receives the message, it will send the message further in order to improve the receiver anonymity. The path ends when all N nodes have been visited.

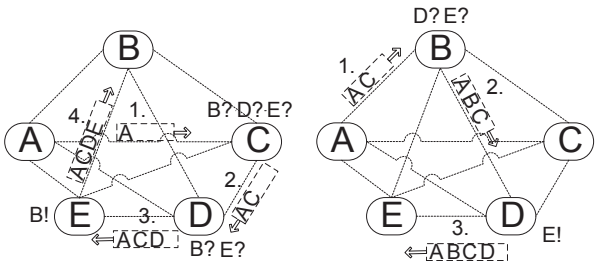

Fig. 1. A simple example of Minstrels with five nodes.

The message, or part of it, is encrypted with the receiver's public key. When a node receives the message, it checks whether it is the receiver by trying to decrypt the encrypted part of the message. If the decrypted part of the message represents valid data, the node is the receiver. Note that a node does not know who the receiver is, but it can check whether it is the receiver itself.

To bound the path length, every message records the set $\mathscr{V}$ of the visited nodes in its header. The set can be implemented, for example, using a Bloom filter, to keep its size small. When a relaying node receives a message, it adds itself to the set $\mathscr{V}$ and relays the message to one of the remaining non-visited nodes. To control the maximum path length (i.e., delay) the sender can initialize the set $\mathscr{V}$ of visited nodes with a number $f \in \{0,...,N-1\}$ of the nodes in the system. These initialized nodes are considered as visited so that the message can not be relayed to them. A message traverses all nodes except for the initialized nodes in the set $\mathscr{V}$ and hence the sender must not include the receiver in the set $\mathscr{V}$. The sender picks the number of initialized nodes at random: it initializes the set with f nodes with probability $P(F=f)$, where $\sum_{f=0}^{N-1} P(F=f) = 1$. For $f=0$ the set is empty, for $f=1$ the set is initialized only with the sender and for $f>1$ the set is initialized with the sender and $f-1$ other nodes. Note that for $f>0$, the sender always includes itself in the set. The distribution of F is a system parameter, and we use it to explore the anonymity-overhead trade-off. In principle the nodes could use different distributions for F, but again, to ease the analysis we consider that all nodes use the same distribution.

Fig. 1 shows two simple examples with five nodes, node A as sender and node D as receiver. Fig. 1 (left) shows a case when the set $\mathscr{V}$ is initialized with the sender node A and the message is forwarded to node C. Node C checks if it is the receiver, puts itself in the set and chooses the next hop uniformly at random among nodes (B,D,E). The next hop, node D, follows the same procedure with only two forwarding options (B,E). Fig. 1 (right) shows another case when the set $\mathscr{V}$ is initialized with the sender and node C, and the message

is forwarded to node B. Node B adds itself to the set and decides to which of the remaining nodes (D,E) to forward the message. Node C is considered as already visited.

IV. Overhead and Anonymity

In the following we derive expressions for the communication overhead and the relationship anonymity provided against inside attackers for MCrowds and for Minstrels.

A. Communication Overhead

We start with calculating the communication overhead of MCrowds and Minstrels. For MCrowds, the mean number of nodes visited by a message is the expected value of a geometric distribution with success probability $1-p_f$ plus the multicast messages, i.e.,

$$E[K] = \frac{p_f}{1-p_f} + 1 + M. \tag{3}$$

For Minstrels and for a given number f of initialized nodes in the set $\mathscr{V}$, the number of nodes visited by a message is equal to $K = N - f$. The mean number of visited nodes depends on the distribution of F and it can be expressed as

$$E[K] = \sum_{f=0}^{N-1} P(F=f)\cdot(N-f). \tag{4}$$

B. Relationship Anonymity Against Inside Attackers

In the following we derive the relationship anonymity expressions for MCrowds and Minstrels.

1) MCrowds: We start the calculation of the relationship anonymity with expressing the probability of having an attacker node on the path. This probability depends on the number of receiver nodes M, and on the number of attacker nodes in the set $\mathscr{M}$ of receiver nodes. We denote by c_M the number of attacker nodes in the receiver set. c_M is a realization of the random variable $C_M \in \{max(0, M-(N-C-1)), ..., min(M-1, C)\}$. For $M=1$ there cannot be attacker nodes in the receiver set, only the receiver r, and therefore $P(C_M = 0) = 1$. For $M > 1$, the sender selects the other $M-1$ nodes uniformly at random from $N-2$ nodes (excluding the sender and the receiver). Thus, once k trusted and j attacker nodes have been selected, the next selected node is a trusted node with probability $\frac{N-C-2-k}{N-2-k-j}$, and is an attacker node with probability $\frac{C-j}{N-2-k-j}$. Observe that it does not matter in what order the c_M attacker nodes were selected, and thus the probability that there are c_M attacker nodes in the set of receiver nodes is

$$P(C_M = c_M) = \binom{M-1}{c_M} \frac{\prod_{k=2}^{M-c_M}(N-C-k)\prod_{k=0}^{c_M-1}(C-k)}{\prod_{k=2}^{M}(N-k)}. \tag{5}$$

Let us denote by H_i the event that the position of the first attacker node is i. The event H_i happens if the message is first relayed $i-1$ times through trusted nodes, i.e., not through attacker nodes in the set $\mathscr{N}\setminus\mathscr{M}$, but the i^{th} relay is an attacker node. A trusted node at position $i-1$ relays the message to one of the $C-c_M$ attacker nodes with probability $\frac{C-c_M}{N-M}$. Therefore, conditioned on $C_M = c_M$ we have

$$P(H_i|c_M, S(a), R(b)) = \frac{C-c_M}{N-M} p_f^{(i-1)} \left(1 - \frac{C-c_M}{N-M}\right)^{(i-1)}, \tag{6}$$

for $a \in \mathscr{N}\setminus(\mathscr{C}\cup\mathscr{M})$ and $b \in \mathscr{M}\setminus\mathscr{C}$. Note that for brevity we use c_M to denote the condition $C_M = c_M$ in (6) and henceforth. If the message is again relayed over an attacker node on any position after i, the attacker does not gain any additional information about the sender-receiver pair (s,r) of the message: any node from the set $\mathscr{N}\setminus\mathscr{M}$ is equally likely to be used as relay, and the receiver is still one of the nodes in $\mathscr{M}$. Hence, the probability assigned to the sender-receiver pair does not change. Thus, it is enough to focus on the position of the first attacker node on the path. Let us now denote by H_{1+} the event that there is an attacker on the path as a relay. This event happens if the event H_i happens for any $i > 0$, and the H_i are mutually exclusive. Therefore, conditioned on $C_M = c_M$, the event H_{1+} happens with probability

$$\begin{aligned} P(H_{1+}|c_M, S(a), R(b)) &= \sum_{i=1}^{\infty} P(H_i|c_M, S(a), R(b)) \\ &= \frac{C-c_M}{N-M-p_f(N-C-M+c_M)}. \end{aligned} \tag{7}$$

When the first attacker node on the path gets the message, the attacker knows the nodes in the set $\mathscr{M}$, the number of attacker nodes c_M in the set, and the node that the message is received from, i.e., the predecessor p. Let us denote by I_a the event that the predecessor is node a ($p = a$). If the attacker node is on position $i = 1$, then the sender of the message is the predecessor, i.e., the event I_a happens if a is the sender.. Otherwise, for $i > 1$ and $S(a)$ we have $P(I_a|H_{2+}, c_M, S(a), R(b)) = \frac{1}{N-C-M+c_M}$ for any $b \in \mathscr{N}\setminus\mathscr{C}$ and $b \neq a$, because any trusted node from the set $\mathscr{N}\setminus\mathscr{M}$ is equally likely to be the predecessor. Observe that I_a can only happen if $a \notin \mathscr{M}$. The event $a \notin \mathscr{M}$ happens with the probability $P(a \notin \mathscr{M}|c_M, S(s), R(b)) = \frac{N-C-M-c_M-1}{N-C-2}$, for any $b \in \mathscr{N}\setminus\mathscr{C}$ and $b \notin \{s, a\}$. Thus, $P(I_a|H_{2+}, c_M, S(s), R(b)) = \frac{P(a\notin\mathscr{M}|c_M, S(s), R(b))}{N-C-M+c_M}$. Finally, the event I_a conditioned on H_{1+}, and $S(a)$ or $S(s)$ such that $s \neq a$ happens with probability

$$\begin{aligned} &P(I_a|H_{1+}, c_M, S(a), R(b)) = P(H_1|c_M, S(a), R(b)) + \\ &P(I_a|H_{2+}, c_M, S(a), R(b)) \cdot P(H_{2+}|c_M, S(a), R(b)), \end{aligned} \tag{8}$$

$$\begin{aligned} &P(I_a|H_{1+}, c_M, S(s), R(b)) = \\ &P(I_a|H_{2+}, c_M, S(s), R(b)) \cdot P(H_{2+}|c_M, S(s), R(b)), \end{aligned} \tag{9}$$

where $P(H_{2+}|c_M, S(a), R(b)) = \sum_{i=2}^{\infty} P(H_i|c_M, S(a), R(b))$. Similarly, we denote by $\bar{I}_a$ the event that the predecessor is not node a, i.e., $p \neq a$.

Let us now consider the case when node s sends a message and the attacker appears as a relay, i.e., the events $S(s)$ and H_{1+} happen. Then, if node s is the predecessor (I_s), the attacker identifies node s being the sender of the message with

probability

$$P(\hat{S}(s)|I_s,H_{1+},c_M,S(s),R(b)) = \frac{\sum_b P(I_s,H_{1+},c_M|S(s),R(b))\cdot P(S(s),R(b))}{\sum_{(a,b)} P(I_s,H_{1+},c_M|S(a),R(b))\cdot P(S(a),R(b))}, \quad (10)$$

where $a \in \mathcal{N} \setminus (\mathcal{M} \cup \mathcal{C})$ and $b \in \mathcal{M} \setminus \mathcal{C}$. $P(S(a),R(b))$ is the a-priori probability that node a sends a message to node b, i.e., the attacker's a-priori belief of the traffic matrix. The probability $P(\hat{S}(s)|\bar{I}_s,H_{1+},c_M,S(s),R(b))$ that the attacker assigns to node s when it is not the predecessor ($\bar{I}_s$) can be expressed in a similar way.

Then, for the BI method the probability that a *relaying* attacker assigns to the actual sender of the message, given H_{1+} and $C_M = c_M$, is

$$P(\hat{S}(s)|H_{1+},c_M,S(s),R(b)) = P(\hat{S}(s)|I_s,H_{1+},c_M,S(s),R(b))\cdot P(I_s|H_{1+},c_M,S(s),R(b)) + P(\hat{S}(s)|\bar{I}_s,H_{1+},c_M,S(s),R(b))\cdot P(\bar{I}_s|H_{1+},c_M,S(s),R(b)), \quad (11)$$

The probability assigned to the receiver is $P(\hat{R}(r)|H_{1+},c_M,S(s),R(r)) = \frac{1}{M-c_M}$. Note that the events are conditionally independent since the receiver is one of the trusted nodes in $\mathcal{M}$, and the sender is one of the trusted nodes in $\mathcal{N} \setminus \mathcal{M}$. Hence, the probability assigned to the sender-receiver pair (s,r) is the product of the two.

What remains is to calculate the probability for a non-relaying attacker. Let us denote by $\overline{H}_{1+}$ the event that a message does not visit any attacker node as a relay, the complement event of H_{1+}. If $\overline{H}_{1+}$ and $C_M = 0$ happens then the attacker does not observe the message. Otherwise, if $\overline{H}_{1+}$ happens but $C_M > 0$ then the attacker nodes in the receiver set $\mathcal{M}$ get the multicast message from the last relay node (the one that decides to send the message to the receivers, with probability $1-p_f$). Observe that any trusted node from the set $\mathcal{N} \setminus \mathcal{M}$ is equally likely to be the last relay (the predecessor), and therefore it holds that $P(I_a|\overline{H}_{1+},c_M,S(a),R(b)) = P(I_a|H_{2+},c_M,S(a),R(b))$, for every $C_M > 0$. Correspondingly, it also holds that $P(I_a|\overline{H}_{1+},c_M,S(s),R(b)) = P(I_a|H_{2+},c_M,S(s),R(b))$ for $s \neq a$ and $C_M > 0$. Consequently, given $\overline{H}_{1+}$, $C_M > 0$, and I_s or $\bar{I}_s$, the probability that the attacker assigns to node s being the sender can be expressed following the same reasoning as in (10). Finally, the probability $P(\hat{S}(s)|\overline{H}_{1+},c_M,S(s),R(b))$ that the attacker assigns to the actual sender, given $\overline{H}_{1+}$ and $C_M > 0$, can be expressed using the law of total probability conditioned on I_s and $\bar{I}_s$, similar to (11).

Since the last relay node removes the receiver set $\mathcal{M}$ from the message, the receiver is hidden among $N-C-1$ trusted nodes (it cannot be the last relay). However, the probability assigned to the receiver depends on whom the attacker guesses to be the sender. If the attacker believes that the predecessor is the sender, each of the other $N-C-1$ trusted nodes is equally likely to be the receiver. If the attacker believes that the predecessor is not the sender then each of the $N-C-2$ trusted nodes apart from the predecessor and the sender is equally likely to be the receiver. Therefore, if I_s happens and the attacker identifies the sender with probability $P(\hat{S}(s)|I_s,H_{1+},c_M,S(s),R(b))$, then the probability assigned to the receiver equals to $P(\hat{R}(r)|\hat{S}(s),I_s,\overline{H}_{1+},c_M,S(s),R(r)) = \frac{1}{N-C-1}$. Otherwise, if $\bar{I}_s$ happens and the attacker identifies the sender with probability $P(\hat{S}(s)|\bar{I}_s,H_{1+},c_M,S(s),R(b))$, then $P(\hat{R}(r)|\hat{S}(s),\bar{I}_s,\overline{H}_{1+},c_M,S(s),R(r)) = \frac{1}{N-C-2}$. Thus, given $\overline{H}_{1+}$ and $C_M = c_M > 0$, the probability assigned to the sender-receiver pair (s,r) can be expressed as

$$P(\hat{S}(s),\hat{R}(r)|\overline{H}_{1+},c_M,S(s),R(r)) = \frac{P(\hat{S}(s)|I_s,\overline{H}_{1+},c_M,S(s),R(r))}{N-C-1}\cdot P(I_s|\overline{H}_{1+},c_M,S(s),R(r)) + \frac{P(\hat{S}(s)|\bar{I}_s,\overline{H}_{1+},c_M,S(s),R(r))}{N-C-2}\cdot P(\bar{I}_s|\overline{H}_{1+},c_M,S(s),R(r)). \quad (12)$$

It can happen that there is an attacker node on the path as a relay (H_{1+}) and there is at least one attacker node specified in the receiver set ($C_M > 0$). In this case it is clear that the attacker assigns higher probability to the actual sender-receiver pair (s,r) when the message is observed by the relaying attacker node than when it is observed by the attacker node in the receiver set. Since the attacker can recognize if the same message visits several compromised nodes, the attacker does not recalculate the assigned probability when it gets the multicast message. Therefore, we do not provide separate expressions for this case.

Finally, we express $P_{relB}(s,r)$ using the law of total probability, accounting for all possible values of C_M, and for all cases when the attacker receives the message, i.e., either H_{1+} or $\overline{H}_{1+}$ and $C_M = c_M > 0$,

$$P_{relB}(s,r) = \sum_{c_M} P(\hat{S}(s),\hat{R}(r)|H_{1+},c_M,S(s),R(r))\cdot P(H_{1+}|c_M,S(s),R(r))\cdot P(C_M = c_M) + \sum_{c_M \neq 0} P(\hat{S}(s),\hat{R}(r)|\overline{H}_{1+},c_M,S(s),R(r))\cdot P(\overline{H}_{1+}|c_M,S(s),R(r))\cdot P(C_M = c_M). \quad (13)$$

In order to calculate the relationship anonymity $P_{relM}(s,r)$ under the MP method, we need to determine the probability that the sender-receiver pair (s,r) is one of the most likely sender-receiver pairs, i.e., $(s,r) \in \mathcal{Q}$. Given particular events, e.g., I_s and H_{1+}, the sender-receiver pair (s,r) is in the set $\mathcal{Q}$ if the probability $P(\hat{S}(s),\hat{R}(r)|I_s,H_{1+},c_M,S(a),R(b))$ for every trusted sender-receiver pair $\forall(a,b)$, s.t. $a \neq b$, is less than or equal to $P(\hat{S}(s),\hat{R}(r)|I_s,H_{1+},c_M,S(s),R(r))$. In the special case when the attacker's a-priori belief is that the traffic matrix is homogeneous, all pairs $(a,b), a \neq b$ of trusted nodes are equally likely to be the sender-receiver pair. Hence, if either H_{1+} or $\overline{H}_{1+}$ and $C_M = c_M > 0$ happens, then the predecessor is the most likely sender. Therefore, the sender-receiver pair (s,r) is in $\mathcal{Q}$ only if I_s happens. If this happens, then the cardinality of the set $\mathcal{Q}$ is equal to $M - c_M$, the number of trusted nodes in the receiver set $\mathcal{M}$.

2) Minstrels: When the first attacker node on the path gets the message, the attacker knows the number c_F of attacker nodes that the set of visited nodes was initialized with by the sender. c_F is a realization of the random variable C_F, whose

distribution depends on the number f of initialized nodes in the set of visited nodes, $\mathscr{V}$.

In Minstrels the probability that the attacker assigns to a sender-receiver pair does not only depend on the node that the message is received from, i.e., the predecessor p, but also on the contents of the set $\mathscr{V}$ of visited nodes that the message carries. Consequently, the attacker distinguishes between three disjoint sets of nodes: the predecessor node ($\{p\}$), nodes in the set of visited nodes except the predecessor ($\mathscr{V} \setminus \{p\}$), and nodes not in the set of visited nodes ($\overline{\mathscr{V} \cup \{p\}}$). These sets form a partition of the set of all nodes in the system, and trusted nodes belonging to the same set are equally likely to be the sender (and the receiver). As a shorthand for the universe of distinguishable events we use the notation $\Omega_s = \{s = p, s \in \mathscr{V} \setminus \{p\}, s \in \overline{\mathscr{V} \cup \{p\}}\}$, where, for example, $s = p$ is the event that the predecessor is the sender. Similarly, we define $\Omega_r = \{r = p, r \in \mathscr{V} \setminus \{p\}, r \in \overline{\mathscr{V} \cup \{p\}}\}$ for the distinguishable events regarding the receiver.

If the message visits multiple attacker nodes on its path then the attacker can identify the nodes that were visited between the different attacker nodes. However, since any node that has not been visited yet is equally likely to be visited by the message, the attacker does not gain more information that could increase the probability assigned to the sender-receiver pair (s,r). Hence, it is enough to consider the first attacker node on the path that gets the message. Given the information on $\mathscr{V}$, c_F, and p available to the attacker, we can use the law of total probability to expand (1) and (2) conditional on the size $||\mathscr{V}|| = v$ of the set of visited nodes, $\omega_s \in \Omega_s$, $\omega_r \in \Omega_r$, and $C_F = c_F$,

$$P_{relB}(s,r) = \sum_{c_F}\sum_{v}\sum_{\omega_s}\sum_{\omega_r} P(\hat{S}(s),\hat{R}(r)|\omega_r,\omega_s,c_F,H_{1+},v,S(s),R(r)) \quad (14)$$

$$\cdot P(\omega_r,\omega_s,c_F,H_{1+},v|S(s),R(r)), \quad (15)$$

$$P_{relM}(s,r) = \sum_{c_F}\sum_{v}\sum_{\omega_s}\sum_{\omega_r} \frac{P((s,r)\in\mathscr{Q}|\omega_r,\omega_s,c_F,H_{1+},v,S(s),R(r))}{||\mathscr{Q}||} \quad (16)$$

$$\cdot P(\omega_r,\omega_s,c_F,H_{1+},v|S(s),R(r)). \quad (17)$$

Note that (15) and (17) are the probability that a message with (s,r) as sender-receiver pair is received by an attacker node and carries particular information. The numerator in eq. (16) is the probability that the sender-receiver pair (s,r) is in the set $\mathscr{Q}$.

Before we turn to the calculation of the probability $P(\omega_r,\omega_s,v,c_F,H_{1+}|S(s),R(r))$ we introduce the notation $H(v,c_F|F=f)$ for the joint event $||\mathscr{V}|| = v$, H_{1+}, and $C_F = c_F$ for a given number of initialized nodes f. Clearly, $v \geq f$. The probability of this event can be expressed as

$$P(H(v,c_F|F=f)) = \begin{cases} \frac{C}{N-1} & v=0, f=0 \\ P(C_F=0|F=f)\frac{N-C-1}{N-1}\frac{C}{N-v}\prod_{z=1}^{v-1}\frac{N-C-z}{N-z} & v\geq 1, f=0 \\ P(C_F=c_F|F=f)\frac{C-c_F}{N-v}\prod_{z=f}^{v-1}\frac{N-C+c_F-z}{N-z} & v\geq 1, f>0, \end{cases} \quad (18)$$

where $P(C_F|F=f)$ is the probability that the set of visited nodes is initialized with c_F attacker nodes, given that it is initialized with f nodes by the sender. Due to the rules of initialization, $c_F \in \{max(0, f-1-(N-2-C)), min(f-1, C)\}$. For $F=0$ and $F=1$ there cannot be any initialized attackers, hence $P(C_F=0|F\in\{0,1\})=1$ and $P(C_F>0|F\in\{0,1\})=0$. For $f>1$ we have

$$P(C_F|F=f) = \binom{f-1}{c_F} \frac{\prod_{k=2}^{f-c_F}(N-C-k)\prod_{k=0}^{c_F-1}(C-k)}{\prod_{k=2}^{f}(N-k)}. \quad (19)$$

We now turn to the calculation of the probability $P(\omega_r,\omega_s,v,c_F,H_{1+}|S(s),R(r))$, i.e., the probability that the attacker would receive a particular message sent by s to r. If the sender is the predecessor ($s=p$) the receiver cannot be the predecessor, hence $P(r=p,s=p,v,c_F,H_{1+}|S(s),R(r))=0$. For the rest of the cases we show the probabilities in a tabular form to improve readability.

For $||\mathscr{V}||=0$ and $||\mathscr{V}||=1$ there can be no attackers in the set of visited nodes (when received by the first attacker), because if the sender initializes the set of visited nodes with $f>0$ nodes, it has to include itself in the set. Hence, for $||\mathscr{V}||=0$ and $||\mathscr{V}||=1$ we have $C_F>0$ with probability 0. Furthermore, for $||\mathscr{V}||=0$ the sender must be the predecessor ($s=p$) and the receiver cannot be in the set of visited nodes ($r\in\overline{\mathscr{V}\cup\{p\}}$). Every other tuple in $\{(\omega_s,\omega_r):\omega_s\in\Omega_s,\omega_r\in\Omega_r\}$ has probability 0. The first row of Table I shows the corresponding probability, i.e., the probability that the sender initializes the message with an empty set, and chooses the attacker as next hop. For $||\mathscr{V}||=1$ the sender and the receiver cannot both be in the set of visited nodes. Furthermore, if the sender or the receiver is in the set of visited nodes, it must be the predecessor, hence $s\in\mathscr{V}\setminus\{p\}$ and $r\in\mathscr{V}\setminus\{p\}$ have probability 0. The probabilities for the remaining cases for $||\mathscr{V}||=1$ are shown in Table I. As an example, the third row in the table is the probability that the sender initializes the set empty, forwards the message to the receiver, which then forwards the message to the attacker.

TABLE I
$P(\Omega_r,\Omega_s,||\mathscr{V}||\in\{0,1\},C_F=0,H_{1+}|S(s),R(r))$

Ω_s,Ω_r	$\|\|\mathscr{V}\|\|$	
$s=p,\ r\in\overline{\mathscr{V}\cup\{p\}}$	0	$P(F=0)P(H(0,0\|F=0))$
$s=p,\ r\in\overline{\mathscr{V}\cup\{p\}}$	1	$P(F=1)P(H(1,0\|F=1))$
$s\in\overline{\mathscr{V}\cup\{p\}},\ r=p$	1	$P(F=0)P(H(1,0\|F=0))\frac{1}{N-C-1}$
$s\in\overline{\mathscr{V}\cup\{p\}},\ r\in\overline{\mathscr{V}\cup\{p\}}$	1	$P(F=0)P(H(1,0\|F=0))\frac{N-C-2}{N-C-1}$

For $||\mathscr{V}||>1$ there may or may not be attackers in the set of initialized nodes. When there are attackers in the set of initialized nodes ($C_F>0$), the sender has to be in the set of visited nodes. Furthermore, if the sender is the predecessor ($s=p$) then the receiver cannot be in the set of visited nodes ($r\in\mathscr{V}\setminus\{p\}$), because this could only happen if the sender had initialized the set of visited nodes with the receiver, but then the receiver would never receive the message. The corresponding probabilities for $||\mathscr{V}||>1$ are shown in Table II and Table III in the Appendix.

Let us now turn to the calculation of the probabilities that the attacker correctly identifies the sender-receiver pair (s,r) used in the Bayesian inference method (14). Given a message

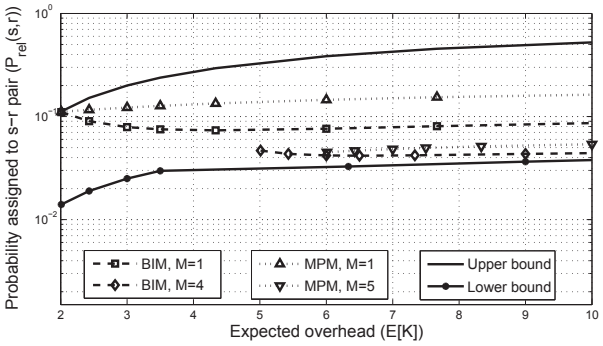

Fig. 2. Relationship anonymity vs. overhead for MCrowds, $N = 10$, $C = 1$

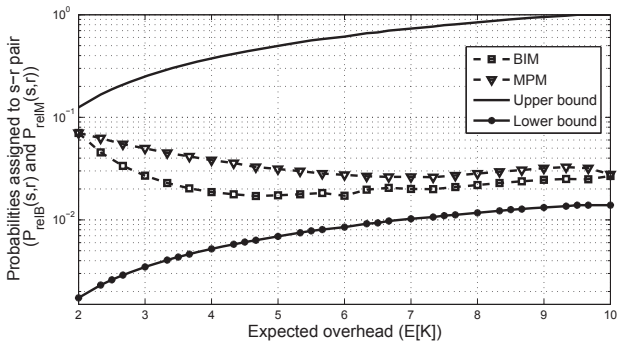

Fig. 3. Relationship anonymity vs. overhead for Minstrels, $N = 10$, $C = 1$

received by an attacker node that contains information ($||\mathscr{V}|| = v$, $\omega_s \in \Omega_s$, $\omega_r \in \Omega_r$, and $C_F = c_F$) the attacker would identify (s,r) as the sender-receiver pair with probability

$$P(\hat{S}(s),\hat{R}(r)|\omega_r,\omega_s,c_F,H_{1+},v) = \frac{P(\omega_r,\omega_s,v,c_F,H_{1+}|S(s),R(r))\cdot P(S(s),R(r))}{\sum_{(a,b)} P(\omega_r,\omega_s,v,c_F,H_{1+}|S(a),R(b))\cdot P(S(a),R(b))} \quad (20)$$

where the summation in the denominator is over all possible non-attacker sender-receiver pairs (a,b), such that $a \neq b$. $P(S(a),R(b))$ is the a-priori probability that node a sends a message to node b, i.e., the attacker's a-priori belief of the traffic matrix. In the special case when the attacker's a-priori belief is that the traffic matrix is homogeneous, $P(S(a),R(b)) = \frac{1}{(N-C)(N-C-1)}$ for all (a,b) such that $a \neq b$, and these probabilities cancel out each other in (20).

We already calculated the numerator of (20), so in order to finish our calculations we only have to express $P(\omega_r,\omega_s,v,c_F,H_{1+}|S(a),R(b))$ and only for the cases when the numerator of (20) is non-zero, and when $a \neq s$ or $b \neq r$.

The attacker can receive a message with an empty set of visited nodes ($||\mathscr{V}|| = 0, C_F = 0$) only if the sender is the predecessor, hence, $P(\omega_r,\omega_s,||\mathscr{V}|| = 0, C_F = 0, H_{1+}|S(a),R(b)) > 0$ only for $a = s$. Nevertheless, the receiver of the message can be any trusted node $b \neq s$ (we use $\forall b$ as a shorthand notation). The corresponding probability $P(\Omega_r,\Omega_s,||\mathscr{V}|| = 0, C_F = 0, H_{1+}|S(a),R(b))$ is given in Table IV in the Appendix.

The attacker can receive a message with only one node in the set of visited nodes ($||\mathscr{V}|| = 1$), in which case the node in the set is the predecessor. The set could have been sent by the predecessor ($a = p$) or by a node not in the set ($a \in \overline{\mathscr{V} \cup \{p\}}$), but in either case there cannot be any attacker node initialized in the set ($C_F = 0$). The receiver could be any other node ($\forall b$). The probability of receiving such a message $P(\Omega_r,\Omega_s,||\mathscr{V}|| = 1, C_F = 0, H_{1+}|S(a),R(b))$ is given in Table V in the Appendix.

The probabilities for $||\mathscr{V}|| > 1$ can be obtained following a similar reasoning. In order to maintain the readability of the paper we describe the probabilities in the Appendix.

We now turn to the calculation of the probability (16) that the sender-receiver pair (s,r) is one of the most likely sender-receiver pairs, i.e. $(s,r) \in \mathscr{Q}$, used in the Maximum posteriori method. The sender-receiver pair (s,r) is in the set $\mathscr{Q}$ if the probability $P(\omega_r,\omega_s,v,c_F,H_{1+}|S(a),R(b))$ for every sender-receiver pair $\forall(a,b)$ is less than or equal to $P(\omega_r,\omega_s,v,c_F,H_{1+}|S(s),R(r))$.

C. Bounds For Relationship Anonymity

In order to have a better understanding of the relationship anonymity provided by the described anonymity networks, we define upper and lower bounds for the relationship anonymity. To obtain the upper bound, we consider that whenever the attacker intercepts a message, it knows the sender-receiver pair with probability $P(\hat{S}(s),\hat{R}(r)|H_{1+},S(s),R(r)) = 1$. Hence, the bound is equivalent to the probability of having an attacker node on the path $P(H_{1+}|S(s),R(r))$. To obtain the lower bound, we consider that whenever the attacker intercepts a message, it assumes that any trusted pair of nodes is equally likely to be the sender-receiver pair with probability $P(\hat{S}(s),\hat{R}(r)|H_{1+},S(s),R(r)) = \frac{1}{(N-C)(N-C-1)}$.

V. Numerical Results

In the following we use the expressions described above for the BI method (denoted by BIM in the figures) and for the MP method (denoted by MPM in the figures) to get insight into the relationship anonymity-overhead trade-off provided by MCrowds and by Minstrels. To explore the trade-off, for MCrowds we use the relaying probability $p_f \in (0,1)$ and $M \in \{1,\ldots,N-2\}$, and for Minstrels we use various uniform, binomial, and triangular distributions to choose the number F of initialized nodes. The attacker's a-priory belief is that the traffic matrix is homogeneous.

Fig. 2 and Fig. 3 show the relationship anonymity under the BI method ($P_{relB}(s,r)$) and the relationship anonymity under the MP method ($P_{relM}(s,r)$) as a function of the expected overhead for $C = 1$ attacker node in a system of $N = 10$ nodes. An expected overhead of $E[K] = 2$ corresponds to one relay on average, while $E[K] = N$ is the maximum expected overhead for Minstrels. Fig. 2 shows results for MCrowds, and Fig. 3 shows results for Minstrels. Higher values of the assigned probabilities $P_{relB}(s,r)$ and $P_{relM}(s,r)$ mean that the sender-receiver pair is more exposed, i.e., has worse relationship anonymity. The upper bound and the lower bound are obtained by finding the distribution of F for Minstrels, and the receiver set size M for MCrowds, that results in the lowest $P(H_{1+}|S(s),R(r))$ for a given overhead.

One would expect that higher overhead always provides better relationship anonymity (i.e., low assigned probability),

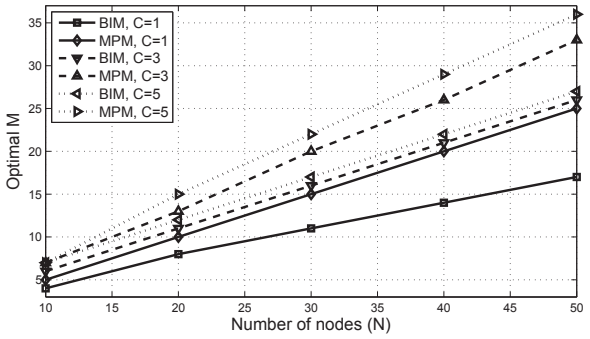

Fig. 4. Optimal M vs. number of nodes in the system

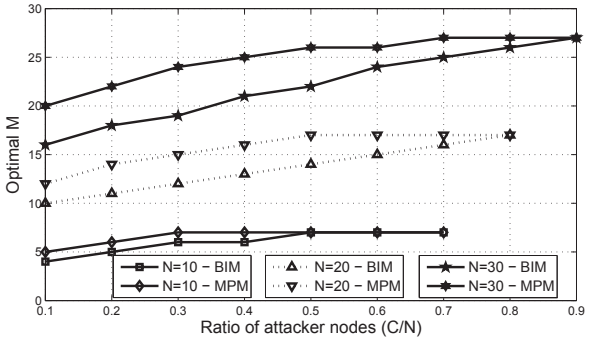

Fig. 5. Optimal M vs. ratio of attacker nodes

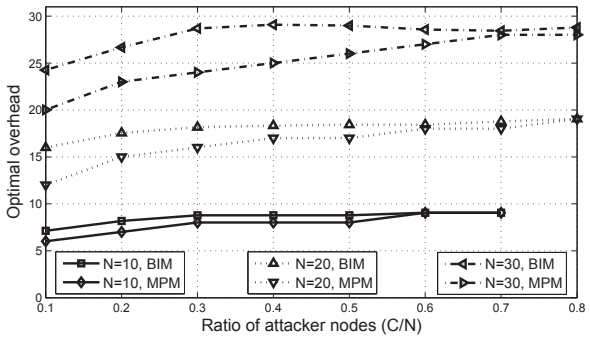

Fig. 6. Optimal overhead vs. ratio of attacker nodes for MCrowds

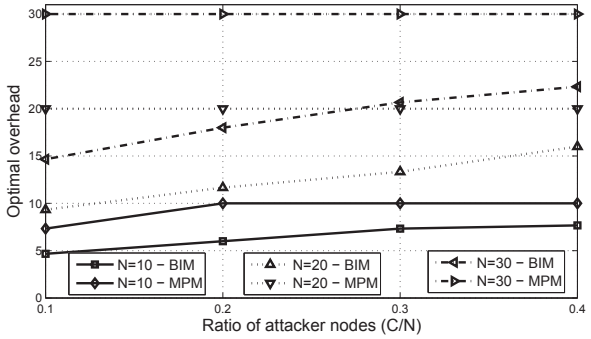

Fig. 7. Optimal overhead vs. ratio of attacker nodes for Minstrels

but surprisingly this is not the case. Above a certain level of overhead a further increase of the overhead (more relaying) has a negative effect on the relationship anonymity under the considered traffic analysis methods for both anonymity networks. The reason is that as the expected number of relays increases, the probability $P(H_{1+}|S(s),R(r))$ of having an attacker node on the path increases faster than the certainty of the attacker about the identity of the sender-receiver pair decreases. Interestingly, for MCrowds and the MP method increased overhead always results in worse relationship anonymity. We also observe that both Minstrels and MCrowds provide worse relationship anonymity under the MP method than under the BI method.

For high overhead, the anonymity provided by both anonymity networks approaches its lower bound. Despite the fact that for Minstrels the probability $P(H_{1+}|S(s),R(r))$ of having an attacker node on the path is higher than for MCrowds, Minstrels provides better relationship anonymity. The reason is that Minstrels hides the sender and the receiver among a bigger subset of nodes.

Fig. 2 suggests that MCrowds performs better for larger values of the receiver set size M. This is not true in general. For a larger M the receiver is better hidden but, at the same time, the sender is more exposed because there are fewer potential relays. Hence there should be an optimal receiver set size M. Fig. 4 shows the optimal value of M as a function of the number N of nodes in the system. The optimal receiver set size M increases both with the number of nodes in the system (almost linearly) and with the ratio $\frac{C}{N}$ of attacker nodes. The value of M used in Fig. 2 ($M = 4$ for both the BI method and the MP method) is in fact optimal for $N = 10$ and $C = 1$.

Fig. 5 shows the optimal receiver set size M as a function of the ratio $\frac{C}{N}$ of attacker nodes in the system. We can see that the optimal value of M is a non-decreasing function of the ratio of attacker nodes. For a given ratio of attacker nodes the optimal receiver set size M for the MP method is always greater or equal than the optimal M for BI method. The optimal M for the MP method and the optimal M for the BI method have the same maximum value. As the system gets larger, the highest optimal value of M for the MP method and for the BI method is reached at higher values of the ratio of attacker nodes. Hence, with more attacker nodes in the system it is better to increase the receiver set size M if it is lower than the highest optimal value.

Fig. 6 and Fig. 7 show the optimal overhead (where the probabilities $P_{relB}(s,r)$ or $P_{relM}(s,r)$ are the lowest) as a function of the ratio of attacker nodes $(\frac{C}{N})$ for MCrowds and for Minstrels, respectively. For MCrowds, the optimal overhead for both the BI method and the MP method increases with the system size N. For a given ratio of attacker nodes $\frac{C}{N}$ the optimal overhead for the BI method is greater than or equal to the optimal overhead for the MP method. It is interesting to note that for the considered system sizes N the optimal overhead is in the interval $\{2..N\}$. For Minstrels, the optimal overhead for the BI method increases with the system size N and it is lower than the optimal overhead for the MP method. The optimal overhead for MP method is equal to the maximum overhead for Minstrels ($E[K] = N$) except for $N = 10$ and $\frac{C}{N} = 0.1$.

Fig. 8 shows the probabilities $P_{relB}(s,r)$ and $P_{relM}(s,r)$ at the

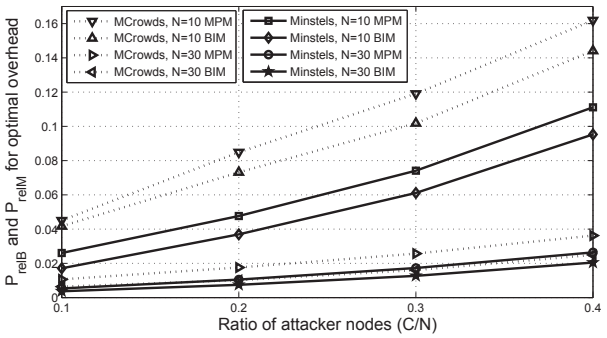

Fig. 8. Relationship anonymity for optimal overhead vs. ratio of attacker nodes

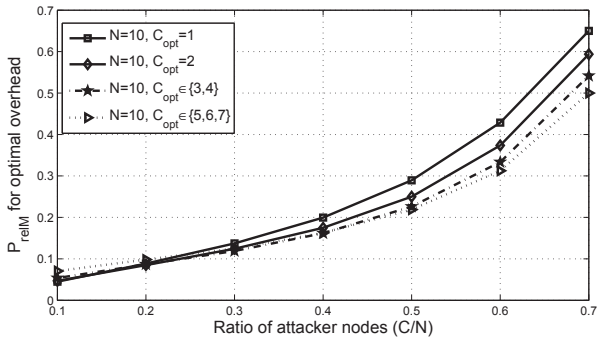

Fig. 9. $P_{relM}(s,r)$ for optimal overhead vs. ratio of attacker nodes for MCrowds

optimal overhead as a function of the ratio of attacker nodes $(\frac{C}{N})$. As the ratio of attacker nodes increases, the probabilities $P_{relB}(s,r)$ and $P_{relM}(s,r)$ increase almost linearly. However, for larger systems the probabilities are lower for the same ratio of attacker nodes. Consequently, with an increase in the system size the attacker needs to corrupt more than proportional number of nodes in order to achieve the same values of $P_{relB}(s,r)$ and $P_{relM}(s,r)$. Hence, both for Minstrels and for MCrowds, it is always beneficial to have more nodes in the network for the same ratio of attacker nodes $\frac{C}{N}$.

In practice the ratio of the attacker nodes is not known by the system designer, hence the anonymity network must be inevitably optimized for an unknown parameter. In Fig. 9 we investigate the sensitivity of the relationship anonymity to misestimating the ratio of attacker nodes. Fig. 9 shows the probability $P_{relM}(s,r)$ (MP method) as a function of the actual ratio $\frac{C}{N}$ of attacker nodes for MCrowds and $N = 10$ nodes. The expected overhead is selected to be optimal for various ratios of attacker nodes, from $\frac{C}{N} = 0.1$ to $\frac{C}{N} = 0.7$. Interestingly, $P_{relM}(s,r)$ is less sensitive to the actual ratio of attacker nodes when the anonymity network is optimized for a higher ratio of attacker nodes. The anonymity network optimized for a lower ratio of attacker nodes performs worse for higher $\frac{C}{N}$ ratios than the anonymity network optimized for a higher ratio of attacker nodes for lower $\frac{C}{N}$ ratios. Therefore, it is better to optimize the anonymity network for a higher ratio of attacker nodes than the actual ratio. We observed similar behavior for bigger system sizes N and the BI method.

The presented results lead us to the following interesting conclusions. First, best relationship anonymity might not be achieved at the highest possible overhead. The optimal overhead depends on the anonymity network, traffic analysis method, system size, and the number of attacker nodes. Second, for an attacker it is always better to use the Maximum posteriori method than the Bayesian inference method for traffic analysis in case of the MCrowds and the Minstrels anonymity networks. Third, MCrowds and Minstrels can achieve better relationship anonymity in bigger systems, but at the price of higher overhead. Fourth, when the number of attacker nodes is unknown MCrowds and Minstrels are less sensitive if they are optimized for a high ratio of attacker nodes. Fifth, for MCrowds it always beneficial to have more than one node specified as the receiver of the message ($M > 1$). Finally, for the considered system sizes N and ratios of attacker nodes $(\frac{C}{N})$, Minstrels achieves better relationship anonymity than MCrowds.

VI. Conclusions

In this paper we considered the problem of providing relationship anonymity for communication among a fixed set of nodes. We described two anonymity networks, MCrowds and Minstrels. MCrowds is an extension of Crowds, and provides unbounded path length, while Minstrels provides bounded path length. We considered two traffic analysis methods, the Bayesian inference method and the Maximum posteriori method. We found that MCrowds provides better relationship anonymity than Crowds, but in order to provide anonymity to the receiver the sender is more exposed than in Crowds. Moreover, we found that Minstrels provides better relationship anonymity than MCrowds. We used the two anonymity systems to study the trade-off between relationship anonymity and communication overhead, and found that increased overhead does not always lead to improved relationship anonymity. When comparing the two traffic analysis methods, we found that the Maximum posteriori method performs always better. We studied the way relationship anonymity scales with the number of nodes, and observed that relationship anonymity improves with the number of nodes but at the price of higher overhead. Our results also show that in practice anonymity systems should be optimized for a higher number of attackers than expected.

References

[1] D. Dzung, M. Naedele, T. V. Hoff, and M. Crevatin "Security for Industrial Communication Systems." *Proc. IEEE*, vol. 93, no. 6, pp. 1152-1177, 2005.

[2] C. W. Ten, C. C. Liu and M. Govindarasu "Vulnerability Assessment of Cybersecurity for SCADA Systems." *IEEE Trans. Power Syst.*, vol. 23, no. 4, 2008.

[3] D. Chaum "Untraceable electronic mail, return addresses and digital pseudonyms" *Commun. of the ACM*, vol. 24, no. 2, pp. 84-88, 1981

[4] A. Pfitzman, M. Köhntopp "Anonymity, unobservability, and pseudonymity - a proposal for terminology" *Anonymity 2000*, pp. 1-9, 2000

[5] P. Syverson, D. Goldschlag, and M. Reed "Anonymous connections and onion routing." *in Proc. IEEE Symp. on Security and Privacy*, pp. 44-54, May 1997.

[6] M. Reiter and A. Rubin "Crowds: Anonymity for Web Transactions." *ACM Trans. Inform. Syst. Secur.*, vol. 1, no. 1, pp. 66-92, 1998.

[7] G. Danezis, C. Díaz, E. Käsper, and C. Troncoso "The wisdom of Crowds: attacks and optimal constructions" in Proc. of ESORICS 2009.

[8] V. Shmatikov and M. H. Wang "Measuring Relationship Anonymity in Mix Networks" in Proc. of Workshop on Privacy in the Electronic Society (WPES) 2006.

[9] J. Feigenbaum, A. Johnson, and P. Syverson "Probabilistic Analysis of Onion Routing in a Black-box Model" in Proc. of Workshop on Privacy in the Electronic Society (WPES) 2007.
[10] M. Wright, M. Adler, B. N. Levine, and C. Shields "The Predecessor Attack: An Analysis of a Threat to Anonymous Communications Systems" *ACM Trans. Inform. Syst. Secur.*, vol. 7, no. 4, pp. 489-522, November 2004
[11] C. Troncoso, B. Gierlichs, B. Preneel, and I. Verbauwhede "Perfect Matching Disclosure Attacks" in Proc. of Privacy Enhancing Technologies Symposium (PETS) 2008.
[12] G. Danezis and C. Troncoso "Vida: How to use Bayesian inference to de-anonymize persistent communications" in Proc. of Privacy Enhancing Technologies Symposium (PETS) 2009.
[13] C. Troncoso and G. Danezis "The Bayesian Traffic Analysis of Mix Networks" In Proc. of Conference on Computer and Communications Security (CCS) 2009.
[14] C. A. Ardagna, S. Jajodia, P. Samarati, and A. Stavrou "Providing Mobile Users' Anonymity in Hybrid Networks" European Symposium on Research in Computer Security (ESORICS) 2010.
[15] C. Diaz, S. J. Murdoch, and C. Troncoso "Impact of Network Topology on Anonymity and Overhead in Low-Latency Anonymity Networks" in Proc. of Privacy Enhancing Technologies Symposium (PETS) 2010.

Ognjen Vuković is a PhD student in the Laboratory of Communication Networks at the KTH Royal Institute of Technology in Stockholm, Sweden. In 2010, he received his M.Sc. degree in Telecommunications, System engineering and Radio Communications, from the Faculty of Electrical Engineering, University of Belgrade.

His research interests include power system communication technologies, communication security and availability, and resource management for networked systems.

György Dán received the M.Sc. degree in computer engineering from the Budapest University of Technology and Economics, Hungary in 1999 and the M.Sc. degree in business administration from the Corvinus University of Budapest, Hungary in 2003. He worked as a consultant in the field of access networks, streaming media and videoconferencing 1999-2001. He received his Ph.D. in Telecommunications in 2006 from KTH Royal Institute of Technology, Stockholm, Sweden, where he currently works as an assistant professor. He was a visiting researcher at the Swedish Institute of Computer Science in 2008.

His research interests include cyber-physical systems security and the design and analysis of distributed and peer-to-peer systems.

Gunnar Karlsson (S'85 - M'89 - SM'99) received his Ph.D. in electrical engineering from Columbia University (1989), New York, and the M.Sc. in electrical engineering from Chalmers University of Technology in Gothenburg, Sweden (1983).

He is Professor since 1998 in the School of Electrical Engineering of KTH, the Royal Institute of Technology, in Stockholm Sweden. He is the director of the Laboratory for Communication Networks and a founding member of the KTH Linnaeus Center ACCESS. He has previously worked as Research Staff Member for IBM Zurich Research Laboratory from 1989 to 1992, and as Senior Researcher at the Swedish Institute of Computer Science (SICS) from 1992 to 1998. He has held the CLUSTER Chair visiting professorship at EPFL, Switzerland, from November 1996 to April 1997; he has been visiting professor at the Helsinki University of Technology, Finland, from June to December 1997, and at ETH Zurich in Switzerland from August 2005 to July 2006. His current research relates to quality of service, wireless LAN developments and delay-tolerant communication.

Prof. Karlsson is senior member of IEEE and member of ACM; he serves on the editorial board of IEEE Journal on Selected Areas in Communication and served on the editorial board of Elsevier Computer Networks during 2005 and 2006. He has been co-chair of the technical program committees of the Fifth International Symposium on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt 2007), the 16th International Workshop on Quality of Service (IWQoS 2008), ITC 19th Specialist Seminar on Network Usage and Traffic, and ACM Workshop on Challenged Networks (CHANTS 2009); he was both general chair and technical co-chair of the 4th COST 263 International Workshop on Quality of Future, Internet Services (QoFIS 2003). He has been guest editor for two issues of IEEE Journal on Selected Areas in Communication and three other journal issues. He serves regularly on program committees, including IEEE Infocom.

APPENDIX

In the following we show calculation of the probabilities introduced in Section IV-B2 in Table II, III, IV, and V. Moreover, we describe the probabilities $P(\Omega_s, \Omega_r, ||\mathscr{V}||, C_F, H_{1+}|S(a), R(b))$ for $||\mathscr{V}|| > 1$.

TABLE II
$P(\Omega_r, \Omega_s, ||\mathscr{V}|| > 1, C_F = 0, H_{1+}|S(s), R(r))$

Ω_s, Ω_r	
$s = p,\ r \in \mathscr{V} \setminus \{p\}$	$P(F=0)P(H(v,0\|F=0)) \frac{v-1}{(N-C-1)^2}$
$s = p,$ $r \in \overline{\mathscr{V} \cup \{p\}}$	$P(F=0)P(H(v,0\|F=0)) \frac{(N-C-v)}{(N-C-1)^2}$ $+P(F=v)P(H(v,0\|F=v))$
$s \in \mathscr{V} \setminus \{p\},$ $r = p$	$P(F=0)P(H(v,0\|F=0)) \frac{v-2}{(N-C-1)^2}$ $+ \sum_{k=1}^{v-1} P(F=k)P(H(v,0\|F=k)) \frac{1}{N-C-k}$
$s \in \mathscr{V} \setminus \{p\},$ $r \in \mathscr{V} \setminus \{p\}$	$P(F=0)P(H(v,0\|F=0)) \frac{(v-2)^2}{(N-C-1)^2}$ $+ \sum_{k=1}^{v-2} P(F=k)P(H(v,0\|F=k)) \frac{v-k-1}{N-C-k}$
$s \in \mathscr{V} \setminus \{p\},$ $r \in \overline{\mathscr{V} \cup \{p\}}$	$P(F=0)P(H(v,0\|F=0)) \frac{(N-C-v)(v-2)}{(N-C-1)^2}$ $+ \sum_{k=1}^{v-1} P(F=k)P(H(v,0\|F=k)) \frac{N-C-v}{N-C-k}$
$s \in \overline{\mathscr{V} \cup \{p\}},\ r = p$	$P(F=0)P(H(v,0\|F=0)) \frac{(N-C-v)}{(N-C-1)^2}$
$s \in \overline{\mathscr{V} \cup \{p\}},\ r \in \mathscr{V} \setminus \{p\}$	$P(F=0)P(H(v,0\|F=0)) \frac{(v-1)(N-C-v)}{(N-C-1)^2}$
$s \in \overline{\mathscr{V} \cup \{p\}},\ r \in \overline{\mathscr{V} \cup \{p\}}$	$P(F=0)P(H(v,0\|F=0)) \frac{(N-C-v)(N-C-v-1)}{(N-C-1)^2}$

TABLE III
$P(\Omega_r, \Omega_s, ||\mathscr{V}|| > 1, C_F > 0, H_{1+}|S(s), R(r))$

Ω_s, Ω_r	
$s = p,\ r \in \overline{\mathscr{V} \cup \{p\}}$	$P(F=v)P(H(v,c_F\|F=v))$
$s \in \mathscr{V} \setminus \{p\},\ r = p$	$\sum_{k=c_F+1}^{v-1} P(F=k)P(H(v,c_F\|F=k)) \frac{1}{N-C+c_F-k}$
$s \in \mathscr{V} \setminus \{p\},$ $r \in \mathscr{V} \setminus \{p\}$	$\sum_{k=c_F+1}^{v-2} P(F=k)P(H(v,c_F\|F=k)) \frac{v-k-1}{N-C+c_F-k}$
$s \in \mathscr{V} \setminus \{p\},$ $r \in \overline{\mathscr{V} \cup \{p\}}$	$\sum_{k=c_F+1}^{v-1} P(F=k)P(H(v,c_F\|F=k)) \frac{N-C+c_F-v}{N-C+c_F-k}$

When there are no initialized attackers ($C_F = 0$) the set could have been initialized with $F \in [0..||\mathscr{V}||]$ nodes. Let us first consider the case when node s is the predecessor ($s = p$) and node r is in the set ($r \in \mathscr{V} \setminus \{p\}$). For any sender-receiver pair (a,b), the prerequisite for this to happen is that node s has to be visited just before the attacker, while node r has to be either initialized or be visited. The corresponding probabilities $P(s = p, r \in \mathscr{V} \setminus \{p\}, ||\mathscr{V}|| = v > 1, C_F = 0, H_{1+}|S(a), R(b))$ are given in Table VI.

The case when node s is the predecessor ($s = p$) but node r is not in the set ($r \in \overline{\mathscr{V} \cup \{p\}}$) is similar to the previous case. The only difference is that node r has to be neither initialized nor be visited. The probabilities $P(s = p, r \in \overline{\mathscr{V} \cup \{p\}}, ||\mathscr{V}|| = v > 1, C_F = 0, H_{1+}|S(a), R(b))$ are given in Table VII.

When we have $s \in \mathscr{V} \setminus \{p\}$ and $r = p$, node s has to be either initialized or be visited, while node r has to be visited just before the attacker. The probabilities $P(s \in \mathscr{V} \setminus \{p\}, r = p, ||\mathscr{V}|| = v > 1, C_F = 0, H_{1+}|S(a), R(b))$ are given in Table VIII.

For $s \in \mathscr{V} \setminus \{p\}$ and $r \in \mathscr{V} \setminus \{p\}$, both nodes ($s$, r) have to be either initialized or be visited before the message reaches the attacker. The probabilities $P(s \in \mathscr{V} \setminus \{p\}, r \in \mathscr{V} \setminus \{p\}, ||\mathscr{V}|| = v > 1, C_F = 0, H_{1+}|S(a), R(b))$ are given in Table IX.

For the case when we have $s \in \mathscr{V} \setminus \{p\}$ and $r \in \overline{\mathscr{V} \cup \{p\}}$, the only difference from the case above is that node r must not have been initialized or visited. The probabilities $P(s \in \mathscr{V} \setminus \{p\}, r \in \overline{\mathscr{V} \cup \{p\}}, ||\mathscr{V}|| = v > 1, C_F = 0, H_{1+}|S(a), R(b))$ are given in Table X.

When we have the opposite case of the above, $s \in \overline{\mathscr{V} \cup \{p\}}$ and $r \in \mathscr{V} \setminus \{p\}$, the same reasoning applies but in this case node s must not have been initialized or visited, and node r has to be either initialized or visited before the message reaches the attacker. The probabilities $P(s \in \overline{\mathscr{V} \cup \{p\}}, r \in \mathscr{V} \setminus \{p\}, ||\mathscr{V}|| = v > 1, C_F = 0, H_{1+}|S(a), R(b)$ are given in Table XI.

For $s \in \overline{\mathscr{V} \cup \{p\}}$ and $r = p$, node s must not have been ini-

TABLE IV
$P(\Omega_r, \Omega_s, ||\mathscr{V}|| = 0, C_F = 0, H_{1+}|S(a), R(b))$

Ω_s, Ω_r, a, b	
$s = p,\ r \in \overline{\mathscr{V} \cup \{p\}},\ a = s,\ \forall b$	$P(F=0)P(H(0,0\|F=0))$

TABLE V
$P(\Omega_r, \Omega_s, ||\mathscr{V}|| = 1, C_F = 0, H_{1+}|S(a), R(b))$

Ω_s, Ω_r, a, b	
$s = p,\ r \in \overline{\mathscr{V} \cup \{p\}},\ a = s,\ \forall b$	$P(F=1)P(H(1,0\|F=1))$
$s = p,\ r \in \overline{\mathscr{V} \cup \{p\}},\ a \neq s,\ \forall b$	$P(F=0)P(H(1,0\|F=0)) \frac{1}{N-C-1}$
$s \in \overline{\mathscr{V} \cup \{p\}},\ r = p,\ a = r,\ \forall b$	$P(F=1)P(H(1,0\|F=1))$
$s \in \overline{\mathscr{V} \cup \{p\}},\ r = p,\ a \neq r,\ \forall b$	$P(F=0)P(H(1,0\|F=0)) \frac{1}{N-C-1}$
$s \in \overline{\mathscr{V} \cup \{p\}},\ r \in \overline{\mathscr{V} \cup \{p\}},$ $a \in \{s,r\},\ \forall b$	$P(F=0)P(H(1,0\|F=0)) \frac{N-C-2}{N-C-1}$
$s \in \overline{\mathscr{V} \cup \{p\}},\ r \in \overline{\mathscr{V} \cup \{p\}},$ $a \notin \{s,r\},\ \forall b$	$P(F=0)P(H(1,0\|F=0)) \frac{N-C-3}{N-C-1}$ $+P(F=1)P(H(1,0\|F=1))$

TABLE VI
$P(s = p, r \in \mathscr{V} \setminus \{p\}, ||\mathscr{V}|| = v > 1, C_F = 0, H_{1+}|S(a), R(b))$

a, b	
$a = s,$ $b \neq r$	$P(F=0)P(H(v,0\|F=0)) \frac{v-1}{(N-C-1)^2}$ $+P(F=v)P(H(v,0\|F=v)) \frac{v-1}{N-C-2}$
$a = r,$ $\forall b$	$P(F=0)P(H(v,0\|F=0)) \frac{v-2}{(N-C-1)^2}$ $+ \sum_{k=1}^{v-1} P(F=k)P(H(v,0\|F=k)) \frac{1}{N-C-1}$
$a \notin \{s,r\},$ $b = s$	$P(F=0)P(H(v,0\|F=0)) \left(\frac{1}{(N-C-1)^2} + \frac{(N-C-3)(v-2)}{(N-C-1)^2(N-C-2)} \right)$ $+ \sum_{k=1}^{v-1} P(F=k)P(H(v,0\|F=k)) \frac{v-2}{(N-C-2)(N-C-k)}$
$a \notin \{s,r\},$ $b = r$	$P(F=0)P(H(v,0\|F=0)) \left(\frac{1}{(N-C-1)^2} + \frac{(N-C-3)(v-2)}{(N-C-1)^2(N-C-2)} \right)$ $+P(F=1)P(H(v,0\|F=1)) \frac{v-2}{(N-C-1)(N-C-2)}$ $+ \sum_{k=2}^{v-1} P(F=k)P(H(v,0\|F=k)) \frac{v-k-1}{(N-C-2)^2}$
$a \notin \{s,r\},$ $b \notin \{s,r\}$	$P(F=0)P(H(v,0\|F=0)) \left(\frac{1}{(N-C-1)^2} + \frac{(N-C-3)(v-2)}{(N-C-1)^2(N-C-2)} \right)$ $+ \sum_{k=1}^{v-1} P(F=k)P(H(v,0\|F=k)) \cdot$ $\left(\frac{(k-1)(N-C-k-1)}{(N-C-2)(N-C-3)(N-C-k)} + \frac{(v-k-1)(N-C-k-2)}{(N-C-2)(N-C-3)(N-C-k)} \right)$

TABLE VII
$P(s = p, r \in \overline{\mathscr{V} \cup \{p\}}, ||\mathscr{V}|| = v > 1, C_F = 0, H_{1+}|S(a), R(b))$

a, b	
$a = s,$ $b \neq r$	$P(F=0)P(H(v,0\|F=0)) \frac{N-C-v}{(N-C-1)^2}$ $+P(F=v)P(H(v,0\|F=v)) \frac{N-C-v-1}{N-C-2}$
$a = r, \forall b$	$P(F=0)P(H(v,0\|F=0)) \frac{N-C-v}{(N-C-1)^2}$
$a \notin \{s,r\},$ $b \in \{s,r\}$	$P(F=0)P(H(v,0\|F=0)) \frac{(N-C-3)(N-C-v)}{(N-C-1)^2(N-C-2)}$ $+ \sum_{k=1}^{v-1} P(F=k)P(H(v,0\|F=k)) \frac{N-C-v}{(N-C-2)(N-C-k)}$
$a \notin \{s,r\},$ $b \notin \{s,r\}$	$P(F=0)P(H(v,0\|F=0)) \frac{(N-C-3)(N-C-v)}{(N-C-1)^2(N-C-2)}$ $+ \sum_{k=1}^{v-1} P(F=k)P(H(v,0\|F=k)) \frac{(N-C-k-2)(N-C-v)}{(N-C-2)(N-C-3)(N-C-k)}$

TABLE VIII
$P(s \in \mathscr{V} \setminus \{p\}, r = p, ||\mathscr{V}|| = v > 1, C_F = 0, H_{1+}|S(a), R(b))$

a,b	
$a = s, \forall b$	$P(F=0)P(H(v,0\|F=0)) \frac{v-2}{(N-C-1)^2}$
	$+\sum_{k=1}^{v-1} P(F=k)P(H(v,0\|F=k)) \frac{N-C-k-1}{N-C-2} \frac{1}{N-C-k}$
$a = r, b = s$	$P(F=0)P(H(v,0\|F=0)) \frac{v-1}{(N-C-1)^2}$
$a = r,$	$P(F=0)P(H(v,0\|F=0)) \frac{v-1}{(N-C-1)^2}$
$b \neq s$	$+P(F=v)P(H(v,0\|F=v)) \frac{v-1}{N-C-2}$
$a \notin \{s,r\},$	$P(F=0)P(H(v,0\|F=0)) \left(\frac{1}{(N-C-1)^2} + \frac{(N-C-3)(v-2)}{(N-C-1)^2(N-C-2)} \right)$
$b = r$	$+\sum_{k=1}^{v-1} P(F=k)P(H(v,0\|F=k)) \frac{v-2}{(N-C-2)(N-C-k)}$
$a \notin \{s,r\},$	$P(F=0)P(H(v,0\|F=0)) \left(\frac{1}{(N-C-1)^2} + \frac{(N-C-3)(v-2)}{(N-C-1)^2(N-C-2)} \right)$
$b = s$	$+P(F=1)P(H(v,0\|F=1)) \frac{v-2}{(N-C-1)(N-C-2)}$
	$+\sum_{k=2}^{v-1} P(F=k)P(H(v,0\|F=k)) \frac{v-k-1}{(N-C-2)^2}$
$a \notin \{s,r\},$	$P(F=0)P(H(v,0\|F=0)) \left(\frac{1}{(N-C-1)^2} + \frac{(N-C-3)(v-2)}{(N-C-1)^2(N-C-2)} \right)$
$b \notin \{s,r\}$	$+\sum_{k=1}^{v-1} P(F=k)P(H(v,0\|F=k))\cdot$
	$\left(\frac{(k-1)(N-C-k-1)}{(N-C-2)(N-C-3)(N-C-k)} + \frac{(v-k-1)(N-C-k-2)}{(N-C-2)(N-C-3)(N-C-k)} \right)$

TABLE IX
$P(s \in \mathscr{V} \setminus \{p\}, r \in \mathscr{V} \setminus \{p\}, ||\mathscr{V}|| = v > 1, C_F = 0, H_{1+}|S(a), R(b))$

a,b	
$a = s,\ b = r$	$P(F=0)P(H(v,0\|F=0)) \frac{(v-2)^2}{(N-C-1)^2}$
$a = r,\ b = s$	$+\sum_{k=1}^{v-2} P(F=k)P(H(v,0\|F=k)) \frac{v-k-1}{N-C-k}$
$a = s,\ b \neq r$	$P(F=0)P(H(v,0\|F=0)) \frac{(v-2)^2}{(N-C-1)^2}$
$a = r,\ b \neq s$	$+\sum_{k=1}^{v-1} P(F=k)P(H(v,0\|F=k))\cdot$
	$\left(\frac{k-1}{N-C-2} + \frac{(v-k-1)(N-C-k-1)}{(N-C-2)(N-C-k)} \right)$
$a \notin \{s,r\},$	$P(F=0)P(H(v,0\|F=0))\cdot$
$b \in \{s,r\}$	$\left(\frac{2(v-2)}{(N-C-1)^2} + \frac{(v-2)(v-3)(N-C-3)}{(N-C-1)^2(N-C-2)} \right)$
$v > 2$	$+P(F=1)P(H(v,0\|F=1)) \frac{(v-2)(v-3)}{(N-C-1)(N-C-2)}$
	$+\sum_{k=2}^{v-3} P(F=k)P(H(v,0\|F=k)) \frac{(v-k-1)^2}{(N-C-2)(N-C-k)}$
	$+P(F=v-2)P(H(v,0\|F=v-2)) \frac{v-3}{(N-C-2)(N-C-v+2)}$
$a \notin \{s,r\},$	$P(F=0)P(H(v,0\|F=0))\cdot$
$b \notin \{s,r\}$	$\left(\frac{2(v-2)}{(N-C-1)^2} + \frac{(v-2)(v-3)(N-C-3)}{(N-C-1)^2(N-C-2)} \right)$
$v > 2$	$\sum_{k=1}^{v-1} P(F=k)P(H(v,0\|F=k)) \left(\frac{(k-1)(k-2)}{(N-C-2)(N-C-3)} \right.$
	$\left. \frac{(v-k-1)(v-k-2)(N-C-k-2)}{(N-C-k)(N-C-k-1)(N-C-3)} + \frac{2(N-C-k-1)(k-1)(v-k-1)}{(N-C-2)(N-C-3)(N-C-k)} \right)$
	$+P(F=v)P(H(v,0\|F=v)) \frac{(v-1)(v-2)}{(N-C-2)(N-C-3)}$

TABLE X
$P(s \in \mathscr{V} \setminus \{p\}, r \in \overline{\mathscr{V} \cup \{p\}}, ||\mathscr{V}|| = v > 1, C_F = 0, H_{1+}|S(a), R(b))$

a,b	
$a = s,$	$P(F=0)P(H(v,0\|F=0)) \frac{(v-2)(N-C-v)}{(N-C-1)^2}$
$b \neq r$	$+\sum_{k=1}^{v-1} P(F=k)P(H(v,0\|F=k)) \frac{(N-C-k-1)(N-C-v)}{(N-C-2)(N-C-k)}$
$a = r,\ \forall b$	$P(F=0)P(H(v,0\|F=0)) \frac{(v-1)(N-C-v)}{(N-C-1)^2}$
$a \notin \{s,r\},$	$P(F=0)P(H(v,0\|F=0))\cdot$
$b = s$	$\left(\frac{N-C-v}{(N-C-1)^2} + \frac{(N-C-v)(N-C-3)(v-2)}{(N-C-1)^2(N-C-2)} \right)$
	$+\sum_{k=1}^{v-2} P(F=k)P(H(v,0\|F=k)) \frac{(v-k-1)(N-C-v)}{(N-C-2)(N-C-k)}$
$a \notin \{s,r\},$	$P(F=0)P(H(v,0\|F=0))\cdot$
$b = r$	$\left(\frac{N-C-v}{(N-C-1)^2} + \frac{(N-C-v)(N-C-3)(v-2)}{(N-C-1)^2(N-C-2)} \right)$
	$+\sum_{k=1}^{v-1} P(F=k)P(H(v,0\|F=k)) \frac{(v-k-1)(N-C-v)}{(N-C-2)(N-C-k)}$
	$+P(F=v)P(H(v,0\|F=v)) \frac{v-1}{N-C-2}$
$a \notin \{s,r\},$	$P(F=0)P(H(v,0\|F=0))\cdot$
$b \notin \{s,r\}$	$\left(\frac{N-C-v}{(N-C-1)^2} + \frac{(N-C-v)(N-C-3)(v-2)}{(N-C-1)^2(N-C-2)} \right)$
	$+\sum_{k=1}^{v-1} P(F=k)P(H(v,0\|F=k))\cdot$
	$\left(\frac{(k-1)(N-C-k-1)(N-C-v)}{(N-C-2)(N-C-3)(N-C-k)} + \frac{(v-k-1)(N-C-k-2)(N-C-v)}{(N-C-2)(N-C-3)(N-C-k)} \right)$
	$+P(F=v)P(H(v,0\|F=v)) \frac{(v-1)(N-C-v-1)}{(N-C-2)(N-C-3)}$

TABLE XI
$P(s \in \overline{\mathscr{V} \cup \{p\}}, r \in \mathscr{V} \setminus \{p\}, ||\mathscr{V}|| = v > 1, C_F = 0, H_{1+}|S(a), R(b))$

a,b	
$a = s,\ \forall b$	$P(F=0)P(H(v,0\|F=0)) \frac{(v-1)(N-C-v)}{(N-C-1)^2}$
$a = r,$	$P(F=0)P(H(v,0\|F=0)) \frac{(v-2)(N-C-v)}{(N-C-1)^2}$
$b = s$	$+\sum_{k=1}^{v-1} P(F=k)P(H(v,0\|F=k)) \frac{N-C-v}{N-C-k}$
$a = r,$	$P(F=0)P(H(v,0\|F=0)) \frac{(v-2)(N-C-v)}{(N-C-1)^2}$
$b \neq s$	$+\sum_{k=1}^{v-1} P(F=k)P(H(v,0\|F=k)) \frac{(N-C-k-1)(N-C-v)}{(N-C-2)(N-C-k)}$
$a \notin \{s,r\},$	$P(F=0)P(H(v,0\|F=0))\cdot$
$b = s$	$\left(\frac{N-C-v}{(N-C-1)^2} + \frac{(N-C-v)(N-C-3)(v-2)}{(N-C-1)^2(N-C-2)} \right)$
	$+\sum_{k=1}^{v-1} P(F=k)P(H(v,0\|F=k)) \frac{(v-k-1)(N-C-v)}{(N-C-2)(N-C-k)}$
	$+P(F=v)P(H(v,0\|F=v)) \frac{v-1}{N-C-2}$
$a \notin \{s,r\},$	$P(F=0)P(H(v,0\|F=0))\cdot$
$b = r$	$\left(\frac{N-C-v}{(N-C-1)^2} + \frac{(N-C-v)(N-C-3)(v-2)}{(N-C-1)^2(N-C-2)} \right)$
	$+\sum_{k=1}^{v-2} P(F=k)P(H(v,0\|F=k)) \frac{(v-k-1)(N-C-v)}{(N-C-2)(N-C-k)}$
$a \notin \{s,r\},$	$P(F=0)P(H(v,0\|F=0))\cdot$
$b \notin \{s,r\}$	$\left(\frac{N-C-v}{(N-C-1)^2} + \frac{(N-C-v)(N-C-3)(v-2)}{(N-C-1)^2(N-C-2)} \right)$
	$+\sum_{k=1}^{v-1} P(F=k)P(H(v,0\|F=k))\cdot$
	$\left(\frac{(k-1)(N-C-k-1)(N-C-v)}{(N-C-2)(N-C-3)(N-C-k)} + \frac{(v-k-1)(N-C-k-2)(N-C-v)}{(N-C-2)(N-C-3)(N-C-k)} \right)$
	$+P(F=v)P(H(v,0\|F=v)) \frac{(v-1)(N-C-v-1)}{(N-C-2)(N-C-3)}$

tialized or visited, while node r has to be visited just before the attacker. The corresponding probabilities $P(s \in \overline{\mathscr{V} \cup \{p\}}, r = p, ||\mathscr{V}|| = v > 1, C_F = 0, H_{1+}|S(a), R(b))$ are given in Table XII.

Finally, for the case when neither s nor r are in the set ($s \in \overline{\mathscr{V} \cup \{p\}}$, $r \in \overline{\mathscr{V} \cup \{p\}}$), they must not have been initialized or visited. The probabilities $P(s \in \overline{\mathscr{V} \cup \{p\}}, r \in \overline{\mathscr{V} \cup \{p\}}, ||\mathscr{V}|| = v > 1, C_F = 0, H_{1+}|S(a), R(b))$ are given in Table XIII.

Until now we considered the cases when there are no initialized attackers in the set of visited nodes ($C_F = 0$). However, the attacker can receive a message with $||\mathscr{V}|| = v > 1$ visited nodes and with $C_F = c_F > 0$ initialized attackers. In this case the sender node must have initialized the set with c_F attackers. Hence $F \in [c_F + 1..v]$. Let us now consider different values of Ω_s and Ω_r. For $s = p$ and $r \in \overline{\mathscr{V} \cup \{p\}}$, node s has to be visited just before the attacker. At the same time, node r must not have been initialized or visited. The corresponding probabilities $P(s = p, r \in \overline{\mathscr{V} \cup \{p\}}, ||\mathscr{V}|| = v > 1, C_F = c_F > 0, H_{1+}|S(a), R(b))$ are given in Table XIV.

A similar reasoning applies when we have $s \in \mathscr{V} \setminus \{p\}$ and $r = p$. Node s has to be either initialized or visited, while node r has to appear as the predecessor. The probabilities $P(s \in \mathscr{V} \setminus \{p\}, r = p, ||\mathscr{V}|| = v > 1, C_F = c_F > 0, H_{1+}|S(a), R(b))$ are given in Table XV.

When nodes s and r are both in the set ($s \in \mathscr{V} \setminus \{p\}$,$r \in \mathscr{V} \setminus \{p\}$), the sender a must have initialized them or the

TABLE XII

$P(s \in \overline{\mathscr{V} \cup \{p\}}, r = p, ||\mathscr{V}|| = v > 1, C_F = 0, H_{1+}|S(a), R(b))$

a,b	
$a = s, \forall b$	$P(F=0)P(H(v,0\|F=0))\frac{N-C-v}{(N-C-1)^2}$
$a = r,$ $b = s$	$P(F=0)P(H(v,0\|F=0))\frac{N-C-v}{(N-C-1)^2}$ $+P(F=v)P(H(v,0\|F=v))$
$a = r,$ $b \neq s$	$P(F=0)P(H(v,0\|F=0))\frac{N-C-v}{(N-C-1)^2}$ $+P(F=v)P(H(v,0\|F=v))\frac{N-C-v-1}{N-C-2}$
$a \notin \{s,r\},$ $b \in \{s,r\}$	$P(F=0)P(H(v,0\|F=0))\frac{(N-C-3)(N-C-v)}{(N-C-1)^2(N-C-2)}$ $+\sum_{k=1}^{v-1}P(F=k)P(H(v,0\|F=k))\frac{N-C-v}{(N-C-2)(N-C-k)}$
$a \notin \{s,r\},$ $b \notin \{s,r\}$	$P(F=0)P(H(v,0\|F=0))\frac{(N-C-3)(N-C-v)}{(N-C-1)^2(N-C-2)}$ $+\sum_{k=1}^{v-1}P(F=k)P(H(v,0\|F=k))\frac{(N-C-k-2)(N-C-v)}{(N-C-2)(N-C-3)(N-C-k)}$

TABLE XIII

$P(s \in \overline{\mathscr{V} \cup \{p\}}, r \in \overline{\mathscr{V} \cup \{p\}}, ||\mathscr{V}|| = v > 1, C_F = 0, H_{1+}|S(a), R(b))$

a,b	
$a \in \{s,r\},\ \forall b$	$P(F=0)P(H(v,0\|F=0))\frac{(N-C-v)(N-C-v-1)}{(N-C-1)^2}$
$a \notin \{s,r\},$ $b \in \{s,r\}$	$P(F=0)P(H(v,0\|F=0))\frac{(N-C-3)(N-C-v)(N-C-v-1)}{(N-C-1)^2(N-C-2)}$ $+\sum_{k=1}^{v}P(F=k)P(H(v,0\|F=k))\frac{(N-C-v)(N-C-v-1)}{(N-C-2)(N-C-k)}$
$a \notin \{s,r\},$ $b \notin \{s,r\}$	$P(F=0)P(H(v,0\|F=0))\frac{(N-C-3)(N-C-v)(N-C-v-1)}{(N-C-1)^2(N-C-2)}$ $+\sum_{k=1}^{v}P(F=k)P(H(v,0\|F=k))\cdot$ $\frac{(N-C-v)(N-C-v-1)(N-C-k-2)}{(N-C-2)(N-C-3)(N-C-k)}$

TABLE XIV

$P(s = p, r \in \overline{\mathscr{V} \cup \{p\}}, ||\mathscr{V}|| = v > 1, C_F = c_F > 0, H_{1+}|S(a), R(b))$

a,b	
$a = s,\ b \neq r$	$P(F=v)P(H(v,c_F\|F=v))\frac{N-C-v-1+c_F}{N-C-2}$
$a \notin \{s,r\},$ $b \in \{s,r\}$	$\sum_{k=c_F+1}^{v-1}P(F=k)P(H(v,c_F\|F=k))\cdot$ $\frac{N-C-v+c_F}{(N-C-k+c_F)(N-C-2)}$
$a \notin \{s,r\},$ $b \notin \{s,r\}$	$\sum_{k=c_F+1}^{v-1}P(F=k)P(H(v,c_F\|F=k))\cdot$ $\frac{(N-C-v+c_F)(N-C-k-2+c_F)}{(N-C-k+c_F)(N-C-2)(N-C-3)}$

TABLE XV

$P(s \in \mathscr{V} \setminus \{p\}, r = p, ||\mathscr{V}|| = v > 1, C_F = c_F > 0, H_{1+}|S(a), R(b))$

a,b	
$a = s,\ b \neq r$	$\sum_{k=c_F+1}^{v-1}P(F=k)P(H(v,c_F\|F=k))\cdot$ $\frac{N-C-k+c_F-1}{(N-C-k+c_F)(N-C-2)}$
$a = r,\ b \neq s$	$P(F=v)P(H(v,c_F\|F=v))\frac{v-1-c_F}{N-C-2}$
$a \notin \{s,r\},\ b = s$	$\sum_{k=c_F+1}^{v-2}P(F=k)P(H(v,c_F\|F=k))\cdot$ $\frac{v-1-k}{(N-C-k+c_F)(N-C-2)}$
$a \notin \{s,r\},\ b = r$	$\sum_{k=c_F+1}^{v-1}P(F=k)P(H(v,c_F\|F=k))\cdot$ $\frac{v-c_F-2}{(N-C-k+c_F)(N-C-2)}$
$a \notin \{s,r\},$ $b \notin \{s,r\}$	$\sum_{k=c_F+1}^{v-1}P(F=k)P(H(v,c_F\|F=k))\cdot$ $\left(\frac{(N-C-k+c_F-1)(k-c_F-1)+(N-C-k+c_F-2)(v-k-1)}{(N-C-k+c_F)(N-C-2)(N-C-3)}\right)$

message must have visited them. The corresponding probabilities $P(s \in \mathscr{V} \setminus \{p\}, r \in \mathscr{V} \setminus \{p\}, ||\mathscr{V}|| = v > 1, C_F = c_F > 0, H_{1+}|S(a), R(b))$ are given in Table XVI.

For $s \in \mathscr{V} \setminus \{p\}$ and $r \in \overline{\mathscr{V} \cup \{p\}}$, the sender a must have initialized node s or the message must have visited it before the attacker received the message. At the same time, node r must not have been initialized or visited. The corresponding probabilities $P(s \in \mathscr{V} \setminus \{p\}, r \in \overline{\mathscr{V} \cup \{p\}}, ||\mathscr{V}|| = v > 1, C_F = c_F > 0, H_{1+}|S(a), R(b))$ are given in Table XVII.

TABLE XVI

$P(s \in \mathscr{V} \setminus \{p\}, r \in \mathscr{V} \setminus \{p\}, ||\mathscr{V}|| = v > 1, C_F = c_F > 0, H_{1+}|S(a), R(b))$

a,b	
$a = s,\ b \neq r$ $a = r,\ b \neq s$	$\sum_{k=c_F+1}^{v-1}P(F=k)P(H(v,c_F\|F=k))\cdot$ $\left(\frac{(N-C-k+c_F-1)(v-k-1)}{(N-C-k+c_F)(N-C-2)}+\frac{k-c_F-1}{N-C-2}\right)$
$a = r,\ b = s$	$\sum_{k=c_F+1}^{v-1}P(F=k)P(H(v,c_F\|F=k))\frac{v-k-1}{N-C-k+c_F}$
$a \notin \{s,r\},$ $b \in \{s,r\}$	$\sum_{k=c_F+1}^{v-2}P(F=k)P(H(v,c_F\|F=k))\frac{v-k-1}{N-C-k+c_F}\cdot$ $\left(\frac{(N-C-k+c_F-1)(v-k-2)}{(N-C-k+c_F-1)(N-C-2)}+\frac{k-c_F-1}{N-C-2}\right)$
$a \notin \{s,r\},$ $b \notin \{s,r\}$	$\sum_{k=c_F+1}^{v}P(F=k)P(H(v,c_F\|F=k))$ $\left(\frac{(k-c_F-1)(k-c_F-2)}{(N-C-2)(N-C-3)}+\frac{(N-C-k+c_F-2)(v-k-1)}{(N-C-k+c_F)(N-C-2)(N-C-3)}\right.$ $\left.+\frac{(N-C-k+c_F-1)(v-k-1)(k-c_F-1)}{(N-C-k+c_F)(N-C-2)(N-C-3)}\right)$

TABLE XVII

$P(s \in \mathscr{V} \setminus \{p\}, r \in \overline{\mathscr{V} \cup \{p\}}, ||\mathscr{V}|| = v > 1, C_F = c_F > 0, H_{1+}|S(a), R(b))$

a,b	
$a = s,\ b \neq r$	$\sum_{k=c_F+1}^{v-1}P(F=k)P(H(v,c_F\|F=k))\cdot$ $\frac{(N-C-k+c_F-1)(N-C+c_F-v)}{(N-C-k+c_F)(N-C-2)}$
$a \notin \{s,r\}, b = s$	$\sum_{k=c_F+1}^{v-2}P(F=k)P(H(v,c_F\|F=k))$ $\frac{(N-C+c_F-v)(v-k-1)}{(N-C-k+c_F)(N-C-2)}$
$a \notin \{s,r\}, b = r$	$\sum_{k=c_F+1}^{v-1}P(F=k)P(H(v,c_F\|F=k))\frac{(N-C+c_F-v)}{(N-C-k+c_F)}\cdot$ $\left(\frac{k-c_F-1}{N-C-2}+\frac{(N-C-k+c_F-1)(v-k-1)}{(N-C-2)(N-C-k+c_F)}\right)$ $+P(F=v)P(H(v,c_F\|F=v))\frac{v-c_F-1}{N-C-2}$
$a \notin \{s,r\},$ $b \notin \{s,r\}$	$\sum_{k=c_F+1}^{v-1}P(F=k)P(H(v,c_F\|F=k))\cdot$ $\frac{(N-C+c_F-v)(N-C+c_F-k-1)}{(N-C-2)(N-C-k+c_F)}\cdot$ $\left(\frac{k-c_F-1}{N-C-3}+\frac{(N-C-k+c_F-2)(v-k-1)}{(N-C-3)(N-C-k+c_F)}\right)$ $+P(F=v)P(H(v,c_F\|F=v))\frac{(N-C-v+c_F-1)(v-c_F-1)}{(N-C-2)(N-C-3)}$