

Mitigating Denial of Service Attacks using Anonymity Networks: Relationship Anonymity-Communication Overhead Trade-off

Ognjen Vuković, György Dán, Gunnar Karlsson

***Abstract*—Denial-of-service attacks are a significant threat to mission critical communication infrastructures, e.g., to industrial control systems. They are relatively easy to perpetrate, as an attacker that has access to communication links or equipment could observe the source and destination addresses for every message, and can identify and discard the messages exchanged between particular communication participants. Mix networks and anonymity networks could render these attacks more difficult by providing anonymous communication via relaying. Nevertheless, relaying introduces overhead and increases the end-to-end message delivery delay, which in practice must often be low. Hence, an important question is how to optimize anonymity for limited overhead and delay. In this paper we address this question by studying 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 on the optimal anonymity network parameters, and the sensitivity of anonymity to the misestimation of the number of attackers.**

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 denial-of-service (DoS) attacks on the communication between any two nodes. It may, for example, drop messages carrying important status or control information. Such an attack can lead to incorrect system operation, e.g., it can destabilize a modern industrial control system [3]–[5], and it may be hard to prevent and to detect [6].

Mix networks [7] are a way to mitigate outside attacks by providing relationship anonymity, i.e., by making it untraceable who communicates with whom [8]. Mix networks consist of a set of mixes that relay messages in such a way that an outside attacker cannot link an outgoing message with an incoming message, and therefore ensures sender-receiver unlinkability against an eavesdropper observing communication links. 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 also mitigate against outside attacks but also provide some level of relationship anonymity against inside attackers (e.g., [9], [10]) 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 [11], 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 challenging.

We consider an attacker that wants to perform a DoS attack on the communication between a particular pair of nodes by dropping the messages that they exchange. To defend against such attacks, we use two anonymity networks that provide relationship anonymity. First, MCrowds, a modification of Crowds [10], 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 [12], but, unlike Crowds, it also hides the receiver among a small subset of

The authors are with ACCESS Linnaeus Center, School of Electrical Engineering, KTH Royal Institute of Technology, Stockholm, Sweden. E-mail: {vukovic,gyuri,gk}@ee.kth.se

anonymity network nodes. 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. We use these two anonymity networks to investigate the inherent trade-off between the communication overhead introduced and the level of provided relationship anonymity. 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.

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

II. Related Work

Early works on traffic analysis attacks against anonymity networks by an external global attacker considered long term intersection attacks [11], [13], [14]. 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 [15] 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 [11], [16], [17]. 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 [18] 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 [12] and for systems inspired by Crowds [18]. In our work, we consider relationship anonymity instead of sender anonymity, and address the trade-off between anonymity and overhead.

Related to our work are studies on DoS attacks [6], particularly DoS attacks in industrial control systems [4], [19]–[22]. In [6], the authors present taxonomies for classifying DoS attacks and defenses in any networked system. DoS attacks against industrial control systems can significantly degrade the performance of such systems [20], and even destabilize them, e.g., power systems in [22]. There have been a number of techniques proposed for detection of DoS attacks caused by malicious communication nodes flooding network with packets to cause congestions [6], [19], [20]. To protect against such attacks, the system can identify the source of the attack, i.e., the flooding node, and filter the traffic coming from the node at the point where the traffic enters the network [6], [19], [20]. In the case of DoS attacks that result in packet loss on links, e.g., due to link jamming or intentional message dropping, the system can optimize the control loop in order to decrease effects of the attacks [4], [21]. In our work, we protect the system against targeted message dropping attacks by using anonymity networks: anonymity networks make the attacker uncertain about the sender-receiver pair for the messages it observes, and therefore, renders the targeted attacks much more difficult.

III. 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: messages can be exchanged between any two nodes without visiting other nodes. 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.

We use s ($s \in \mathcal{N}$) to denote the node that originates a message, i.e., the sender, and r ($r \in \mathcal{N} \setminus \{s\}$) to denote the node for which the message is intended, i.e., the receiver. We use a and b to denote any two nodes in the network ($a \in \mathcal{N}, b \in \mathcal{N} \setminus \{a\}$), including the sender and the receiver, and we use $(a \rightarrow b)$ for a sender-receiver pair ($a \in \mathcal{N}, b \in \mathcal{N} \setminus \{a\}$).

A. Attack Model

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 $T(S(a), R(b))$ 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). Entry $T(S(a), R(b))$ of the traffic matrix is the message sending rate from a to b normalized by the total message rate. For example, the distribution T could be uniform if the attacker has no a-priori knowledge of the actual traffic matrix.

The aim of the attacker is to perform a targeted attack on the communication between a particular pair of nodes,

which we refer to as the *targeted s-r pair*. In principle, the attacker could drop every message that gets relayed over the nodes $\mathscr{C}$ it controls to maximize the effect of the attack, but then such an attack could be detected easier as no message would ever been successfully relayed over the nodes in $\mathscr{C}$. Instead, for every message it observes, the attacker decides whether to drop or to continue relaying the message based on its belief that the message is part of the communication between the targeted s-r pair. The belief can be formulated as the *a-posteriori* probability $P(\hat{S}(a),\hat{R}(b))$ that the attacker assigns to the targeted s-r pair being the sender and the receiver of the observed message. The attacker decides to drop the message with a probability that is a function of the belief, i.e., the message is dropped with probability $g(P(\hat{S}(a),\hat{R}(b)))$ for the targeted s-r pair. We consider two attack methods that differ in the function $g(P(\hat{S}(a),\hat{R}(b)))$, and we describe them in Section III-B1 and Section III-B2. We show the efficiency of the targeted attack as a part of numerical results in Section VI.

B. Overhead and Anonymity Metrics

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 probability that a message sent from s to r is dropped when the (s,r) pair is the targeted s-r pair, i.e., the expected true positive rate. The lower the relationship anonymity is, the more difficult it is for the attacker to perform a successful targeted message dropping attack. Note that the relationship anonymity may not be the same for (s,r) and for (r,s). In general, the relationship anonymity depends on three 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. Third, on the function $g(P(\hat{S}(s),\hat{R}(r)))$. The first two factors are functions of the anonymity protocol, the number of nodes N and the number of inside attacker nodes C. The function $g(P(\hat{S}(s),\hat{R}(r)))$ depends on the method used by the attacker. We consider the following two methods.

1) Maximum posteriori method: Using the Maximum Posteriori (MP) method, when the attacker intercepts a message, it populates the set $\mathscr{Q} = \{(a,b) : P(\hat{S}(a),\hat{R}(b)) \geq P(\hat{S}(a),\hat{R}(b))\}$ of most likely sender-receiver pairs. If $(s,r) \in \mathscr{Q}$ then the attacker drops the message with probability $1/||\mathscr{Q}||$. Thus, $g(P(\hat{S}(a),\hat{R}(b))) = 1/||\mathscr{Q}||$ if $(s,r) \in \mathscr{Q}$, and $g(P(\hat{S}(a),\hat{R}(b))) = 0$ otherwise. The set $\mathscr{Q}$ may be a singleton, $||\mathscr{Q}|| = 1$, in which case the anonymity is likely to be low, but it may just as well contain all possible sender-receiver pairs, $||\mathscr{Q}|| = (N-C)\cdot(N-C-1)$, which would correspond to perfect relationship anonymity. Note that $(a,b) \in \mathscr{Q}$ does not imply that (a,b) is the actual sender-receiver pair, not even when $||\mathscr{Q}|| = 1$.

Let us denote by H_{1+} the event that there is an attacker node on the path that the message traverses. If H_{1+} and $(s,r) \in \mathscr{Q}$ happen then the attacker drops the message with probability $1/||\mathscr{Q}||$, otherwise it does not drop the message. We can thus express the relationship anonymity under the MP method as

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

2) Bayesian inference method: Using the Bayesian Inference (BI) method, when the attacker intercepts a message (i.e., H_{1+} happens) it drops the message with the a-posteriori probability $P(\hat{S}(s),\hat{R}(r))$, i.e., $g(P(\hat{S}(s),\hat{R}(r))) = P(\hat{S}(s),\hat{R}(r))$. Unlike under the MP method, the attacker may drop a message even in (s,r) is not the most likely sender-receiver pair.

Using the above notation we can express the relationship anonymity under the BI method as

$$A_{BI}(s,r) = P(\hat{S}(s),\hat{R}(r) | H_{1+}, S(s), R(r)) \cdot P(H_{1+}|S(s),R(r)). \quad (2)$$

IV. Anonymity system descriptions

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

A. MCrowds system description

MCrowds is an anonymity network inspired by Crowds [10], which was proven to provide optimal sender anonymity [12]. 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

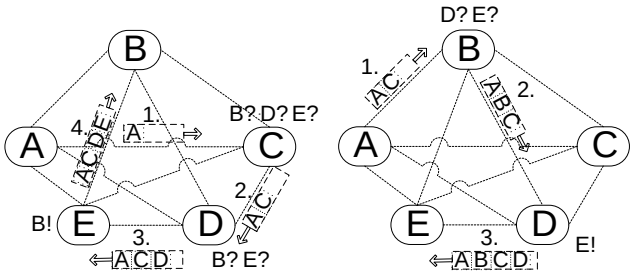

Fig. 1. A simple example of Minstrels with five nodes.

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.

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.

V. 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 for 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

$$P(C_M=c_M) = \left(\begin{array}{c} M-1 \\ c_M \end{array}\right) \frac{\prod_{k=2}^{M-c_M}(N-C-k)\prod_{k=0}^{c_M-1}(C-k)}{\prod_{k=2}^{M}(N-k)}. \tag{5}$$

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. Since a message is relayed to one of the $C-c_M$ attacker nodes with probability $\frac{C-c_M}{N-M}$ and the sender must relay the message initially, 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 $a \in \mathcal{N} \setminus (\mathcal{C} \cup \mathcal{M})$ and $b \in \mathcal{M} \setminus \mathcal{C}$. 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 $\mathcal{N} \setminus \mathcal{M}$ is equally likely to be used as relay, and the receiver is still one of the nodes in $\mathcal{M}$. 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

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

This expression is obtained using the same approach as in [10], but considering that the number of attacker nodes is $C - c_M$ and that the total number or relaying nodes is $N - M$. We omit the derivation for brevity.

Predecessor Node: Consider now that there is an attacker on the path. When the first attacker node on the path gets the message, the attacker knows the nodes in the set $\mathcal{M}$, 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 ($p = a$), and by $\bar{I}_a$ the event that the predecessor is not node a ($p \neq a$).

If H_1 happens and thus the attacker node is on position $i = 1$, then the sender of the message is the predecessor and the event I_a happens if a is the sender. Otherwise, if H_{2+} happens, i.e., the attacker is at position $i > 1$, we have to distinguish two cases. If $S(a)$ then any trusted node from the set $\mathcal{N} \setminus \mathcal{M}$ is equally likely to be the predecessor, and we have $P(I_a|H_{2+}, c_M, S(a), R(b)) = \frac{1}{N-C-M+c_M}$ for any $b \in \mathcal{N} \setminus \mathcal{C}$ and $b \neq a$. If $S(s)$ then I_a for $a \neq s$ can only happen if $a \notin \mathcal{M}$, but any $a \notin \mathcal{M}$ is equally likely to be the predecessor. The event $a \notin \mathcal{M}$ conditioned on $S(s)$ ($a \neq s$) happens with probability $P(a \notin \mathcal{M}|c_M, S(s), R(b)) = \frac{N-C-M-c_M-1}{N-C-2}$, for any $b \in \mathcal{N} \setminus \mathcal{C}$ and $b \notin \{s, a\}$. Thus, $P(I_a|H_{2+}, c_M, S(s), R(b)) = \frac{P(a \notin \mathcal{M}|c_M, S(s), R(b))}{N-C-M+c_M}$. Putting it all together, the event I_a conditioned on H_{1+} and $S(a)$ ($s \neq a$) happens with probability

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

and for $S(s)$ with probability

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

Anonymity with Attacker as Relay: 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. If node s is the predecessor (I_s) then the probability that the attacker assigns to node s being the sender of the message is

$$\begin{aligned} &P(\hat{S}(s)|I_s, H_{1+}, c_M, S(s), R(b)) = \\ &\frac{\sum_b P(I_s, H_{1+}, c_M|S(s), R(b)) \cdot T(S(s), R(b))}{\sum_{(a,b)} P(I_s, H_{1+}, c_M|S(a), R(b)) \cdot T(S(a), R(b))}, \end{aligned} \tag{10}$$

where $a \in \mathcal{N} \setminus (\mathcal{M} \cup \mathcal{C})$ and $b \in \mathcal{M} \setminus \mathcal{C}$. Recall that $T(S(a), R(b))$ is the attacker's a-priori belief of the traffic matrix, which it uses as the probability that node a sends a message to node b. 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.

Based on the above, the probability that a *relaying* attacker assigns to the actual sender of the message, given H_{1+} and $C_M = c_M$, is

$$\begin{aligned} &P(\hat{S}(s)|H_{1+}, c_M, S(s), R(b)) = \\ &P(\hat{S}(s)|I_s, H_{1+}, c_M, S(s), R(b)) \cdot P(I_s|H_{1+}, c_M, S(s), R(b)) + \\ &P(\hat{S}(s)|\bar{I}_s, H_{1+}, c_M, S(s), R(b)) \cdot P(\bar{I}_s|H_{1+}, c_M, S(s), R(b)). \end{aligned} \tag{11}$$

The probability the attacker assigns 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.

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 ($C_M > 0$). Nevertheless, the attacker does not gain more information about the actual sender-receiver pair (s,r) upon receiving the message as a member of the receiver set. We therefore do not have to consider this case separately.

Anonymity with no Attacker as Relay: Let us now consider the case when there is no attacker on the path. We 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 for $C_M > 0$ we have

$$\begin{aligned} &P(I_a|\overline{H}_{1+}, c_M, S(a), R(b)) = P(I_a|H_{2+}, c_M, S(a), R(b)), \\ &P(I_a|\overline{H}_{1+}, c_M, S(s), R(b)) = P(I_a|H_{2+}, c_M, S(s), R(b)), \quad a \neq s. \end{aligned}$$

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 similar to (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 $\mathscr{M}$ 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. Therefore, if I_s happens and the attacker assumes $\hat{S}(s)$ then it assigns probability $P(\hat{R}(r)|\hat{S}(s),I_s,\overline{H}_{1+},c_M,S(s),R(r)) = \frac{1}{N-C-1}$ to 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. Thus, if $\bar{I}_s$ happens and the attacker assumes $\hat{S}(s)$ then the probability assigned to the receiver is $P(\hat{R}(r)|\hat{S}(s),\bar{I}_s,\overline{H}_{1+},c_M,S(s),R(r)) = \frac{1}{N-C-2}$. 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

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

Tying it together: We are now ready to express the relationship anonymity $A_{BI}(s,r)$ under the *BI* method 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 $C_M = c_M > 0$,

$$\begin{aligned}A_{BI}(s,r) = &\sum_{c_M} P(\hat{S}(s),\hat{R}(r)|H_{1+},c_M,S(s),R(r)) \\ &\cdot P(H_{1+}|c_M,S(s),R(r))\cdot P(C_M = c_M) \\ &+ \sum_{c_M\neq 0} P(\hat{S}(s),\hat{R}(r)|\overline{H}_{1+},c_M,S(s),R(r)) \\ &\cdot P(\overline{H}_{1+}|c_M,S(s),R(r))\cdot P(C_M = c_M).\end{aligned} \tag{13}$$

In order to calculate the relationship anonymity $A_{MP}(s,r)$ 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., $(s,r)\in\mathscr{Q}$. This can be easily done for an arbitrary traffic matrix T given particular events, e.g., I_s and H_{1+}. In the special case when the attacker's a-priori belief is that the traffic matrix is homogeneous, all pairs $(a\rightarrow b)$ 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 sole most likely sender. Therefore, $(s,r)\in\mathscr{Q}$ only if I_s happens. If this happens, then every trusted node in the receiver set $\mathscr{M}$ is equally likely to be the receiver, thus $||\mathscr{Q}|| = M - c_M$.

C. Relationship Anonymity for 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 additional information that it could use to assign higher probability 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 $C_F = c_F$,

$$\begin{aligned}A_{BI}(s,r) = &\sum_{c_F}\sum_{v}\sum_{\omega_s}\sum_{\omega_r} \\ &P(\hat{S}(s),\hat{R}(r)|\omega_r,\omega_s,c_F,H_{1+},v,S(s),R(r)) &(14)\\ &\cdot P(\omega_r,\omega_s,c_F,H_{1+},v|S(s),R(r)), &(15)\end{aligned}$$

$$\begin{aligned}A_{MP}(s,r) = &\sum_{c_F}\sum_{v}\sum_{\omega_s}\sum_{\omega_r}\frac{P((s,r)\in\mathscr{Q}|\omega_r,\omega_s,c_F,H_{1+},v,S(s),R(r))}{||\mathscr{Q}||} &(16)\\ &\cdot P(\omega_r,\omega_s,c_F,H_{1+},v|S(s),R(r)). &(17)\end{aligned}$$

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) corresponds to the probability that the sender-receiver pair $(s,r)\in\mathscr{Q}$.

The key to calculate $A_{BI}(s,r)$ and $A_{MP}(s,r)$ is to calculate the probability that the attacker assigns to the sender-receiver pair (s,r) in (14), for which we have to rely on the information available to the attacker upon receiving a message. A message contains the information ($||\mathscr{V}|| = v$, $\omega_s\in\Omega_s$, $\omega_r\in\Omega_r$, and $C_F = c_F$), and based on these the attacker would compute the probability that (s,r) is the sender-receiver pair as

$$\begin{aligned}&P(\hat{S}(s),\hat{R}(r)|\omega_r,\omega_s,c_F,H_{1+},v) = \\ &\frac{P(\omega_r,\omega_s,v,c_F,H_{1+}|S(s),R(r))\cdot T(S(s),R(r))}{\sum_{(a,b)}P(\omega_r,\omega_s,v,c_F,H_{1+}|S(a),R(b))\cdot T(S(a),R(b))}\end{aligned} \tag{18}$$

where the summation in the denominator is over all possible non-attacker sender-receiver pairs $(a\rightarrow b)$. $T(S(a),R(b))$

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, $T(S(a),R(b)) = \frac{1}{(N-C)(N-C-1)}$ for all $(a \to b)$, and these probabilities cancel out each other in (18). In what follows we compute the probabilities in (18).

Number of Initialized Attacker Nodes: 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

$$P(H(v,c_F|F=f)) = \begin{cases} \frac{C}{N-1} & v=0, f=0 \\ P(C_F=0|F=f)\frac{N-C-1}{N-1}\frac{C}{N-v}\prod_{z=1}^{v-1}\frac{N-C-z}{N-z} & v\geq 1, f=0 \\ P(C_F=c_F|F=f)\frac{C-c_F}{N-v}\prod_{z=f}^{v-1}\frac{N-C+c_F-z}{N-z} & v\geq 1, f>0, \end{cases} \quad (19)$$

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 in Minstels, $c_F \in \{max(0,f-1-(N-2-C)),min(f-1,C)\}$. 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

$$P(C_F|F=f) = \begin{pmatrix} f-1 \\ c_F \end{pmatrix} \frac{\prod_{k=2}^{f-c_F}(N-C-k)\prod_{k=0}^{c_F-1}(C-k)}{\prod_{k=2}^{f}(N-k)}. \quad (20)$$

Visited nodes and the Predecessor: 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.

We already calculated the numerator of (18), 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 (18) 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.

D. Bounds for the 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

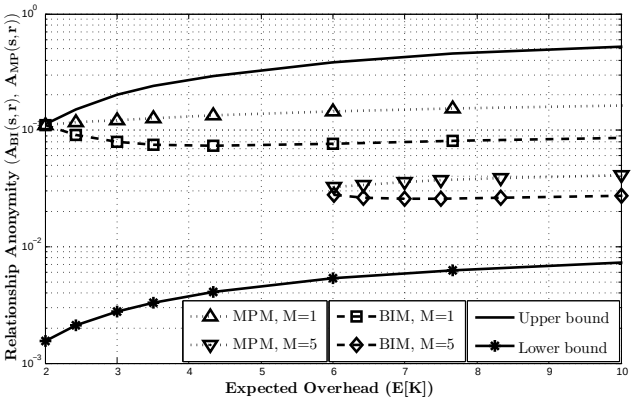

Fig. 2. Relationship anonymity vs. overhead for MCrowds, $N = 10$, $C = 1$

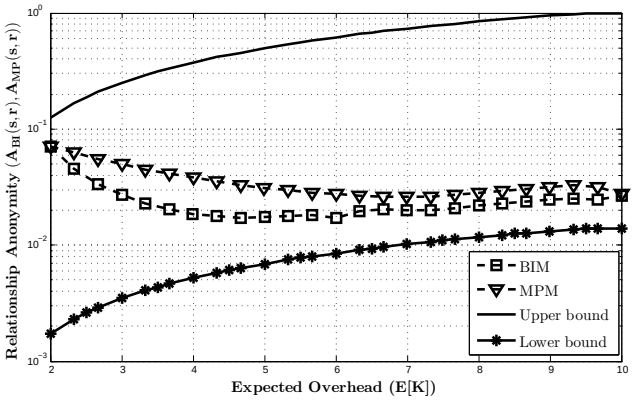

Fig. 3. Relationship anonymity vs. overhead for Minstrels, $N = 10$, $C = 1$

a message, it assumes that any trusted pair of nodes is equally likely to be the sender-receiver pair with probability $P(\hat{S}(s),\hat{R}(r)|H_{1+},S(s),R(r)) = \frac{1}{(N-C)(N-C-1)}$.

VI. Numerical Results

In the following, we first use the analytical results to investigate the relationship anonymity-overhead trade-off provided by MCrowds and by Minstrels. We then show simulation results that confirm the analytical results.

A. Relationship anonymity-overhead trade off

We use the analytical results developed in Section V to explore the trade-off between relationship anonymity and overhead for MCrowds and for Minstrels. For MCrowds we use a 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 ($A_{BI}(s,r)$) and the relationship anonymity under the MP method ($A_{MP}(s,r)$) 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 $A_{BI}(s,r)$ and $A_{MP}(s,r)$ 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), 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 ($M = 5$ 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

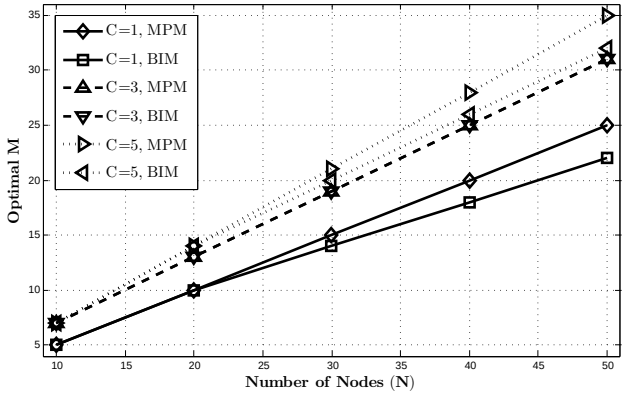

Fig. 4. Optimal receiver set size M vs. number of nodes for MCrowds

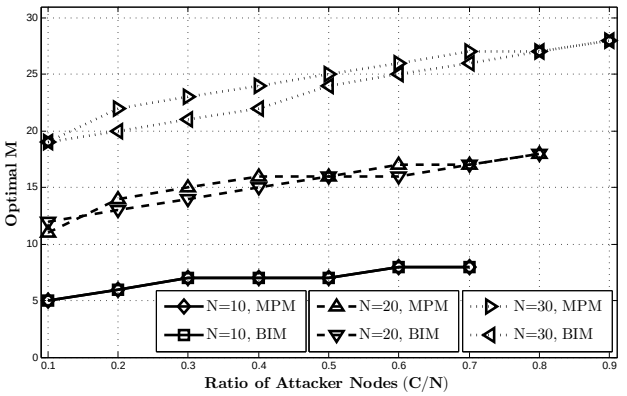

Fig. 5. Optimal receiver set size M vs. ratio of attacker nodes for MCrowds

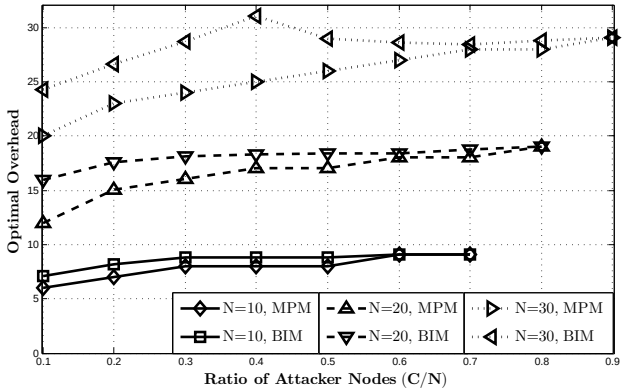

Fig. 6. Optimal overhead vs. ratio of attacker nodes for MCrowds

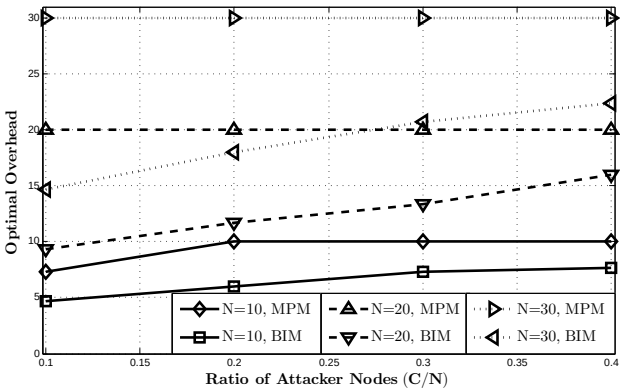

Fig. 7. Optimal overhead vs. ratio of attacker nodes for Minstrels

value.

Fig. 6 and Fig. 7 show the optimal overhead (where the probabilities $A_{MP}(s,r)$ or $A_{BI}(s,r)$ 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 $A_{MP}(s,r)$ and $A_{BI}(s,r)$ at the optimal overhead as a function of the ratio of attacker nodes $(\frac{C}{N})$. As the ratio of attacker nodes increases, the probabilities $A_{MP}(s,r)$ and $A_{BI}(s,r)$ 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 $A_{MP}(s,r)$ and $A_{BI}(s,r)$. 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 $A_{MP}(s,r)$ (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, $A_{MP}(s,r)$ 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

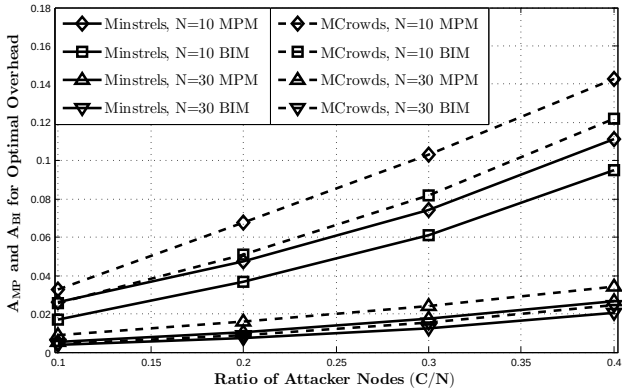

Fig. 8. Relationship anonymity for optimal overhead vs. ratio of attacker nodes

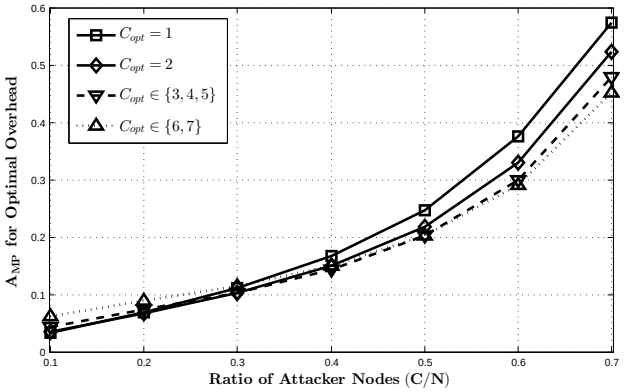

Fig. 9. $A_{MP}(s,r)$ for optimal overhead vs. ratio of attacker nodes for MCrowds

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.

B. Trade off between Relationship Anonymity and Sender-Receiver Anonymity

In the following, we explore the trade off between the relationship anonymity and the sender or receiver anonymity in order to justify our approach to consider the relationship anonymity instead of the sender and the receiver anonymity separately. We quantify the sender (receiver) anonymity similarly to the relationship anonymity: the probability that a message sent from s (sent to r) is dropped when s (r) is the targeted sender (receiver), i.e., the expected true positive rate.

Fig. 10 shows the trade-off between the sender or receiver anonymity and the relationship anonymity for a system with $N = 10$ nodes that uses MCrowds with $M \in \{1,3,5\}$ and $p_f \in (0.1, 0.9)$. The attacker is in control of one node ($C = 1$), and it uses the MP method assuming that $T(S(a),R(b))$ is uniform. To calculate the sender (receiver) anonymity, we used the analytical results developed in Section V-B while assuming that the probability assigned to the receiver (sender) equals to 1. Both sender and receiver anonymity increase with the relationship anonymity as a function of p_f. However, the best relationship anonymity is not achieved together with the best sender or receiver anonymity. The best relationship anonymity is achieved for $M = 5$, while the best sender anonymity and the best receiver anonymity are achieved for $M = 1$ and $M = 3$, respectively. Thus, the results show that it is better to consider the relationship anonymity instead of the sender and the receiver anonymity separately when optimizing an anonymity network to protect pair-wise communication.

C. Attack Efficiency

As a validation of the model, we finish with simulation results for a system with $N = 10$ nodes. The actual traffic matrix of the system is: every pair of nodes exchanges 50 messages per time unit. To anonymize the communication, we use MCrowds with $M = 1$. The attacker wants to drop messages sent from node n_1 to node n_4. The attacker is in control of node n_3 ($\mathscr{C} = \{n_3\}$, $C = 1$), and it does not have any a-priori knowledge of the system traffic matrix; it assumes $T(S(a),R(b))$ is uniform. Recall that for MCrowds $M = 1$, the attacker is always certain about the receiver r but it is uncertain about the actual sender of the messages it observes. Therefore, it may drop some messages that are not actually sent from n_1, but it does not drop messages for which the receiver is not n_4.

Fig. 11 shows the fraction of dropped messages sent from node n_1 to node n_4 (true-positive) and the average of the fractions of dropped messages sent from nodes in $\mathscr{N} \setminus (\{n_1,n_3,n_4\})$ to node n_4 (false-positive) as a function of the relaying probability p_f. The results are the averages of 1000 simulations. The 95% confidence intervals are within 3.25% of the results, and they are omitted in the figure since they would be hardly visible. The scenario is the same as that considered in Fig. 2, and the results show a perfect match.

VII. Conclusions

In this paper we considered the problem of mitigating denial of service attacks by providing relationship anonymity 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 attack 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

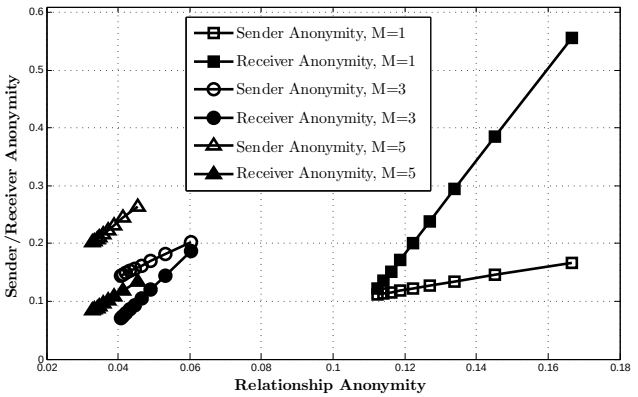

Fig. 10. Sender and receiver anonymity vs. relationship anonymity under the MP method for MCrowds.

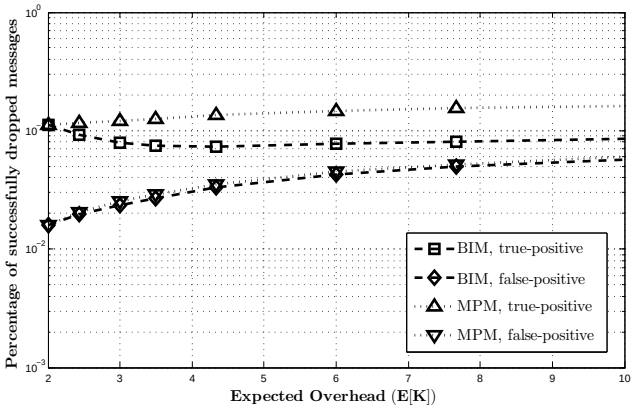

Fig. 11. Fraction of dropped messages sent from n_1 to n_4 (true-positive) and between other pairs (average of false positives) where the receiver is n_4 (false-positive) vs. relaying probability p_f. The targeted s-r pair is (n_1, n_4), $N = 10$, $C = 1$, and MCrowds is used with $M = 1$.

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," in *Proc. of IEEE*, vol. 93, no. 6, 2005, pp. 1152–1177.

[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] R. J. Turk, "Cyber incidents involving control systems," Idaho National Laboratory, Tech. Rep., 2005.

[4] S. Amin, A. A. Cárdenas, and S. S. Sastry, "Safe and secure networked control systems under denial-of-service attacks," in *Proc. of the 12th International Conference on Hybrid Systems: Computation and Control*. Springer-Verlag, 2009, pp. 31–45.

[5] K. Stouffer, J. Falco, and K. Scarfone, "Guide to industrial control systems (ICS) security," NIST SP 800-82, Tech. Rep., 2011.

[6] J. Mirkovic and P. Reiher, "A taxonomy of DDoS attack and DDoS defense mechanisms," *ACM SIGCOMM Computer Communications Review*, vol. 34, no. 2, pp. 39–53, April 2004.

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

[8] A. Pfitzmann and M. Köhntopp, "Anonymity, unobservability, and pseudonymity - a proposal for terminology," in *Designing Privacy Enhancing Technologies*. Springer Berlin Heidelberg, 2001, vol. 2009, pp. 1–9.

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

[10] M. Reiter and A. Rubin, "Crowds: Anonymity for web transactions," *ACM Trans. Inform. Syst. Secur.*, vol. 1, no. 1, pp. 66–92, 1998.

[11] V. Shmatikov and M. H. Wang, "Measuring relationship anonymity in mix networks," in *Proc. of Workshop on Privacy in the Electronic Society (WPES)*, 2006.

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

[13] 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.

[14] 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.

[15] C. Troncoso, B. Gierlichs, B. Preneel, and I. Verbauwhede, "Perfect matching disclosure attacks," in *Proc. of Privacy Enhancing Technologies Symposium (PETS)*, 2008.

[16] 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.

[17] C. Troncoso and G. Danezis, "The bayesian traffic analysis of mix networks," in *Proc. of Conference on Computer and Communications Security (CCS)*, 2009.

[18] 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.

[19] M. Long, C.-H. Wu, J. Hung, and J. Irwin, "Mitigating performance degradation of network-based control systems under denial of service attacks," in *In Proc. of IEEE Industrial Electronics Society Conference IECON*, vol. 3, November 2004, pp. 2339–2342.

[20] M. Long, C.-H. Wu, and J. Y. Hung, "Denial of Service Attacks on Network-Based Control Systems: Impact and Mitigation," *IEEE Transactions on Industrial Informatics*, vol. 1, no. 2, pp. 85–96, May 2005.

[21] H. Foroush and S. Martinez, "On event-triggered control of linear systems under periodic denial-of-service jamming attacks," in *In Proc. of IEEE 51st Annual Conference on Decision and Control (CDC)*, December 2012, pp. 2551–2556.

[22] L. Shichao, X. Liu, and A. El Saddik, "Denial-of-service (DoS) attacks on load frequency control in smart grids," in *In Proc. of IEEE PES Innovative Smart Grid Technologies (ISGT)*, February 2013, pp. 1–6.

Ognjen Vuković received his M.Sc. degree in Telecommunications, System engineering and Radio Communications in 2010 from the Faculty of Electrical Engineering, University of Belgrade. He received his Licentiate Degree in Electrical Engineering in 2013 from KTH The Royal Institute of Technology, Stockholm, Sweden, where he is currently pursuing his PhD in the Laboratory of Communication Networks.

His research interests include cyber-physical security of power systems, 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 V-C 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.

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

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 $P(s\in$

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

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

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

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

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

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

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

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