SECURITY AND COMMUNICATION NETWORKS
Security Comm. Networks 0000; **00**:1–17
DOI: 10.1002/sec

SPECIAL ISSUE PAPER

On Collaborative Anonymous Communications in Lossy Networks

David Rebollo-Monedero, Jordi Forné*, Esteve Pallarès, Javier Parra-Arnau, Carolina Tripp, Luis Urquiza and Mónica Aguilar

Department of Telematics Engineering, Universitat Politècnica de Catalunya (UPC), C. Jordi Girona 1-3, E-08034 Barcelona, Spain

ABSTRACT

Message encryption does not prevent eavesdroppers from unveiling who is communicating with whom, when, or how frequently, a privacy risk wireless networks are particularly vulnerable to. The Crowds protocol, a well-established anonymous-communication system, capitalizes on user collaboration to enforce sender anonymity.
This works formulates a mathematical model of a Crowds-like protocol for anonymous communication in a lossy network, establishes quantifiable metrics of anonymity and quality of service (QoS), and theoretically characterizes the trade-off between them. The anonymity metric chosen follows the principle of measuring privacy as an attacker's estimation error. By introducing losses, we extend the applicability of the protocol beyond its original proposal. We quantify the intuition that anonymity comes at the expense of both delay and end-to-end losses. Aside from introducing losses in our model, another main difference with respect to the traditional Crowds is the focus on networks with stringent QoS requirements, for best-effort anonymity, and the consequent elimination of the initial forwarding step.
Beyond the mathematical solution, we illustrate a systematic methodology in our analysis of the protocol. This methodology includes a series of formal steps, from the establishment of quantifiable metrics all the way to the theoretical study of the privacy-QoS trade-off.

KEYWORDS

anonymous communications; quality of service; Crowds; lossy, wireless and vehicular networks

***Correspondence**
Jordi Forné, Universitat Politècnica de Catalunya (UPC), Campus Nord, Módulo C3, C. Jordi Girona 1-3, E-08034 Barcelona, Spain
E-mail: jforne@entel.upc.edu

Received . . .

1. INTRODUCTION

Recent times have witnessed the emergence of a wide variety of wireless and mobile technologies, such as WiFi (IEEE 802.11a/b/g/n), Bluetooth, Wireless USB, 3G and 4G mobile telephony (HSDPA+, LTE), which enable and respond to an increasingly clear trend towards a completely interconnected world, an Internet of Things where individuals and objects are seamlessly integrated into a global information network, where physical entities gradually acquire a virtual counterpart, and where vast quantities of data are generated and processed directly by users and automatically by software on their behalf. Widespread forms of online access, such as Web browsing, social networking and virtual shopping, are becoming a common occurrence on mobile devices, which in addition allow new forms of online services, such as those based on GPS location. And yet free or inexpensive mobile ad hoc (MANETs) and vehicular networks (VANETs), along with the presence of hot spots on the streets, are making Internet more accessible than it ever was. These and many other information and communication technologies (ICTs) make it possible to envision smart cities where people share experiences everywhere, and actively contribute to the betterment of their environment. Thanks to this ubiquitous connectivity, citizens may report incidences such as traffic jams or violations and poor traffic signaling, or emergency situations such as car accidents and crimes.

But many of these unquestionably useful technologies come at a price, as the availability of potentially sensitive information about all sorts of individual preferences and behavior across a diversity of services, translates into numerous, increasingly prominent privacy risks. Further, in applications such as reporting of traffic violations, users may strongly prefer remaining anonymous in order to avoid personal repercussions. The implementation of

Prepared using secauth.cls [Version: 2010/06/28 v2.00]

mechanisms to protect user privacy, key to a sustainable development of such technologies, cannot disregard any impact on service quality due to any form of traffic or processing overhead, a particularly delicate issue in wireless networks. Of particular importance for user acceptance is the ability to protect the anonymity of the sender in applications involving both two-way communication and reporting, including any the examples aforementioned.

Message encryption is notably insufficient to mitigate all possible kinds of privacy risks derived from network eavesdropping. Concealing the content of data packets hinders attackers in their efforts to learn the information exchanged, but does not prevent those attackers from unveiling who is communicating with whom, when, or how frequently. *Anonymous-communication systems* emerged to mitigate the serious privacy risk posed by illicit traffic analysis based on routing, size, timing and frequency patterns of messages between identified senders and receivers, beyond the mere protection of the confidentiality of message content via encryption.

The fundamental strategies to counter traffic analysis based on message routing, size and timing commonly resort to a network architecture involving trusted nodes or user collaboration, relying on rerouting, header encryption, message padding and splitting, dummy traffic insertion, and message delay and reordering, with varying degrees of sophistication. A constant in all of these strategies is that any anonymity gain comes at a price in processing and communication overhead, often causing a measurable degradation of the *quality of service* (QoS) offered by the network. Users and system designers are thus faced with a dilemma in the form of a *trade-off* between the contrasting aspects of privacy and usability, of inescapable relevance in any practical, modern communication system.

Of great importance in this context are wireless, mobile ad hoc, and vehicular networks. While their rapid expansion obeys to unquestionable advantages in innumerable fields of application, these type of networks are especially vulnerable to the traffic analysis risks aforementioned. Further, these networks are subject to packet losses, mainly owing to signal attenuation and message collision in the wireless medium. Consequently, any traffic overhead incurred by the anonymous-communication mechanisms enumerated is likely to translate into message losses and delays.

An archetypical example of anonymous-communication system is the *Crowds protocol* [27], which builds upon the principle of user collaboration with a limited degree of trust. Crowds is particularly helpful to minimize requirements for infrastructure and trusted intermediaries such as pseudonymizers, or to simply provide an additional layer of anonymity. In this protocol, a group of users will collaborate to submit their messages to a specified recipient, from whose standpoint they wish to remain completely anonymous. In simple terms, the protocol works as follows. When sending a message, a user flips a biased coin to decide whether to submit it directly to the recipient, or to send it to another user, who will then repeat the randomized decision. In the end, anonymity comes at the expense of traffic overhead and delay.

It is our contention that if we wish to propose usable privacy solutions, we must contemplate both the gain in anonymity offered and the cost in QoS demanded. A systematic approach consists in first establishing quantifiable measures not only of QoS, but also of anonymity, to then assess, compare, improve and ultimately optimize anonymous-communication systems, in terms of the inherent trade-off discussed. In certain cases, including but not limited to ad hoc networks, the specific requirements of the network architecture may lead to prefer solutions based on user collaboration in lieu of those involving infrastructure with trusted intermediaries. The object of this work is to apply this systematic approach to the theoretical analysis of a specific yet representative privacy application.

1.1. Contribution and Organization

Because of its paramount importance, the trade-off between anonymity and QoS in anonymous-communication systems has been frequently addressed when proposing and assessing solutions, either through theoretical analysis or experimental evaluation [17, 35, 24, 20, 30]. However, to the best of our knowledge, this is the first theoretical analysis of the anonymity-QoS trade-off for Crowds in the presence of losses.

More precisely, in this work, we

- formulate a mathematical model of a Crowds-like protocol for anonymous communication in a lossy network,
- establish appropriate metrics of anonymity and QoS, and
- theoretically characterize the trade-off between them.

A further dimension of our contribution lies beyond the mathematical solution to the specific problem formulated. Namely, the general, systematic methodology applied to the analysis of the protocol as a privacy-enhancing mechanism, from the establishment of quantifiable metrics all the way to the theoretical study of the trade-off. This paper constitutes an illustration of said methodology.

Two important differences in our contribution with respect to the original Crowds protocol must be stressed.

- First, the possibility of losses, and more generally the focus of our work on the compromise posed by the price of anonymity in the form of violation of stringent QoS requirements.
- Secondly and concordantly, we do not introduce a mandatory initial forwarding step. One reason is that such initial step would double the minimum possible message forward count from 1 to 2, imposing a price on average delay and loss probability which, in the context of the intended

Security Comm. Networks 0000; **00**:1–17 © 0000 John Wiley & Sons, Ltd.
DOI: 10.1002/sec
Prepared using secauth.cls

applicability of our work, we deem more than significant. Another reason is that, while the present study attributes any anonymity attacks to a common, untrusted receiver, the benefit in anonymity of said initial forwarding would not be as pronounced in a more general setting where collaborating users were not fully trusted.

On a more practical note, recall that in *single-hop* wireless networks, all nodes are within transmission range and messages are thus sent directly, as opposed to *multihop networks*, where in order to attain the desired coverage and throughput, messages may be relayed a number of times before they reach their intended destination. This preliminary contribution on the subject of collaborative anonymous communications in lossy networks is restricted to the former case, single-hop networks, which already offers a rich interplay of issues that translate into a sufficiently complex mathematical analysis, but should constitute a first step towards the understanding of the more intricate case of multihop networks.

Finally, we also exclude from the necessarily limited scope of this preliminary contribution on the subject, the issue of analysis of the entire forwarding path based on message timing or length. Details on privacy and security assumptions and their justification are provided later. These and other restrictions in our study, along with applicability considerations, are the object of discussion of an entire section prior to our theoretical analysis. Far from presenting a complete analysis of a comprehensive anonymous-communication solution to all possible forms of attacks based on traffic analysis, along with detailed configuration guidelines and implementation details, the current work addresses a partial albeit sufficiently rich aspect of Crowds in lossy networks.

From a mathematical perspective, it must be pointed out that, despite the apparent simplicity of the Markov chain modeling the main problem of the paper, the proof corresponding to its full-fledged version with losses but without self-forwarding, requires an intricate deconstruction into a series of preliminarily lemmas. These lemmas, specifically developed here for the problem at hand, should greatly facilitate its understanding. In particular, the proof of the second theorem resorts to two consecutive, nontrivial transformations into a simpler version; it is the two transformations themselves, not the simpler, reduced version, which draw upon the lemmas.

The rest of the paper is organized as follows. After a quick note on the main causes of packet losses in wireless networks, Sec. 2 succinctly places the Crowds protocol in the context of the state of the art on anonymous-communication systems and related anonymity metrics. Sec. 4 describes our main assumptions, formalizes the problem investigated in this paper, and presents and discusses our theoretical analysis. The theorems laying the foundation of our disquisition are proven in Sec. 5. Our main results are validated and illustrated by means of a numerical example in Sec. 6, and briefly summarized in the conclusions of Sec. 7.

2. BACKGROUND

Before delving into the state of the art on anonymous-communication systems and related anonymity metrics, we briefly enumerate the causes of packet losses in wireless networks.

2.1. Packet Losses in Wireless Networks

Roughly speaking, the causes of packet losses in a wireless network are as follows:

- *Saturated link*. Whenever the available bandwidth to a node in a link becomes (nearly) nonexistent, that node never gains access to the radio medium or already emits frames at a rate that saturates the medium.
- *Collision*. If the medium is busy on the receiver's side, frames systematically experience collisions and communication cannot succeed. The likelihood of collision increases with packet size and frequency of attempts to access the medium.
- *Fading and attenuation*. Buildings and a variety of structures, specially in urban scenarios, may attenuate or even impede signal propagation, due not only to shadowing from objects blocking the line of sight, but also to multipath destructive interference.
- *Link breakage*. Particularly in vehicular ad hoc networks, the moving speed of the nodes can be high, thereby quickly altering topology and link effective lifetime.

2.2. State of the Art on Anonymous-Communication Systems

The concept and purpose of anonymous-communication systems have already been introduced in Sec. 1. Next, we offer a glimpse into the extensive literature on the subject, while placing the Crowds protocol, also defined in that section, in the context of this type of systems. Before we proceed, however, we must stress that the focus of our overview captures only a fraction of a wide spectrum of existing forms of user privacy risks and mechanisms in communication systems [14], beyond those directly related to traffic analysis. A simple yet notable way of enforcing message anonymity employs a *trusted third party* (TTP) acting as a pseudonymizer between user and information service provider, effectively hiding the identity of the user. An appealing twist that does not require that the TTP be online is that of digital credentials [8, 3]. Needless to say, many existing alternative privacy-enhancing technologies, far from being mutually exclusive, may in fact be combined synergically.

Timing analysis [19, 21, 2], essentially traffic analysis that infers the correspondence between incoming and outgoing messages for a given node on the basis of the arrival and departure times, has already been motivated in the introduction. The first anonymous-communication system attempting to also counter timing analysis was the *Chaum mix* [7], essentially a trusted node that delays and reorders messages with the purpose of providing unlinkability, as defined by [23], between incoming and outgoing messages.

A wide range of sophisticated variations on the original mix shortly ensued [29], with the same purpose. One of the most relevant varieties is a family of mixes known as threshold pool mixes. The leading idea is for the mix to collect a number of incoming messages, store them in the internal memory of the mix, and output some of them when the number of messages kept in its memory reaches a certain threshold. In order to reduce the correlation between outgoing and incoming messages, the mix modifies the flow of messages by resorting to two strategies, namely delay and reordering.

Naturally, chains of mixes can be implemented to distribute trust, as Chaum already suggested in his original work [7] but, certainly, delaying messages may seriously affect the usability of these systems. Nevertheless, higher delays provide users with a higher degree of message unlinkability. In short, mix systems pose an inherent trade-off between anonymity and delay, in addition to the overheads derived from any encryption or padding.

Alternative low-latency anonymous-communication systems appeared later to provide routing anonymity on the Internet to a certain extent, without the price of message delay. Onion routing, and subsequent improvements termed *the* second-generation version of *onion routing* (Tor) [15], consist in networks of trusted routing nodes which, unlike mixes, do not insert artificial delays. In a nutshell, a user wishing to send a message chooses a chain of onion routers, and encrypts the message in a multilayered manner; hence the onion metaphor. This multilayered encryption is such that each router, after decrypting —peeling off a layer of encryption—, retrieves the address in plaintext of the node immediately subsequent in the path, along with an encrypted portion meant for said next node, all the way to the final recipient. We would like to stress that, as these systems boil down to anonymously relaying messages without introducing delays, they are susceptible to traffic analysis based on timing comparisons.

Yet another type of anonymous-communication systems builds upon the principle of user collaboration with a limited degree of trust. We already mentioned in the introductory section the *Crowds protocol* [27], according to which a group of users will collaborate to submit their messages to a specified recipient. As we explained, when sending a message, a user flips a biased coin to decide whether to submit it directly to the recipient, or to send it to another user, who will then repeat the same randomized decision. In fact, Crowds provides anonymity from the perspective of not only the final recipient, but also the intermediate nodes. Therefore, trust assumptions are essentially limited to fulfillment of the protocol. The original proposal suggests adding an initial forwarding step, which substantially increases the uncertainty of the first sender from the point of view of the final receiver, at the cost of an additional hop. In addition, we remarked that, in Crowds as in most anonymous-communication systems, anonymity comes at the expense of traffic overhead and delay. Just as the rest of low-latency systems described, Crowds only addresses part of the risks derived from traffic analysis, excluding attacks based on timing.

Anonymous-communication systems in general are vulnerable to a number of attacks based on traffic analysis. When striving to reveal the recipient of a communication for a given sender, attackers may perpetrate what is known as *disclosure attack* [18], based on the intersection of successive sets of possible candidate receivers for a given sender throughout extended periods of time. A refinement of this attack, the *statistical disclosure attack* [12], considers not only possibilities, but also probabilities.

Apart from these attacks, considerable research effort has been devoted to investigate more specific weaknesses of the Crowds protocol itself. Possibly the best-known attack is the *predecessor attack*, which was suggested in the original paper [27]. Such an attack contemplates that the most likely initiator of a communication is the immediate node preceding the first attacker. A generalized version of said attack assumes that an originator node will communicate with a certain destination more than once. In this more general attack, malicious collaborators can track communication flows over a number of rounds; at every round, the communication path between the originator and destination nodes is reconfigured.

A closely related work is [33]. Here, it is assumed that the attackers are able to track a given session, that is, a communication between a sender and a receiver. The cited work conducts an analysis of complexity in terms of the number of rounds, size of the crowd and number of malicious collaborators, with the aim of ascertaining the originator with high probability. However, in the special case when several originators establish a communication with a single destination, the authors find that the attackers cannot link specific data streams to each originator. This is unless there exists information in at least one packet per round that distinguishes the sessions from each other.

Another study regarding this same attack [22] determines how many rounds the attackers need to calculate, with arbitrary precision, the frequency with which a certain user communicates with the receiver. For that purpose, the authors use a Poisson distribution that enables them to model the flow messages to a destination. To counter this attack, they propose that honest users passively monitor the network to estimate both the sending rate and the peers of other users, in order to adapt their behavior to not be detected by an attacker.

Security Comm. Networks 0000; **00**:1–17 © 0000 John Wiley & Sons, Ltd.
DOI: 10.1002/sec
Prepared using secauth.cls

Also in relation to the subject of attacks against Crowds, [31] defines participant payload as the amount of messages sent or forwarded by a given node. Based on this concept, the cited work presents a study of the participant payload in Crowds as a function of the length of the forwarding paths. The study concludes that the expected participant payload is, on the one hand, equal to the expected length of forwarding paths, and on the other it is independent of the size of the crowd. Furthermore, the authors perform tests to get the relationship between the number of rounds needed for the predecessor attack to succeed, and the maximum length of forwarding paths. The authors did not find a remarkable dependence.

With regard to recent implementations of Crowds for providing sender anonymity, an example over Bluetooth with Java technology is described in [32]. The system developed is one-way only, on account of the fact that the mobility of nodes cannot guarantee a valid two-way path. In [34], Crowds is implemented over a wireless network, using the NTRUEncrypt public-key cryptosystem. More specifically, the authors propose a scheme with lower latency and CPU consumption, more suited to wireless networks, which only performs one decryption operation per path.

VIPER [6] is a modification of Crowds for vehicle-to-infrastructure communications, in which vehicles forward messages directed to a common infrastructure access point. Messages are sent in batches in predetermined time slots to counter timing attacks. The efficiency of VIPER is measured in terms of buffer occupancy and delivery time.

Lastly, we would like to remark that hybrid privacy protocols leverage not only user collaboration, but also query forgery, such as the private location-based information retrieval protocol via user collaboration in [25].

2.3. Related Anonymity Metrics

We argued in the introductory section that quantifiable measures of privacy and usability are undoubtedly essential to the assessment, comparison, improvement and optimization of any privacy-enhancing technologies. In the special case of anonymous-communication systems, the knowledge of the privacy attacker may be modeled by a probability distribution on the possible senders of a given message.

Certainly, one could measure the degree of anonymity attained by the mere cardinality of the set of candidate senders, that is, the size of the *anonymity set* [9]. The logarithm of such cardinality is in fact called *Hartley's entropy*. Loosely speaking, Hartley's entropy is a *possibilistic* metric, in the sense that it disregards the likelihood associated to each candidate.

Recall that *Shannon's entropy* is a measure of the uncertainty of a random event, associated with a probability distribution across the set of possible outcomes. Informally, Shannon's entropy is a good approximation to the minimum of the average number of binary questions regarding the nature of possible outcomes of an event, to determine which one in fact has come to pass, intelligently exploiting their known probabilities [11]. Its particular significance and wide application in the fields of information theory, statistics and engineering is unquestionable. Inspired by the interpretation of Shannon's entropy as the effective uncertainty within a set endowed with a probability distribution, [28, 13] proposed it as a measure of anonymity.

Recall also that *maximum a posteriori estimation* (MAP) is that in which the estimate is the most likely outcome, thereby minimizing the probability of estimation error in a finite set of candidates. In [26] a number of privacy metrics are studied under a unifying conceptual perspective, namely that of an attacker's estimation error in ascertaining the outcome of a random event, or effort in removing any residual uncertainty. The cited work includes, in addition to the two entropies aforementioned, *min-entropy* as a measure equivalent to the probability of error in MAP.

Because both Shannon's entropy and min-entropy, unlike Hartley's, take into consideration the probability distribution, thereby exploiting its potential skewness, they constitute *probabilistic* metrics. All three belong to the family of *Rényi entropies*, interpreted under the perspective of privacy measures in [26]. Additional surveys on information-theoretic quantities as privacy measures and novel proposals include [1]. Incidentally, [26] illustrates some of the entropies discussed with a vastly simplified example of Crowds, albeit considering neither losses nor QoS metrics.

The trade-off between anonymity and QoS has been frequently addressed in the literature. In [17], the authors illustrate the trade-off between anonymity and QoS for solutions implementing location privacy in wireless networks, and propose a new technique named silent cascade to enhance a user' location privacy without QoS degradation. Anonymity is measured as the Shannon and Hartley entropies of a mix network, while QoS is measured as the share of time a user spends on location privacy protection and as the delay introduced by the mix network.

The Scalar Anonymity System (SAS) is proposed in [35] in order to provide a trade-off between anonymity and cost for different users with different requirements. In SAS, by selecting the level of anonymity, a user obtains the corresponding anonymity and QoS and also sustains the corresponding load of traffic rerouting for other users. Anonymity is studied in terms of the predecessor attack while QoS is measured by the length of the rerouting path.

The QoS of Tor is systematically analyzed in [24]. The TCP throughput is used as the QoS metric and extensive experiments on the real-world Tor network are presented.

The impact of using established standard anonymity mechanisms on selected Quality of Service (QoS) parameters for Web services in real networks is evaluated in [20]. QoS is measured as the response time, consisting

of the network latency for the message transport and the service's execution time on the provider side. The authors set up a measurement infrastructure and evaluate the response time of different anonymity systems, including Tor, I2P and JonDo (free and commercial).

Finally, [30] addresses the need for applications such as VoIP to provide anonymity to clients while maintaining low latency to satisfy quality of service (QoS) requirements. They describe different triangulation based timing attacks and show that even when a small fraction of the network is malicious, the adversary can infer the source (caller) with reasonably high probability. The QoS property of an on-demand route set up protocol can be characterized by route latency and route set up latency

3. PRIVACY, SECURITY AND QOS REQUIREMENTS, APPLICABILITY AND IMPLEMENTATION

We would like to preface the mathematical analysis of the next section with more practical considerations, specifically, with a discussion of anonymity, security and QoS requirements, assumptions, limitations, implementation and configuration choices concerning our study, with the purpose of delimiting the most immediate real-world applicability of our work.

3.1. Requirements and Applicability

The introductory section already motivated the interest of this proposal, pertaining to the subject of anonymous-communication systems through user collaboration, which encompasses all information systems built on computer networks in which the disclosure of the identity of the sender of a message, by means of traffic analysis, represents a privacy risk. This includes both anonymous querying of an untrusted information provider and the delivery of the corresponding reply, and mere reporting or one-way communication. Examples may be found in the contexts of Web browsing, location-based and general recommendation systems, online social networks, online shopping, reporting of traffic conditions or violations, posting of reviews or opinions, and detailed power consumption in a household for smart grid optimization, to name a few application scenarios of key relevance in a future, completely interconnected world. Conceivably, messages may be sent either by people, or automatically by devices on their behalf, although the term user may informally be employed throughout the text for senders, receivers and intermediate, collaborating entities.

Our overview of the literature on anonymous-communication systems in Sec. 2.2 described a number of privacy attacks. Far from presenting a comprehensive, definitive solution to all forms of privacy and security risks that may arise in any given type of network with packet losses, the current study focuses solely on anonymity attacks perpetrated by the final, intended recipient of the message. These attacks are of the utmost relevance in behavioral profiling inferred from statistically matching the contents of said queries or reports with user identities, or from the observation of who —or more generally, which entity— is communicating with whom, when, or how frequently. Table I below gathers up in a conceptual manner the fundamental elements of the adversarial model assumed in this work, details of which are provided mainly in the current subsection and in Sec. 4.1.

We may add, to the probabilistic forwarding strategy in the Crowds protocol, message encryption to reinforce the confidentiality of messages in either direction, against collaborating users and external observers. Beyond those preliminary security measures, in the necessarily limited scope of this contribution on the subject we shall assume that the collaborating nodes in the network properly follow the forwarding protocol, thus disregarding denial-of-service attacks, and they will not be viewed as attackers against the anonymity of the messages, be it individually or in collusion with other forwarders or with their final recipient.

We stressed in Sec. 1.1 that one of the main differences in this work with respect to the classical formulation of the Crowds protocol is the incorporation of message losses into the theoretical model. This enables us to extend the applicability of Crowds, as an anonymous-communication protocol, which capitalizes on user collaboration to reduce infrastructure requirements, more realistically to wireless, mobile ad hoc, and vehicular networks.

Another notable difference already pointed out is the elimination of the initial, mandatory forwarding step in the original proposal of Crowds. The purpose of this initial step is to substantially increase the anonymity of messages from the perspective of their final, intended receiver. We justified its suppression in terms of a focus shift towards QoS-sensitive applications, say voice-based or emergency-related, as the minimum message forward count is halved, which translates into a significant reduction in end-to-end delay, message losses and traffic overhead. Despite our focus on the receiver as the potential anonymity attacker, we also noted that the benefit in anonymity of said initial forwarding would not be as pronounced in a more general setting where collaborating users were not fully trusted.

It was also anticipated in the introductory section that our theoretical model is restricted to single-hop networks, mainly because these networks already offer a rich interplay of issues that translate into a sufficiently complex theoretical analysis. Due to the significance of multihop networks, we should hasten to stress that our mathematical analysis on the modified Crowds protocol with message losses, with emphasis on the anonymity-QoS trade-off, may very well lay part of the fundamental principles to approach the theoretical study of the more intricate case of multihop networks in future research.

In wireless networks, particularly prone to eavesdropping, and especially in the single-hop case, the final receiver of a message may attempt to unveil the identity of

Security Comm. Networks 0000; **00**:1–17 © 0000 John Wiley & Sons, Ltd.
DOI: 10.1002/sec
Prepared using secauth.cls

Adversarial Model Highlights	
Who is the privacy attacker?	The scope of this work is limited to anonymity attacks perpetrated by a common receiver a group of collaborating users communicate with. Notable examples of such receivers are untrusted information providers and recipients of anonymous reports.
What is the attacker's goal?	The immediate goal is to identify the identity of the sender of a message. Ultimate purposes include profiling of user interests and behavior inferred from statistically matching the contents of queries with sender identities, and violation of anonymity in sensitive reporting.
What are the attacker's capabilities?	The receiver is assumed to know the specifics of the anonymous-communication, Crowds-like protocol employed by the users. Additionally, the receiver observes the identity of the last sender of a message in an incoming forwarding path. From all this information, the receiver may estimate the most likely identity of the sender of a message, although with limited certainty.

Table I. Main conceptual highlights of the adversarial model assumed in this work. Additional details appear in Secs. 3.1 and 4.1.

its original sender, and inferring the entire forwarding path, by examining the timing and length of packets through intermediate nodes, provided that they are within reception range. In our review of the literature in anonymous-communication systems we saw that mixes resort to the introduction of artificial delays and padding to counter this form of traffic analysis, strategies that could very well be implemented by trusted collaborating nodes. Alternatively, one may consider application scenarios where collaborating users have direct visibility with each other, say within a Bluetooth network or within a single vehicular network cell, but a common information provider remains accessible only through a separate network, say a GSM or a UMTS cellular network, or through an access point wired to the Internet. In any case, as we already stated, our preliminary theoretical model will exclude the form of traffic analysis described.

3.2. Metrics, Implementation and Configuration Choices

Clearly, the choice of specific, quantifiable metrics of anonymity and QoS, necessary to systematically assess, compare and optimize usable privacy mechanisms, should reflect the particular requirements of the underlying information systems and the privacy preferences and concerns of users and system designers.

The establishment of QoS metrics such as average delay, jitter, probability of message loss and number of hops are a necessary step in order to compare and improve routing protocols [10, 4]. Average delay is a wildly popular measure of QoS which reflects the intent of privacy and general system designers to tune performance according to the principle of average-case optimization, and enjoys the advantages of simplicity and mathematical tractability. Averages may be replaced by medians, a more complex quantity, for increased robustness against statistical outliers. In time-sensitive applications where significantly delayed packets may have to be discarded or their value is severely diminished, such as those involving voice or reporting of emergencies, examples of suitable QoS metrics comprise high delay percentiles —interpretable as robust maxima—, and the probability that a delay exceeds a given threshold, to be established in accordance with the application at hand. These more pessimistic metrics adhere to the principle of worst-case minimization, which may indeed yield less extreme delays, but cannot possibly improve over the average values resulting from average-case optimization. The probability of end-to-loss of a message is also a traditional measure of QoS, commonly accompanying average delay in order to offer a more informative picture. A simple measure combining the effects of message delays and losses is the aforementioned probability that a given time threshold is exceeded.

In our mathematical analysis we shall measure QoS jointly as end-to-end loss probability and as the average number of times a message is forwarded due to the modified Crowds protocol, and argue that the latter is an indirect measure of average delay in time units. Note that delay is also an indicator of traffic overhead and congestion, as in Crowds forwarded messages translate into repeated packets. Later, in order to represent the anonymity-QoS delay trade-off more simply, as a two-dimensional curve, we shall resort to the combined QoS metric described.

Similar considerations of dependence on specific user preferences and system requirements affect the choice of privacy metrics. Our review of the state of the art, more precisely the subsection on related anonymity metrics,

Sec. 2.3, succinctly described the examination in [26] of a variety of information-theoretic privacy measures under a unifying perspective, which considers privacy as an attacker's estimation error in ascertaining the outcome of a random event. In the current paper, we are concerned with the statistical estimation of the identity of the original sender of a message, carried out by the final receiver, from the observed identity of the last forwarding node. In the theoretical analysis in the next section, the anonymity measure chosen will be the probability of error in the attacker's assumption that the most likely sender is the correct one. As explained in the state of the art section of this paper, this choice is in keeping with the MAP estimation strategy, which [26] discusses in the context of privacy and shows to be equivalent to measuring anonymity as the min-entropy of the probability distribution across possible senders. The cited work investigates alternative privacy measures, including

- Hartley's entropy as a possibilistic metric that simply counts the number of candidate identities, irrespective of their probability,
- Shannon's entropy, as a measure of uncertainty in the set of possible identities that is interpreted as an average quantity involving all probabilities, not just the highest, and
- the parametric family of Rényi entropies that includes the three aforementioned as special cases.

Additional measures of anonymity and considerations regarding their applicability can be found in [1].

The anonymity metric chosen here, based on the most likely identity and equivalent to min-entropy, one among the numerous alternative measures in the literature, reflects a specific concern of the user or designer of the privacy system and its parametrization. Namely, this anonymity metric reflects the most exposed or vulnerable candidate sender, and in this regard may be viewed as a worst-case measure. Obviously, both our choices for QoS and anonymity metrics partly owe to their mathematical tractability and in future research, alternatives might be considered.

We shall quantitatively characterize the optimal anonymity-QoS curve, which represents maximum anonymity for a given QoS, and viceversa, parametrized in terms of the sending probability. In practice, users or system designers may select a desired QoS goal, dependent on the application at hand, and refer to the optimal trade-off curve to find out the best possible level of anonymity attainable, and the corresponding sending probability. Users of privacy-sensitive applications may feel alternatively inclined to fix an anonymity level first. Additionally, we shall define two points of operation within the curve, called absolute and relative equilibria, which consist in points where a small increment in QoS corresponds to an equivalent increment in anonymity, and where small relative increments or percentages match, respectively. Said equilibria may come in handy as quantitative reference values to further assist users and system designers in their decisions.

Finally, the implementation of our modified Crowds protocol would entail the decision and communication of its main working parameter, namely the probability of direct sending to the intended recipient. Regarding its decision, the metrics proposed would inform users and system designers of the impact in terms of QoS and anonymity. As far as communication is concerned, the sending probability could simply be agreed upon as the group of collaborating nodes is formed, or accompany the recipient's header if chosen by the sender on a per message basis.

4. FORMAL PROBLEM STATEMENT AND MAIN RESULTS

In this section we formulate our modification of the Crowds protocol in a lossy network, and present the main theoretical result characterizing the trade-off between quantifiable measures of anonymity and QoS.

Throughout the paper, we shall follow the convention of uppercase letters for *random variables* (r.v.'s), and lowercase letters for particular values they take on. For compactness, for any probability expression p, we write $1-p$ as $\bar{p}$.

4.1. Formal Problem Statement

Consider $n \geqslant 2$ collaborating users wishing to communicate with a common, untrusted receiver. In order to attain a certain degree of anonymity, each user flips a biased coin and depending on the outcome chooses to send the message directly to the receiver Rx or else to another user, who is asked to perform the same exact probabilistic forwarding. More precisely, we suppose that for each forwarding operation, with (*link*) *loss probability* $q < 1$, the message in question is lost. If no loss occurs, with *sending probability* $p > 0$, the message is sent directly to the receiver. Otherwise, it is forwarded to any of the other users with equal probability $1/(n-1)$, where the entire probabilistic process will be repeated. This process is depicted by a Markov chain in Fig. 1, with two absorbing states modeling the receipt and the loss of a message. We avoided cluttering the figure with arrows such as those coming out from the rest of the users, entirely analogous to the first's.

Define the (*extended*) *delay* Δ as an r.v. in $\{1, 2, \dots\} \cup \{\infty\}$, equal to the number of times the message is sent from its origin to the final recipient when it is not lost, and infinity otherwise. Natural measures of QoS are the *average delay* $\delta_{\text{avg}} = \mathrm{E}[\Delta | \Delta < \infty]$, and the *end-to-end loss probability* $q_{\text{end}} = \mathrm{P}\{\Delta = \infty\}$. Note that both Δ and δ_{avg} are delays in terms of message *hops* rather than direct time units, that is, hops due to the forwarding protocol described in order to improve anonymity, even in networks

DOI: 10.1002/sec
Prepared using secauth.cls

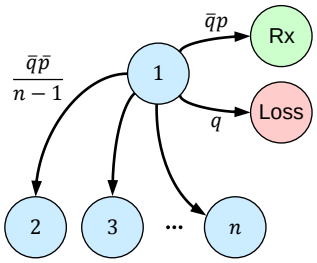

Figure 1. Markov model of a Crowds-like forwarding mechanism for anonymous communication in a lossy network.

where direct communication would otherwise require a single hop.

Let F be an r.v. represent the *first*, original sender of a message, and L, the *last* sender in the forwarding path to the receiver. When a loss occurs, simply define $L = \infty$. The untrusted receiver, who is assumed to know the forwarding protocol, the loss model and their precise parameters, strives to estimate the original sender F from observing the last L, as the most likely. Let $\hat{F}$, be such MAP estimate. A reasonable measure of *anonymity* is the probability that the estimate, defined for received messages, is erroneous:

$$a = \mathrm{P}\{\hat{F} \neq F \,|\, \Delta < \infty\}.$$

This probability of error may be equivalently written as an average error across all messages received,

$$a = \mathrm{E}_{L \,|\, \Delta<\infty} \, \mathrm{P}\{\hat{F} \neq F \,|\, L, \, \Delta < \infty\} =$$
$$= \sum_{i=1}^{n} \mathrm{P}\{L = i \,|\, \Delta < \infty\} \; \mathrm{P}\{\hat{F} \neq F \,|\, L = i, \, \Delta < \infty\}.$$

We shall assume that the receiver assigns equal probability $\mathrm{P}\{F = i\} = 1/n$ to each possible first sender F of a message $i = 1, \ldots, n$, prior to the observation of the last L, although this assumption may be easily relaxed, as we remark later.

We argued in Secs. 1.1 and 3.1 that our modification of the Crowds protocol focused on stringent QoS requirements and that concordantly eliminated the first forwarding step in the traditional proposal. In terms of the model introduced, the independent, uniform choice of forwarding node in this initial step would render F and L statistically independent (conditionally on successful reception $\Delta < \infty$). This means that if the first forwarding step were enforced, we would equate priors and posteriors, having

$$\mathrm{P}\{F = i \,|\, L = j, \, \Delta < \infty\} = \mathrm{P}\{F = i\},$$

and under the assumption of equal prior probability aforementioned, we would attain perfect anonymity $a = 1 - 1/n$ from the perspective of the untrusted receiver. However, such anonymity would come at the cost of doubling the minimum delay, that is, at the cost of making $\Delta \geqslant 2$ (with probability 1) for any sending probability p, no matter how large, which would impact both δ_{avg} and q_{end} negatively.

4.2. Fundamental Theorems

The following results theoretically characterize the anonymity-QoS trade-off in our model of Crowds in lossy networks. Proofs are provided in the next section.

In these results, we define the *effective sending probability* $p_{\text{eff}} = 1 - \bar{p}\bar{q}$ ($\overline{p_{\text{eff}}} = \bar{p}\bar{q}$). It is clear that if $q = 0$, then $p_{\text{eff}} = p$. Interestingly, we shall discover that part of the behavior of the protocol in a lossy network with sending probability p and loss probability q replicates that of a lossless network with sending probability precisely p_{eff}.

Theorem 1 (QoS)

i. Δ conditioned on $\Delta < \infty$ is geometrically distributed with parameter p_{eff}.
ii. $\delta_{\text{avg}} = 1/p_{\text{eff}}$.
iii. $q_{\text{end}} = q/p_{\text{eff}} = q\,\delta_{\text{avg}}$.

Theorem 2 (Anonymity)

i. $\mathrm{P}\{F = i \,|\, L = i\} =$
$= \mathrm{P}\{L = i \,|\, F = i, \, \Delta < \infty\} =$
$= \mathrm{P}\{F = L \,|\, \Delta < \infty\}$ for any user $i = 1, \ldots, n$.
ii. $\mathrm{P}\{F = L \,|\, \Delta < \infty\} = \frac{1+(n-2)p_{\text{eff}}}{n-p_{\text{eff}}}$.
iii. $\hat{F} = L$ and $a = \mathrm{P}\{F \neq L \,|\, \Delta < \infty\} = \frac{(n-1)\overline{p_{\text{eff}}}}{n-p_{\text{eff}}}$.

In light of the above theorems, from the point of view of the receiver, the most likely identity of the original sender of a message turns out to be the last's, which justifies the *predecessor attack* cited in Sec. 2.2. (Careful inspection of our proofs will show that the uniformity assumption on the message generation rate is only needed to conclude that the MAP estimate is $\hat{F} = L$. Precisely, if such estimation rule were taken as a starting hypothesis rather than a consequence, all of the results in the theorems in this section, except for (i) in Theorem 2, would still hold true. Further, for any prior message generation probability $\mathrm{P}\{F = i\}$, no matter how unequal, there exists a sufficiently high p for which $\hat{F} = L$ remains the best attacker's strategy.)

Further, one may regard a network with loss probability q and sending probability p as a lossless network with a higher effective sending probability $p \leqslant p_{\text{eff}} < 1$, where the left inequality holds with equality when $q = 0$, and the right one in the limit as $q \to 1$. This is consistent with the intuition that higher link loss probability decreases the likelihood of longer message forwarding. Lastly, careful inspection of the proofs shows that allowing self-forwarding would make no difference in terms of anonymity, at the expense of worse QoS.

Thus far, we have presented two separate, traditional QoS metrics on the extended delay Δ. However, we may

combine both the effects of end-to-end losses and delay in a single quantity, for a simpler representation of the anonymity-QoS trade-off on a unique plane. One example of such combined (inverse) QoS metric is

$$c = \mathrm{P}\{\Delta > \delta_{\max}\}$$

for some maximum delay $\delta_{\max}$ tolerable by a given messaging application, which we may regard as the cost of anonymity. It is routine[(a)] to check that

$$c = \mathrm{P}\{\Delta > \delta_{\max}\} = q_{\mathrm{end}} + \overline{q_{\mathrm{end}}}\,\overline{p_{\mathrm{eff}}}^{\,\delta_{\max}}. \qquad (1)$$

Additional examples are percentiles of Δ, such as the median or the 90$^{\mathrm{th}}$ percentile, suitable for average-case and worst-case scenarios, respectively.

4.3. Further Trade-Off Analysis

To complete our characterization of the anonymity-QoS trade-off, we proceed to draw a series of consequences of the above theorems. As they only require straightforward notions of algebra and calculus, proofs are omitted or merely hinted at. In short, these results shed some light on the usability in a lossy network, from the perspective of impact on QoS, of the Crowds-like protocol for anonymity just described. We succinctly and conceptually recapitulate the main conclusions in Sec. 7.

Our initial consequences are graphically summarized in Fig. 2. For each $q \in [0, 1)$, p_{eff} is an increasing, affine function of $p \in (0, 1]$, with infimum q and maximum 1. As p vanishes, i.e., in the *high-anonymity region*, p_{eff} approaches q, and a tends to its supremum $\frac{(n-1)\bar{q}}{n-q}$, upper bounded by $1 - 1/n$, corresponding to a uniform posterior probability distribution on F given L, ideal from the anonymity standpoint, but reachable only for $q = 0$ in the limit of small p. In this region, δ_{avg} and q_{end} approach their suprema, $1/q$ and 1, respectively. For $p \simeq 1$, corresponding to the *high-QoS region*[(b)],

$$a \simeq \delta_{\mathrm{avg}} - 1 = q_{\mathrm{end}}/q - 1.$$

For $p = 1$, $\delta_{\mathrm{avg}} = 1$ and $q_{\mathrm{end}} = q$, their respective minimum values. The trade-off itself has the same shape, regardless of whether QoS is measured as δ_{avg} or q_{end}, and a is an increasing, strictly concave function, which means

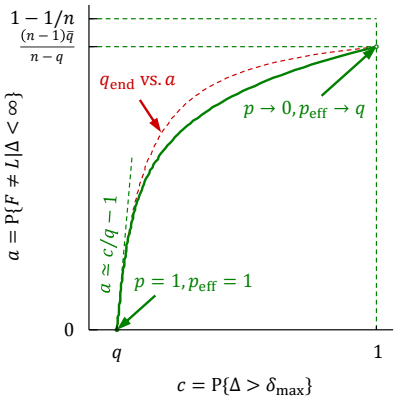

Figure 3. Trade-off between anonymity a and QoS, the latter expressed by means of the metric $c = \mathrm{P}\{\Delta > \delta_{\max}\}$, combining both the effects of delay and end-to-end loss.

that the protocol has diminishing returns, albeit always positive.

Fig. 3 provides a snapshot of the anonymity-QoS trade-off as a single curve, simply by merging the effects of both delay and end-to-end losses into the QoS metric c defined in (1). Since

$$c = \mathrm{P}\{\Delta > \delta_{\max}\} > \mathrm{P}\{\Delta = \infty\} = q_{\mathrm{end}},$$

the trade-off with respect to c appears shifted towards the right of the one with respect to q_{end}, plotted in Fig. 2(b) and superimposed in Fig. 3. It is clear from its definition that c will approach the end-to-end loss metric q_{end} in the limit of increasing maximum delay tolerance $\delta_{\max}$, i.e., $\lim_{\delta_{\max}\to\infty} c = q_{\mathrm{end}}$. Less obvious is the fact that for any fixed $\delta_{\max}$, c becomes asymptotically equivalent to q_{end} in the high-QoS region, i.e., $\lim_{p\to 1} c/q_{\mathrm{end}} = 1$.

A striking observation is that[(c)] in the high-QoS region, $a \simeq c/q - 1$. This means that whenever the link loss probability q is small, the rate of anonymity gain with respect to QoS degradation is highly favorable, as the slope of the curve in this region approaches $1/q$.

4.4. Absolute and Relative Equilibria

The characterization of the anonymity-QoS trade-off carried out thus far may very well suffice when making an informed decision regarding the specific point of operation within the curves analyzed. Typically, a user or system designer might simply specify a desired QoS, which immediately determines the best anonymity attainable, or vice versa. Still, there exists a couple of natural points

---

[(a)] Write

$$\mathrm{P}\{\Delta > \delta_{\max}\} = \mathrm{P}\{\Delta = \infty\} + \mathrm{P}\{\Delta < \infty\}\ \mathrm{P}\{\Delta > \delta_{\max} \mid \Delta < \infty\}.$$

In light of Theorem 1(i), the conditioned event $\Delta > \delta_{\max}$ now represents $\delta_{\max}$ failures of a geometric r.v.

[(b)] This is a first-order Taylor approximation to a as a function of δ_{avg} for $p \simeq 1$, equivalent to $p_{\mathrm{eff}} \simeq 1$, where we compute

$$\frac{\mathrm{d}a}{\mathrm{d}\delta_{\mathrm{avg}}} = \frac{\frac{\mathrm{d}a}{\mathrm{d}p_{\mathrm{eff}}}}{\frac{\mathrm{d}\delta_{\mathrm{avg}}}{\mathrm{d}p_{\mathrm{eff}}}},$$

for $p_{\mathrm{eff}} = 1$ from the formulas in Theorems 1 and 2.

[(c)] As before, a first-order Taylor approximation to a is computed for $p \simeq 1$, this time in terms of c, given in (1), after carefully calculating, for the equivalent condition $p_{\mathrm{eff}} = 1$,

$$\frac{\mathrm{d}a}{\mathrm{d}c} = \frac{\frac{\mathrm{d}a}{\mathrm{d}p_{\mathrm{eff}}}}{\frac{\mathrm{d}c}{\mathrm{d}p_{\mathrm{eff}}}}.$$

DOI: 10.1002/sec
Prepared using secauth.cls

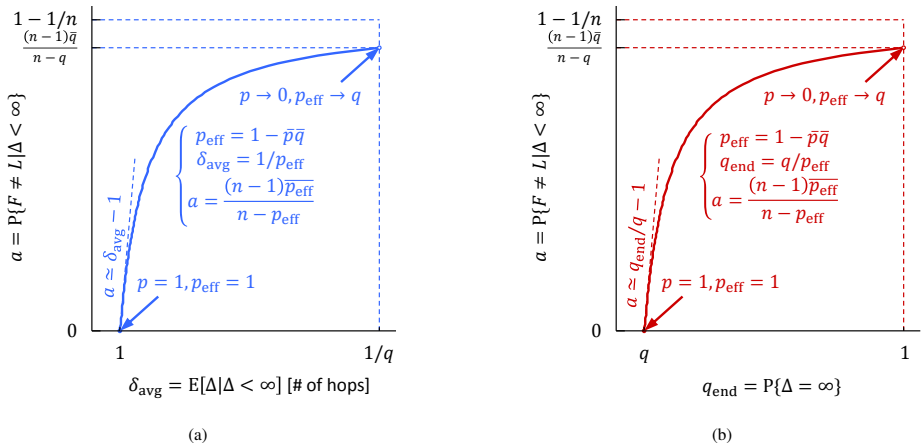

Figure 2. Trade-off between anonymity a and QoS, the latter in terms of (a) average delay δ_{avg} and (b) end-to-end loss probability q_{end}.

of equilibrium within the trade-off we would like to consider here, even if merely as additional information to assist in said decision. Throughout this subsection, we content ourselves with the special case of combined QoS metric (1).

The problem of maximizing the anonymity a while minimizing the QoS degradation c may be approached from the perspective of multiobjective optimization. In essence, when one wishes to minimize several competing costs, as it would technically be the case of $-a$ and c here, it is customary to consider the minimization of their *Lagrangian cost* [5]. This is effectively a weighted sum of those individual costs, modeling their overall impact, where the nonnegative weights, called *Lagrangian multipliers*, represent the importance of one optimization objective with respect to the rest.

Accordingly, define the Lagrangian cost $-a + \lambda\, c$, which we wish to minimize, and in which the multiplier λ quantifies the importance of QoS degradation in relation to anonymity, clearly application dependent. Because of the simple one-to-one correspondence between the sending probability p and its effective version p_{eff}, we may think of a and c as functions of either, the latter leading to somewhat simpler expressions; and of course, we may view a directly as a function of c.

The Lagrangian-optimal operation point within the trade-off is given by any of the following equivalent conditions:

$$\frac{\mathrm{d}a}{\mathrm{d}p_{\text{eff}}} = \lambda \frac{\mathrm{d}c}{\mathrm{d}p_{\text{eff}}}, \quad \frac{\mathrm{d}a}{\mathrm{d}c} = \lambda. \tag{2}$$

Graphically, this corresponds to the point of the c-a curve in Fig. 3 with slope λ. We shall refer to the solution, whether in p or p_{eff}, as the *absolute equilibrium*, in the sense that it represents the sending probability for which a small increment in cost would lead to an increment in anonymity with the same overall impact. Informally, $\mathrm{d}a = \mathrm{d}(\lambda c)$.

We already mentioned at the end of Sec. 4.3 that in the high-QoS region the steep slope of the curve could be interpreted as an argument in favor of implementing our protocol, or at the very least against unprotected, direct delivery ($p = 1$). Indeed,

$$\left.\frac{\mathrm{d}a}{\mathrm{d}c}\right|_{p=1} = \frac{1}{q}, \tag{3}$$

a large gain under the mild assumption of a small link loss probability q. Mathematically, there exists an absolute equilibrium $p < 1$ for any weight $\lambda < 1/q$.

Alternatively, we might be interested in a point of operation within the trade-off such that relative increments in both objectives, rather than absolute increments, match. Informally, $\mathrm{d}a/a = \mathrm{d}c/c$, where either member of the equation may be thought of as a small percentage. More formally, we define the *relative equilibrium* as the solution, whether in p or p_{eff}, to any of the following equivalent conditions:

$$\frac{1}{a}\frac{\mathrm{d}a}{\mathrm{d}p_{\text{eff}}} = \frac{1}{c}\frac{\mathrm{d}c}{\mathrm{d}p_{\text{eff}}}, \quad \frac{\mathrm{d}a}{\mathrm{d}c} = \frac{a}{c}. \tag{4}$$

Note that this equilibrium is invariant with respect to scaling of either of the objective functions. Under the perspective of relative gains and the assumption that losses exist, since at $p = 1$, $a = 0$ but $c = q > 0$, there is a strong incentive to avoid message delivery without any anonymity protection, mathematically reflected by the fact that for $q > 0$,

$$\left.\frac{1}{a}\frac{\mathrm{d}a}{\mathrm{d}c}\right|_{p=1} = \infty. \tag{5}$$

We would like to remark that the notion of logarithmic derivatives enables us to connect this type of equilibrium with the previous one. Precisely, define $\tilde{a} = -\ln a$ and $\tilde{c} = -\ln c$. Since in this case both a and c are probabilities, $\tilde{a}$ and $\tilde{c}$ are nonnegative. Because

$$\frac{\mathrm{d}\tilde{a}}{\mathrm{d}p_{\text{eff}}} = -\frac{1}{a}\frac{\mathrm{d}a}{\mathrm{d}p_{\text{eff}}},$$

$$
\begin{aligned}
\frac{\mathrm{d}a}{\mathrm{d}p_{\text{eff}}} &= -\left(\frac{n-1}{n-p_{\text{eff}}}\right)^2 \\
\frac{\mathrm{d}c}{\mathrm{d}p_{\text{eff}}} &= \frac{1}{p_{\text{eff}}^2}\left(-q+\left(q\,\overline{p_{\text{eff}}}-\delta_{\max}(p_{\text{eff}}-q)\,p_{\text{eff}}\right)\,\overline{p_{\text{eff}}}^{\,\delta_{\max}-1}\right) \\
\frac{1}{a}\frac{\mathrm{d}a}{\mathrm{d}p_{\text{eff}}} &= \frac{1-n}{(n-p_{\text{eff}})\,\overline{p_{\text{eff}}}} = \frac{1}{n-p_{\text{eff}}} - \frac{1}{\overline{p_{\text{eff}}}} \\
\frac{1}{c}\frac{\mathrm{d}c}{\mathrm{d}p_{\text{eff}}} &= \frac{-q+\left(q\,\overline{p_{\text{eff}}}-\delta_{\max}(p_{\text{eff}}-q)\,p_{\text{eff}}\right)\,\overline{p_{\text{eff}}}^{\,\delta_{\max}-1}}{p_{\text{eff}}\left(q+(p_{\text{eff}}-q)\,\overline{p_{\text{eff}}}^{\,\delta_{\max}}\right)}
\end{aligned}
\tag{6}
$$

and similarly for the rest of variables, the condition for relative equilibrium becomes that for absolute equilibrium, with $\lambda = 1$, in terms of the transformed objectives. In a doubly logarithmic graphical representation, this equilibrium would correspond to the point in the curve with unit slope.

We would like to finish with explicit expressions (6) of the traditional and logarithmic derivatives involved in these equilibria, obtained after careful simplification but ultimately straightforward. Replacing the members of the first form of each of the equilibria equations (2) and (4) by their corresponding expression in (6), polynomial equations in p_{eff} are obtained. These polynomial equations may be solved numerically with mathematical and computational software such as Matlab or Mathematica. The former, for instance, provides the function `roots`, which exploits the fact that finding the roots of a polynomial is equivalent to finding the eigenvalues of its companion matrix [16].

5. THEORETICAL ANALYSIS OF THE FUNDAMENTAL THEOREMS

We first develop a couple of lemmas on a certain type of Markov chains with absorbing states we call exit states, which will serve as a stepping stone towards the theoretical resolution of the problem formulated in Sec. 4, conducted next.

5.1. Markov Chains with Exit States

Consider the binary Markov chain of Fig. 4(a). In this chain, no arrow is depicted to represent that 0 is an absorbing state, meaning that, once entered, it cannot be left. Denote the transition probability from state 1 to state 0 by e, so that the return probability to 1 is $\bar{e}$. Assuming that the starting state is 1, and considering only the portion $1\,1\ldots1\,0$ of the sequence of states until 0 is reached, we may refer to 0 as an *exit state*, and define the *exit time* as the r.v. T determined by the length of such finite subsequence, including 0. Clearly, T is a *geometric r.v.* with parameter e, the *exit probability*, and recall that, consequently, $\mathrm{E}[T] = 1/e$.

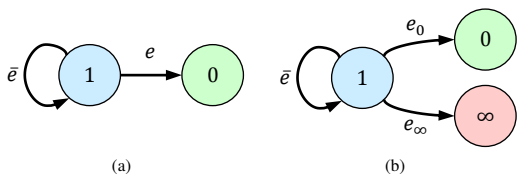

Figure 4. Markov chain representing (a) a geometric r.v., and (b) its generalization with an additional exit state.

We generalize this Markov chain by introducing the additional exit state ∞, as depicted in Fig. 4(b), with exit probability e_∞, for a total exit probability $e = e_0 + e_\infty$. Assuming again that the chain is started at state 1, the exit time T is redefined analogously, considering now the length of subsequences of the form $1\,1\ldots1\,0$ or $1\,1\ldots1\,\infty$. From these definitions, the probabilities of each of such subsequences are $\bar{e}^{\,t-1}e_0$ and $\bar{e}^{\,t-1}e_\infty$, respectively, for any given exit time $t = 1, 2, \ldots$ Finally, define the *exit outcome* as a binary r.v. E taking values on $\{0, \infty\}$, corresponding to the last state of the aforementioned subsequences. The following lemma characterizes the *geometric r.v. with an additional exit state* represented in Fig. 4(b).

Lemma 3
Consider a geometric r.v. with an additional exit state, under the previous assumptions, with exit time T and exit outcome E.

i. T and E are statistically independent.
ii. The distribution of T, whether conditioned on the event $E = 0$ or not, is geometric with parameter e.
iii. $\mathrm{P}\{E = 0\} = e_0/e$ and $\mathrm{P}\{E = \infty\} = e_\infty/e$.

Proof: The statistical independence between T and E is immediate from their definition, which also implies that T conditioned on $E = 0$ is distributed exactly as its unconditioned version. To see that the latter T is geometrically distributed with parameter $e = e_0 + e_\infty$, simply regard the two exit states as a macrostate. Lastly, due to the symmetry in the definition of the two exit states, it suffices to show the third statement for one of them:

$$
\mathrm{P}\{E = 0\} = \sum_{t=1}^{\infty} \mathrm{P}\{E = 0,\, T = t\} =
$$

DOI: 10.1002/sec
Prepared using secauth.cls

$$= \sum_{t=1}^{\infty} \bar{e}^{\,t-1} e_0 = \frac{e_0}{e}.$$

Alternatively, apply independence to write $\mathrm{P}\{E=0\} = \mathrm{P}\{E=0 \,|\, T=1\}$, proportional to $\mathrm{P}\{E=0,\, T=1\}$, and conclude that $\mathrm{P}\{E=0\}/\,\mathrm{P}\{E=\infty\} = e_0/e_\infty$. ■

We may now proceed to extend Lemma 3 to a *Markov chain with two exit states*, enabling us to address the theoretical analysis of the problem formulated in Sec. 4. Specifically, consider a Markov chain with finite state space $\{1,\ldots,n\}$ and transition matrix $P=(p_{ij})_{ij}$. Assume we enlarge this chain with two exit states 0 and ∞, i.e., absorbing states, with equal transition probabilities e_0 and e_∞ from each of the original n states, adding up to a total exit probability $e = e_0 + e_\infty$. The new transition probabilities between the original states are obtained from the original p_{ij} simply by multiplying by $\bar{e}$, representing that transitions occur with the original likelihood in the absence of exit. Suppose further that the chain is started at one of the original states $i_0 = 1,\ldots,n$ with probability π_{i_0}. As previously, we are only interested in the *sequence of states until the exit event*, which we denote by $\boldsymbol{I} = I_0\, I_1\, \ldots\, I_{T-1}$, where I_0 is the initial state, T the exit time, and the exit outcome E would occur immediately after I_{T-1}. Thus,

$$\mathrm{P}\{\boldsymbol{I}=\boldsymbol{i},\, E=0\} = \\ = \pi_{i_0}\, p_{i_1\, i_0}\, \cdots\, p_{i_{t-1}\, i_{t-2}}\, \bar{e}^{\,t-1}\, e_0, \quad (6)$$

and similarly for $E=\infty$. The following lemma characterizes this type of Markov chain.

Lemma 4
Consider a Markov chain with two exit states, under the previous assumptions, with initial state probabilities π_{i_0}, random sequence of states until the exit event $\boldsymbol{I}$, exit time T, and exit outcome E. Viewing the set of n original states as a single macrostate, it is clear that T is geometrically distributed with parameter e, and that all properties of Lemma 3 hold for T and E, in particular. More generally,

i. $\boldsymbol{I}$ and E are statistically independent.
ii. For any sequence $\boldsymbol{i}$ until the exit event,

$$\mathrm{P}\{\boldsymbol{I}=\boldsymbol{i} \,|\, E=0\} = \mathrm{P}\{\boldsymbol{I}=\boldsymbol{i}\} = \\ = \pi_{i_0}\, p_{i_1\, i_0}\, \cdots\, p_{i_{t-1}\, i_{t-2}}\, \bar{e}^{\,t-1}\, e.$$

Proof: The statistical independence between $\boldsymbol{I}$ and E is immediate from their definition, which also implies that $\mathrm{P}\{\boldsymbol{I}=\boldsymbol{i} \,|\, E=0\} = \mathrm{P}\{\boldsymbol{I}=\boldsymbol{i}\}$. The last equation in the lemma follows from rewriting (6) for $\mathrm{P}\{\boldsymbol{I}=\boldsymbol{i}\}$, viewing the two exit states as a single macrostate with exit probability $e = e_0 + e_\infty$. Alternatively, the equation in question can be shown by writing

$$\mathrm{P}\{\boldsymbol{I}=\boldsymbol{i} \,|\, E=0\} = \mathrm{P}\{\boldsymbol{I}=\boldsymbol{i},\, E=0\}/\,\mathrm{P}\{E=0\},$$

and then applying (6) and Lemma 3(iii). ■

The importance of the statistical independence results stated in (i) of both Lemma 3 and 4 is best understood under the well-known fact that for any two events, the latter with positive probability, independence is equivalent to requiring that the prior on the first be equal to its posterior given the second. Mathematically, A and B are statistically independent, that is, $\mathrm{P}(A\cap B) = \mathrm{P}(A)\,\mathrm{P}(B)$, if and only if, $\mathrm{P}(A|B) = \mathrm{P}(A)$, under the mild assumption that $\mathrm{P}(B) > 0$. We would also like to stress that statements (ii) in both Lemma 3 and 4 mean that conditioning on $E = 0$ preserve the behavior of the corresponding stochastic processes. That is, the role of the total exit probability e is preserved in the distribution of both T and $\boldsymbol{I}$, running contrary to any intuition that might suggest replacing e by the exit probability e_0 of the conditioning outcome.

5.2. Proofs of the Main Theorems

We proceed to prove our main results, stated in the theorems in Sec. 4. Both proofs resort to the lemmas in the previous part of this section, identifying the exit states 0 and ∞ depicted in Fig. 4(b), with the events of sending and losing a message shown in Fig. 1, respectively. Under this correspondence, the exit probabilities are $e_0 = \bar{q}\,p$, $e_\infty = q$ and

$$e = q + \bar{q}\,p = 1 - \bar{q} + \bar{q}\,p = 1 - \bar{p}\,\bar{q} = p_{\text{eff}}.$$

Proof of Theorem 1: Recall the geometric r.v. with an additional exit state of Lemma 3, represented in Fig. 4(b). The extended delay Δ defined in Sec. 4 may be expressed as T when $E=0$, and ∞ otherwise. Bearing in mind the exit state correspondence aforementioned, observe that the distribution of Δ conditioned on $\Delta<\infty$ precisely coincides with that of T conditioned on $E=0$, which the lemma asserts to be geometric with parameter $e = p_{\text{eff}}$, proving (i) in the theorem. Statement (ii) is an immediate consequence of the well-known fact that the expectation of a geometric r.v. is the inverse of its parameter. The last statement of the theorem follows from its counterpart in Lemma 3, by identifying $\mathrm{P}\{E=\infty\} = q_{\text{end}}$, $e_\infty = q$ and $e = p_{\text{eff}}$. ■

Proof of Theorem 2 (Sketch): We proceed by considering decreasingly simplified variations of the model represented in Fig. 1. Consider first the special case without losses, $q=0$, and concordantly disregard any conditioning on $\Delta<\infty$. Suppose further that users were allowed to forward messages to themselves, so that the transition probabilities in the corresponding Markov chain became $\bar{p}/n$ in lieu of $\bar{p}\,\bar{q}/(n-1)$. In this variation of the problem, regardless of i, we claim that

$$\mathrm{P}\{L=i \,|\, F=i\} = p + \bar{p}/n \geqslant 1/n. \quad (7)$$

Indeed, either i sends directly the message to the receiver, with probability p, or else, with probability $\bar{p}$, the message is forwarded with equal probability $1/n$ to any user, including the original sender, and any consecutive hops

will maintain this uniform uncertainty. Thus

$$\mathrm{P}\{L = i \,|\, F = i\} = \mathrm{P}\{F = L \,|\, F = i\}$$

is a constant quantity. To verify the bound, simply apply $p \geqslant p/n$.

We assumed equal prior sending rate among users. Hence, by symmetry, $\mathrm{P}\{F = i\} = 1/n = \mathrm{P}\{L = i\}$, and consequently[(d)],

$$\begin{aligned}\mathrm{P}\{F = i \,|\, L = i\} &= \frac{\mathrm{P}\{F = i\}}{\mathrm{P}\{L = i\}}\,\mathrm{P}\{L = i \,|\, F = i\} = \\ &= \mathrm{P}\{L = i \,|\, F = i\} = \mathrm{P}\{F = L \,|\, F = i\} = \\ &= \frac{1}{n}\sum_i \mathrm{P}\{F = L \,|\, F = i\} = \mathrm{P}\{F = L\} = \bar{a}.\end{aligned}$$

Again by symmetry, $\mathrm{P}\{L = j \,|\, F = i\}$ will remain the same for any $j \neq i$. For that reason, the bound in (7) implies that $\mathrm{P}\{L = j \,|\, F = i\}$ is maximized at $j = i$. Concordantly, the MAP estimator is $\hat{F} = L$.

Still in the special case $q = 0$, suppose now that self-forwarding is no longer allowed, so that transition probabilities between distinct nodes corresponding to users are $\bar{p}/(n-1)$. Although this may seem a problem fairly different from the completely symmetric simplification with self-forwarding above, a clever application of Lemma 3 will enable us to transform it into the former simplification.

Before proceeding, we need the immediate generalization of statement (iii) in the lemma to n exit states with transition probabilities e_j, $j = 0$ representing direct sending, and $j > 0$ forwarding to any of the other $n-1$ users. For a total $e = \sum_j e_j$, the return probability modeling self-forwarding in the former simplification would be $r = \bar{e}$. Said generalization of (iii) would guarantee that the probability of each exit outcome $E = i$ would be $e_i/e = e_i/\bar{r}$.

Back to the argument relating the earlier, symmetric simplification allowing self-forwarding, and the variation without self-forwarding, it may help thinking of the latter strategy as implemented exactly as the former, with the caveat that the self-forwarding event remains hidden from an external observer and yields no delay.

Define p' as the sending probability in the earlier, symmetric simplification allowing self-forwarding. Then, from the point of view of the statistics involving F and L, both strategies are utterly equivalent under the transformation given by $\mathrm{P}\{E = 0\} = e_0/\bar{r}$ for the new forwarding probability $p = p'/(1 - \overline{p'}/n)$.

After routine algebraic manipulation, $p' = (n-1)p/(n-p)$, and on account of (7) and the hypothesis of uniform prior message generation, we conclude

$$\mathrm{P}\{L = i \,|\, F = i\} = \mathrm{P}\{F = L\} = \frac{1 + (n-2)p}{n-p}, \quad (8)$$

and

$$\mathrm{P}\{F \neq L\} = \overline{\mathrm{P}\{F = L\}} = (n-1)\bar{p}/(n-p).$$

Finally, we move to the most general case, with $q \geqslant 0$ and without self-forwarding. To complete the proof, it will suffice to apply Lemma 4 to the previous variation for $q = 0$. Specifically, we identify the absence of loss represented by the event $\Delta < \infty$ with the exit outcome $E = 0$, and apply assertion (ii) in the lemma, on the fact that after conditioning, the behavior of the underlying stochastic process remains the same, in terms of the total exit probability $e = p_{\text{eff}}$ in lieu of the exit probability of the conditioning outcome, e_0. Accordingly, we replace p by p_{eff} in (8). The two consecutive transformations of the problem prove all three statements in the theorem, in the most general version. ■

6. NUMERICAL EXAMPLE

In order to confirm and illustrate the theoretical results in Secs. 4.2 through 4.4, we conduct a simulation of the Markov chain corresponding to the full-fledged version of the problem, with losses and without self-forwarding.

In our simulation, a total of $n = 10$ users follow our variation of the Crowds protocol on a network with link loss probability $q = 0.1$. The combined QoS metric (1) is used, for a maximum delay tolerance of $\delta_{\max} = 5$ hops. The protocol is repeated for each of the sending probabilities $p = 0.05, 0.25, 0.5, 0.75, 1$, and for 10^4 messages uniformly generated by the users, for each of those probabilities. The anonymity a and the QoS cost c are estimated directly as the corresponding relative frequencies found empirically.

The results of the simulation are shown in Fig. 5, in which the simulated points quite accurately lie along the theoretical trade-off curve, verifying the analysis in Fig. 3. Further, we numerically compute the absolute and relative equilibria, the former for equal weight ($\lambda = 1$), following the method explained at the end of Sec. 4.4. As expected, at the absolute equilibrium, corresponding to $p \simeq 0.373$, the linear curve has unit slope, and so does the logarithmic curve at the relative equilibrium, $p \simeq 0.536$.

The equilibrium condition (2), together with the steep slope in the high-QoS region of Fig. 5(a), given by (3), suggest that unless QoS cost is weighted 10 times more than anonymity, our protocol should be used ($p \neq 1$). Anonymity is bound by the supremum

$$\frac{(n-1)\,\bar{q}}{n-q} \simeq 0.818,$$

[(d)] For any events A and B with positive probability,

$$\mathrm{P}(A|B) = \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)} = \frac{\mathrm{P}(A)}{\mathrm{P}(B)}\frac{\mathrm{P}(A \cap B)}{\mathrm{P}(A)} = \frac{\mathrm{P}(A)}{\mathrm{P}(B)}\,\mathrm{P}(B|A).$$

DOI: 10.1002/sec
Prepared using secauth.cls

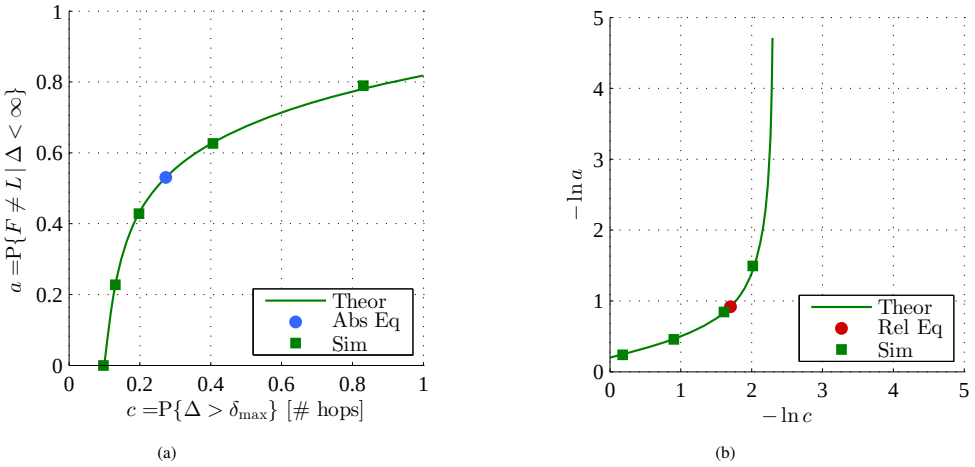

Figure 5. Trade-off between anonymity a and combined QoS cost c, represented linearly (a) and after logarithmic transformation (b), with $n = 10$, $q = 0.1$ and $\delta_{\max} = 5$. We conduct simulations for 10^4 messages and $p = 0.05, 0.25, 0.5, 0.75, 1$, whose results are superimposed onto the theoretical curve. At the absolute equilibrium $p \simeq 0.373$ with weight $\lambda = 1$, the linear curve has unit slope, and so does the logarithmic curve at the relative equilibrium $p \simeq 0.536$.

below the ideal value of $1 - 1/n = 0.9$, attainable in the limit of $p \to 0$ only when $q = 0$. Loosely speaking, losses degrade anonymity, as $p_{\text{eff}} > q$ means that the effective sending probability cannot be made arbitrarily small.

If we are concerned with relative gains in lieu of absolute increments, $q > 0$ suffices to argue strongly in favor of using the protocol, as (5) mathematically reflects. In this regard, the vertical asymptote at $-\ln q \simeq 2.30$ in Fig. 5(b) responds to the fact that as $p \to 1$, $c \to q$ but $a \to 0$.

7. CONCLUSION

We propose a theoretical model of the trade-off between anonymity and QoS of a Crowds-like protocol, suitable for lossy networks. The anonymity metric chosen adheres to and illustrates the recently established principle of pragmatically measuring privacy as an attacker's estimation error.

The necessarily limited scope of this preliminary contribution on the subject of Crowds in networks with losses contemplates only the special case of single-hop wireless networks, thus excluding the multihop case, largely because the former already requires a sufficiently intricate analysis.

Still, by introducing the presence of message losses, we extend the applicability of the protocol beyond the types of networks considered in the original Crowds proposal. We quantify the intuition that anonymity now comes at the expense of, not only delay, but additional end-to-end losses. Focusing on stringent QoS requirements, we concordantly eliminate the initial forwarding step of the original version of the protocol.

Our analysis shows that packet losses lead to a higher effective sending probability, as longer forwarding paths lead to end-to-end loss. Decreasing the sending probability yields significant albeit diminishing returns in terms of anonymity.

When measuring QoS in terms of the probability that message delay exceeds a maximum tolerance threshold, we find that the gain in anonymity per QoS cost in the high-QoS region is inversely proportional to the link loss probability, and thus potentially very favorable under small values of such loss likelihood. In addition, absolutely no anonymity is provided if direct delivery is enforced, while losses will impose imperfect QoS even in this case. This strongly argues in favor of anonymity mechanisms, even if we are only willing to accept minimal QoS degradation.

Notwithstanding the limitation of our work to single-hop networks, we expect that our first steps towards introducing losses in the modified Crowds protocol, while analyzing the contrasting aspects of anonymity and QoS jointly, may very well lay part of the fundamental principles to approach the theoretical study of the more intricate case of multihop networks in future endeavors. Additional future research avenues include more extensive simulations with precise wireless network models, both single-and multihop, comparing different anonymous-communication protocols or variations thereof, along with alternative measures of anonymity and QoS reflecting diverse application requirements.

Last but not least, and beyond the mathematical details specific to the problem at hand, this work may be construed as an illustration of a systematic approach to privacy-enhancing strategies. In this approach, we contemplate both privacy and utility in a quantifiable manner that

enables us to address the important issue of their inherent trade-off.

ACKNOWLEDGMENT

We would like to thank J. Moreira for a number of helpful comments. This work was partly supported by the Spanish Government through projects Consolider Ingenio 2010 CSD2007-00004 "ARES" and TEC2010-20572-C02-02 "Consequence", by the Government of Catalonia under grants 2009 SGR 1362 and FI-AGAUR, and by the UAS in Mexico. D. Rebollo-Monedero is the recipient of a Juan de la Cierva postdoctoral fellowship, JCI-2009-05259, from the Spanish Ministry of Science and Innovation.

REFERENCES

1. R. Bagai and N. Jiang, "Measuring anonymity by profiling probability distributions," in *Proc. IEEE Int. Conf. Trust, Secur., Priv., Comput., Commun. (TRUSTCOM)*, Liverpool, UK, Jun. 2012, pp. 366–374.
2. K. Bauer, D. McCoy, D. Grunwald, T. Kohno, and D. Sicker, "Low-resource routing attacks against anonymous systems," University of Colorado, Tech. Rep., 2007.
3. G. Bianchi, M. Bonola, V. Falletta, F. S. Proto, and S. Teofili, "The SPARTA pseudonym and authorization system," *Sci. Comput. Program.*, vol. 74, no. 1–2, pp. 23–33, 2008.
4. M. Boban, G. Misek, and O. K. Tonguz, "What is the best achievable QoS for unicast routing in VANETs?" in *Proc. IEEE Global Telecomm. Conf. (GLOBECOM)*, New Orleans, LA, Dec. 2008, pp. 1–10.
5. S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, UK: Cambridge University Press, 2004.
6. P. Cencioni and R. D. Pietro, "VIPER: A vehicle-to-infrastructure communication privacy enforcement protocol," in *Proc. IEEE Int. Conf. Mob. Ad hoc, Sensor Syst. (MASS)*, Pisa, Italy, 2007, pp. 1–6.
7. D. Chaum, "Untraceable electronic mail, return addresses, and digital pseudonyms," *Commun. ACM*, vol. 24, no. 2, pp. 84–88, 1981.
8. ——, "Security without identification: Transaction systems to make big brother obsolete," *Commun. ACM*, vol. 28, no. 10, pp. 1030–1044, Oct. 1985.
9. ——, "The dining cryptographers problem: Unconditional sender and recipient untraceability," *J. Cryptol.*, vol. 1, no. 1, pp. 65–75, 1988.
10. S. Chen and K. Nahrstedt, "On finding multi-constrained paths," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Atlanta, GA, Jun. 1998, pp. 874–879.
11. T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New York: Wiley, 2006.
12. G. Danezis, "Statistical disclosure attacks: Traffic confirmation in open environments," in *Proc. Secur., Priv., Age Uncertainty, (SEC)*, Athens, Greece, May 2003, pp. 421–426.
13. C. Díaz, S. Seys, J. Claessens, and B. Preneel, "Towards measuring anonymity," in *Proc. Workshop Priv. Enhanc. Technol. (PET)*, ser. Lecture Notes Comput. Sci. (LNCS), vol. 2482. Springer-Verlag, Apr. 2002, pp. 54–68.
14. R. Dingledine, "Free Haven's anonymity bibliography," 2009. [Online]. Available: www.freehaven.net/anonbib/
15. R. Dingledine, N. Mathewson, and P. Syverson, "Tor: The second-generation onion router," in *Proc. Conf. USENIX Secur. Symp.*, Berkeley, CA, 2004, pp. 21–21.
16. R. A. Horn and R. J. Charles, *Matrix Analysis*. Cambridge, UK: Cambridge Univ. Press, 1985.
17. L. Huang, K. Matsuura, H. Yamane, and K. Sezaki, "Silent cascade: Enhancing location privacy without communication QoS degradation," in *Proc. Int. Conf. Secur. Pervas. Comput. (SPC)*, ser. Lecture Notes Comput. Sci. (LNCS), vol. 3934. York, UK: Springer-Verlag, Apr. 2006, pp. 165–180.
18. D. Kesdogan, D. Agrawal, and S. Penz, "Limits of anonymity in open environments," in *Proc. Inform. Hiding Workshop (IH)*, ser. Lecture Notes Comput. Sci. (LNCS). Noordwijkerhout, The Netherlands: Springer-Verlag, Oct. 2002.
19. B. N. Levine, M. K. Reiter, C. Wang, and M. Wright, "Timing attacks in low-latency mix systems," in *Proc. Int. Financial Cryptogr. Conf.* Springer-Verlag, 2004, pp. 251–265.
20. A. Miede, U. Lampe, D. Schuller, J. Eckert, and R. Steinmetz, "Evaluating the QoS impact of Web service anonymity," in *Proc. IEEE Euro. Conf. Web Serv. (ECOWS)*, Ayia Napa, Cyprus, Dec. 2010, pp. 75–82.
21. S. J. Murdoch and G. Danezis, "Low-cost traffic analysis of Tor," in *Proc. IEEE Symp. Secur., Priv. (SP)*, 2005, pp. 183–195.
22. A. Panchenko and L. Pimenidis, "Crowds revisited: Practically effective predecessor attack," in *Proc. Nordic Workshop Secure IT Syst.*, 2007, pp. 1–6.
23. A. Pfitzmann and M. Hansen, "A terminology for talking about privacy by data minimization: Anonymity, unlinkability, undetectability, unobservability, pseudonymity, and identity management," Aug. 2010, v0.34. [Online]. Available: http://dud.inf.tu-dresden.de/literatur/Anon_Terminology_v0.34.pdf
24. R. Pries, W. Yu, S. Graham, and X. Fu, "On performance bottleneck of anonymous communication networks," in *Proc. IEEE Int. Parallel, Distrib. Process. Symp. (IPDPS)*, Miami, FL, Apr. 2008.
25. D. Rebollo-Monedero, J. Forné, A. Solanas, and T. Martnez-Ballesté, "Private location-based

information retrieval through user collaboration," *Comput. Commun.*, vol. 33, no. 6, pp. 762–774, 2010. [Online]. Available: http://dx.doi.org/10.1016/j.comcom.2009.11.024

26. D. Rebollo-Monedero, J. Parra-Arnau, C. Diaz, and J. Forné, "On the measurement of privacy as an attacker's estimation error," *Int. J. Inform. Secur.*, 2012, to appear. [Online]. Available: http://arxiv.org/pdf/1111.3567v3.pdf
27. M. K. Reiter and A. D. Rubin, "Crowds: Anonymity for Web transactions," *ACM Trans. Inform. Syst. Secur.*, vol. 1, no. 1, pp. 66–92, 1998.
28. A. Serjantov and G. Danezis, "Towards an information theoretic metric for anonymity," in *Proc. Workshop Priv. Enhanc. Technol. (PET)*, vol. 2482. Springer-Verlag, 2002, pp. 41–53.
29. A. Serjantov, R. Dingledine, and P. Syverson, "From a trickle to a flood: Active attacks on several mix types," in *Proc. Inform. Hiding Workshop (IH)*. Springer-Verlag, 2002, pp. 36–52.
30. M. Srivatsa, L. Liu, and A. Iyengar, "Preserving caller anonymity in voice-over-IP networks," in *Proc. IEEE Symp. Secur., Priv. (SP)*, May 2008, pp. 50–63.
31. H. Sui, J. Wang, J. Chen, and S. Chen, "An analysis of forwarding mechanism in Crowds," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2003, pp. 261–265.
32. A. Vaha-Sipila and T. Virtanen, "BT-Crowds: Crowds-style anonymity with Bluetooth and Java," in *Proc. IEEE Annual Hawaii Int. Conf. Syst. Sci. (HICSS)*, Washington, DC, 2005, pp. 1–11.
33. M. K. Wrigh, M. Adler, B. 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, Nov. 2004.
34. G. Xu, "Wireless Crowds based on NTRU," Iowa State Univ., Tech. Rep., 2001.
35. H. Xu, X. Fu, Y. Zhu, R. Bettati, J. Chen, and W. Zhao, "A scalar anonymous communication system," in *Proc. Int. Conf. Netw., Mob. Comput. (ICCNMC)*, ser. Lecture Notes Comput. Sci. (LNCS), vol. 3619. Zhangjiajie, China: Springer-Verlag, Aug. 2005, pp. 452–461.