Improving Sender Anonymity in a Structured Overlay with Imprecise Routing [*]

Giuseppe Ciaccio

DISI, Università di Genova
via Dodecaneso 35, 16146 Genova, Italy
ciaccio@disi.unige.it

Abstract. In the framework of peer to peer distributed systems, the problem of anonymity in structured overlay networks remains a quite elusive one. It is especially unclear how to evaluate and improve *sender* anonymity, that is, untraceability of the peers who issue messages to other participants in the overlay. In a structured overlay organized as a chordal ring, we have found that a technique originally developed for *recipient* anonymity also improves sender anonymity. The technique is based on the use of imprecise entries in the routing tables of each participating peer. Simulations show that the sender anonymity, as measured in terms of average size of anonymity set, decreases slightly if the peers use imprecise routing; yet, the anonymity takes a better distribution, with good anonymity levels becoming more likely at the expenses of very high and very low levels. A better quality of anonymity service is thus provided to participants.

1   Introduction and motivation

Overlay networks are receiving a lot of attention by the research community, as flexible and scalable low-level infrastructures for distributed applications of many kinds: network storage [18, 13, 39], naming [12], content publication [16, 11, 3, 37, 46, 40], multicast/anycast [36, 6, 31], and communication security [33, 47]. They have also been proposed as general networking infrastructures [17, 44, 20, 19], because of their potential ability to decouple network addresses from physical placements of cooperating hosts, an important feature for privacy and mobility.

The vast population of existing or proposed overlay systems can be broadly divided into two families, namely, unstructured overlays and structured overlays.

Structured overlays [35, 14, 30, 23, 38, 29, 48] are receiving far more attention lately, because of performance guarantees they can in principle provide thanks to their regular topologies. Regular topologies allow routing algorithms to provably converge, and a careful choice of entries in routing tables can reduce the number of routing hops to even a constant quantity, independent of the overlay size [22, 27]. The most known example of a structured overlay is the chordal ring [45]

(Figure 1): N peers are arranged in a circle, and each can route messages via its own *successor* in the ring as well as a small ($O(\log(N))$) number of other peers, called *fingers*, whose "distances" increase according to a geometric progression. With this organization, a message can be delivered in $O(\log(N))$ hops according to a so called "greedy" routing (Figure 2 and Section 3.1).

On the other hand, unstructured overlays like Freenet [11] and GNUnet [2] first leveraged techniques to enhance identity privacy or *anonymity* of participant entities.

Both families of overlays share a common goal, namely, to implement a layer of virtual addressing and message routing on top of the Internet addressing and packet routing infrastructure. Each host participating to the overlay is said to be *responsible* for (or *owner* of) a range of overlay virtual addresses. Messages can be issued by any participant, and are targeted to overlay addresses rather than Internet addresses; the routing algorithm of the overlay implements the correspondence between the target address (an overlay address) and the destination host (an Internet address).

In this respect we easily identify at least two anonymization possibilities. Mostly researched upon is *sender* anonymity, namely, the untraceability of the Internet address of a host which issued a given message. Indirection based on source rewriting, usual cryptographic machinery, or, even better, mix chains [7, 4], can help hide the identity of a message sender, that is, improve sender anonymity. But there is another face of the coin, namely, *recipient* anonymity, which in this context means hiding the correspondence between any given overlay address A (the target of a message) and the Internet address of the peer who is responsible for A (the actual receiver of the message).

In a distributed system for content publication, the actions of producing and making use of a content are implemented by letting each participant send suitable messages (respectively "write" and "read") to other entities in the network which happen to store the information. In such a kind of systems, sender anonymity is thus a key ingredient for protecting the privacy of those people who are either producing or making use of any contents. On the other side, recipient anonymity is a key ingredient for censorship resistance, in that it makes it difficult for a censor to locate and then attack the physical place where the target piece of information is stored. In a distributed storage system (but also in the real world), censorship resistance without user privacy makes no sense: readers of unlawful information, when identified, can be prosecuted. Thus, sender anonymity and recipient anonymity may not live separated from each other, and any potential trade-off between these two features must be considered with the greatest care.

The overall goal of our investigation is to understand and improve both the user privacy and the censorship-resistance properties of structured overlay networks. In a previous work of ours [10] we have proposed a technique, that we have called *imprecise routing*, aimed at enhancing the censorship resistance of a chordal ring. The technique, based on the use of deliberately inaccurate entries in the routing tables of all peers, has been shown to be effective in hiding, to some extent, the correspondence between overlay addresses and Internet addresses,

without compromising the nice routing properties of this family of overlay networks. In other words we were able to enforce recipient anonymity in the overlay, thus providing a necessary condition for censorship resistance, without sacrificing too much the routing efficiency. In this paper we report about the subsequent step, namely, a study of the interplay between recipient anonymity (related to the censorship resistance) and sender anonymity (related to privacy of users) when the technique of imprecise routing is in place. We have carried out simulations which shows that imprecise routing is beneficial to sender anonymity as well: with imprecise routing, the amount of sender anonymity takes a better distribution, with good anonymity levels becoming more likely at the expenses of very high and very low levels. A more uniform and effective quality of anonymity service is thus provided to participants.

The paper is structured as follows: after defining the adversarial model and the anonymity metrics (Section 2), we recall the ideas behind imprecise routing (Section 3). Then, we report the results of our simulation study on sender anonymity with imprecise routing (Section 4). After a brief survey on the few existing works in the field (Section 5), the paper closes with a summary of conclusions and open issues, in which we also mention NEBLO, a working implementation of the concepts accounted in this paper.

2   Preliminary assumptions

We believe that the entire work presented here could be adapted to any structured overlay. Nevertheless, for practical purposes we had to choose a reference model of overlay network. We focused on the most successful such model, namely, the aforementioned chordal ring.

The overlay supports the abstraction of a generic *address space*, consisting of the set of 2^k binary words of k bits ordered as a circle modulo 2^k. This space is mapped onto the ring of peers in consecutive chunks or *address intervals*; thus, each peer *owns* a well defined address interval. For the purpose of our work, it is uninteresting to give meaning to the data possibly "stored" at each overlay address. In other words we choose an application-neutral standpoint, and therefore prefer the terms "overlay network" and "address space" to the more popular "distributed hash table" and "key space".

Our discussion assumes an adversarial model that, following Diaz et al. [15], we term "internal, local, and passive"; that is, the adversary controls and can orchestrate a number of peers in the system, each of which complies to the overlay protocol and does not generate malicious traffic, but can maliciously gather information from its internal routing tables as well as any messages it happens to forward.

In order to enforce some sender anonymity, our system relies on pure indirection with no mixes nor cover traffic; in such a case, the adversary has no convenience in injecting extra traffic in the system. Violations of the routing algorithm can be excluded from our adversarial model, because in a structured overlay the routing choices are constrained by the overlay graph, and thus any

violation could be easily detected. Global adversaries, either external (capable of observing possibly any message across the entire overlay) or internal (capable of controlling possibly any peer in the network) appear to be unrealistic in a large peer to peer system. So, we conclude that our model of an "internal, local, and passive" adversary is reasonable in a stable overlay. However, when a peer first joins the overlay, or whenever it tries to rebuild its own routing table, an "active" adversary is given chance to take over by playing a suitable protocol; this shall be briefly discussed in Section 6.

We also assume that the overlay protocol does not explicitly disclose the identity of any participant.

Various metrics for sender anonymity have been proposed so far [8, 2, 5, 15, 41]. In this paper we conform to other existing studies on structured overlays [26, 32, 42] by adopting the size of the anonymity set [8] as a metrics. The anonymity set is the set of those participants who are considered as being possible senders for a given message. The adversary will make its best to narrow down the anonymity set, usually by making use of routing information concerning the intercepted message. If a message is not intercepted by the adversary, the anonymity set is conventionally the whole set of those participants not colluding with the adversary.

Some of the proposed anonymity metrics are based on the entropy within the anonymity set and thus might be more accurate in some cases. We now show why these metrics are unneeded in our scenario. The first adversarial peer P_A who happens to intercept a given message M has the shortest distance from the sender of M. P_A directly knows the peer P_l which it has received M from, whereas the possible predecessors of P_l in the routing path followed by M are unknown. Based on the knowledge of the (greedy) routing algorithm of the chordal ring (Section 3.1), the best P_A can do is to compute how many well-formed routes cross with one another at P_l, and hence the size of the smallest set of possible originators of M (which P_l indeed belongs to), with no chance of discriminating any better within that set (whose members are unknown to P_A, with the only exception of P_l). Later interceptions of M by other adversarial peers are of no help: they occur at greater distance from the sender, and the routing rules does not depend upon the sender, so a later interception cannot gather more information than an earlier one.

3 Imprecise routing

3.1 Generalized chordal rings

Let us consider a set of peers logically organized into an overlay shaped as a ring. Each peer has a link to its own *successor* in the ring; "to have a link" means to store $\langle IP\,address, listening\,port \rangle$ of the linked peer in the own routing table. If peer P owns the address interval from A_l to A_u in the address space, and peer Q is the successor of P, then all addresses owned by Q are greater than A_u (modulo 2^k). For better resiliency, each peer has a *successor list*, rather that just

one immediate successor. This allows a peer to talk directly to its successor's successor to seal the ring in case the successor has gone (the extension to the case of multiple adjacent faulty peers is straightforward).

In order to keep the routing path below an acceptable size, each peer also knows additional peers called the *fingers*. We present here a generalized version of the concept originally introduced by Stoica et al. [45]. [1] A finger is a link (an entry in the own routing table), pointing to a distant peer in the overlay. The distance is measured between (one of the bounds of) the local address interval and (the corresponding bound of) the address interval owned by the remote peer. Each peer maintains its own list of fingers, the elements of which are ordered by increasing distance. Finger distances obey a mathematical requirement that we call the *distance rule*. The distance rule is often geometric on base 2. Given a bottom value C, called *cutoff*, the first finger has the largest possible distance $\leq C$ from local peer, the second finger has the largest possible distance $\leq 2C$, the third finger has distance $\leq 4C$, and so on, up to spanning half of the address space. The finger at distance $C \cdot 2^m$ is said to have *magnitude m*; we will also call it the "finger m" for brevity. Clearly, each peer can have at most $O(\log(N))$ fingers. A ring of peers, enhanced by fingers, becomes what we call a chordal ring. Figure 1 illustrates this concept.

The routing algorithm takes advantage of fingers in a so-called "greedy" way (Figure 2). When a peer P gets an incoming message whose destination address is A, it acts as follows:

1. P checks out if A is locally owned; if so, the message has arrived and no routing is needed;
2. otherwise, P computes the residual distance D yet to be travelled by the message, as the difference between A and (one of the bounds of) the locally owned address interval;
3. P chooses the finger of largest magnitude whose distance does not exceed D, and forwards the message to it. If no such finger is found, P forwards to successor.

In a chordal ring with complete finger tables conforming to a geometric distance rule, a total travel distance of D is covered in $O(\log_2(D/C))$ hops.

The most efficient way to build and maintain a finger table takes advantage from the recursive nature of the geometric distance rule. To find the finger 0, P sends a suitable request along its successor chain, until the most distant peer still within cutoff distance C is found. To find a finger of magnitude $m > 0$, P asks its own current finger $m - 1$ to be contacted by its finger $m - 1$. [2] Such an *incremental procedure* minimizes the number of contacted peers, so it should be preferred when anonymity is of concern, because it can minimize the information leak towards potential adversaries.

---

[1] Similar concepts are found in every scalable overlay.

[2] In case the address interval of P spans the entire cutoff distance, the finger of magnitude 0 could not be found. In this case P starts by directly searching its finger of magnitude n along successor chain within distance $C \cdot (n + 1)$, where n is such that $C \cdot (n + 1)$ is larger than the size of P's address interval.
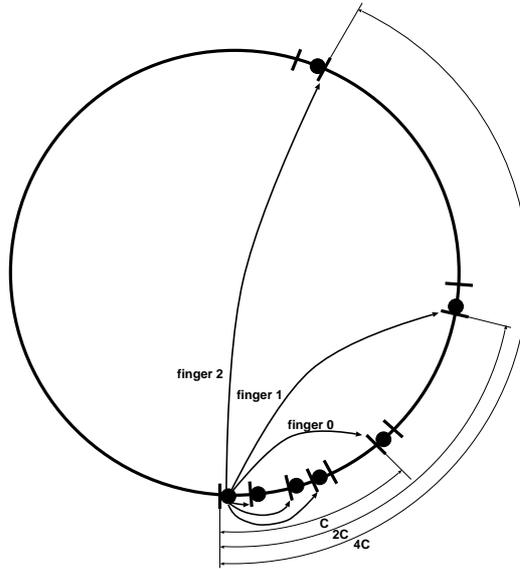
Fig. 1. An instance of a chordal ring. Each peer is responsible for a contiguous interval of overlay addresses. Each peer has links to some successors, and other links called "fingers" pointing to peers at distances C, 2*C, 4*C, etc., where C is a parameter of the system. With N participants, each peer "knows" $O(log(N))$ other peers, and can easily infer their overlay addresses thanks to the above geometric progression.

3.2 Improving recipient anonymity with imprecise routing

Recipient anonymity is broken when the adversary knows which peer (identified by IP address) is responsible for which overlay address. Clearly, the above (traditional, after Chord [45]) definition of fingers poses two serious threats on recipient anonymity, namely:

1. If peer P has peer Q as its own finger of magnitude m, then P knows that Q's address interval is more or less at distance $C \cdot 2^m$ from itself. Thus, Q's address interval is indirectly disclosed to P. In general, in a ring counting N peers, each participant has $O(\log(N))$ fingers and thus can deduce the address intervals of as many other peers. A malicious coalition counting $O(N/\log(N))$ peers can thus build a *map of the overlay*, namely, a map where all participants (identified by their IP addresses) are related to the overlay addresses they are responsible for.
2. When searching finger 0, peer P exposes its own address interval to the whole successor chain up to the finger. This can help an adversarial coalition to harvest useful information for building the aforementioned map of the overlay.

The two anonymity flaws above are impossible to fix, because they are implied by the traditional definition of fingers. To improve recipient anonymity
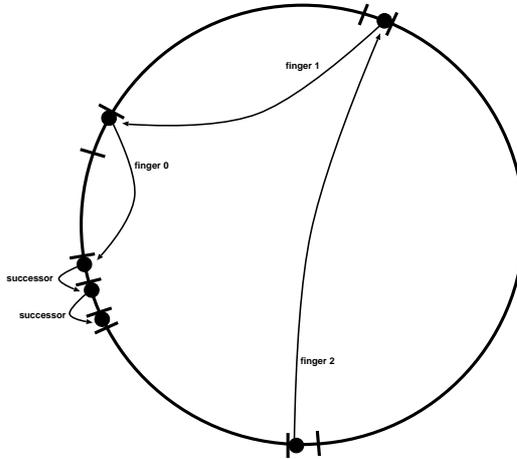
Fig. 2. "Greedy" routing in a chordal ring. With N participants and complete routing tables, $O(\log(N))$ hops are sufficient.

we must shift to a slightly different definition. Our goal is to obfuscate part of the topological information conveyed by traditional fingers, and to protect peers against excessive exposure when they search their fingers of magnitude 0. The solution envisaged in our previous work [10], is that a routing table should only be allowed to contain a small and fixed amount of exact addressing information, whereas most of the information in the table should be deliberately made *imprecise* by construction. Such construction, whose details are reported in [10], ensures that the distance of any finger of generic magnitude m is never fully known; the optimal distance of $C \cdot 2^m$ is affected by a random and unknown error in $[0, C \cdot 2^{m-1}[$, so that the actual distance is an unknown random value in $[C \cdot 2^{m-1}, C \cdot 2^m[$. [3] (Figure 3).

Such an amount of finger imprecision is a good device for recipient anonymity. The higher a finger's magnitude, the lesser the information the finger conveys about the remote peer it points to. As a result, only large adversarial coalitions can harvest sufficiently exact information from finger tables. In [10] we have also shown that routing convergence in a logarithmic number of hops is preserved even with imprecise routing.

4   Imprecise routing and sender anonymity

Imprecise routing is aimed at recipient anonymity, yet its use would be impractical, if sender anonymity was compromised by this. But we come to the

---

[3] The distribution of distance corresponds to the convolution of $m+2$ uniform random variables over $[0, C/4[$.
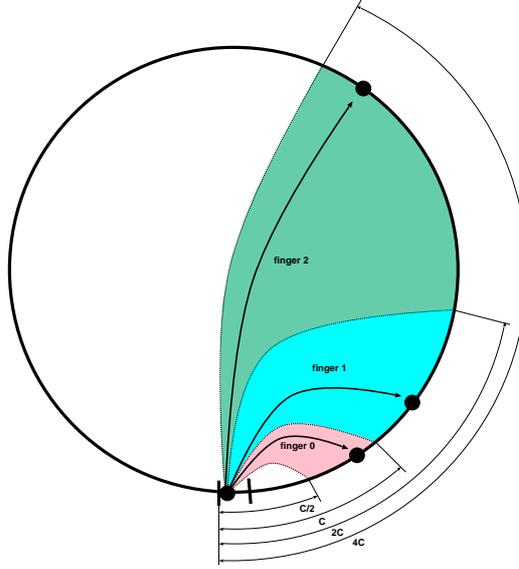
Fig. 3. A chordal ring (successors omitted) in which the fingers are affected by an unknown random error. The average error increases proportionally with the distance of the peer. This way, no peer can infer much about the overlay addresses of other peers, and this improves recipient anonymity. Yet, $O(log(N))$ hops are still sufficient to route messages towards an arbitrary destination.

main contribution of this paper: not only imprecise routing does not compromise sender anonymity, it is even beneficial in improving the quality of sender anonymity provided to participants.

To validate such a claim we have built a simulator for a chordal ring over an address space made of 32-bit addresses. By acting upon a handful of parameters, we could simulate rings with imprecise fingers as well as traditional rings with exact fingers.

In the systems with imprecise fingers, all fingers are built according to the incremental procedure outlined in Section 3.1. The cutoff distance C is a critical parameter because it affects the average number of successors in between each peer and its corresponding finger 0. If C was too small, the inaccuracy affecting the routing tables would be small as well, with lesser guarantees of recipient anonymity. Therefore, C should depend on the number N of participants in the overlay. For a ring with N participants, our simulator initializes the parameter C as follows: the initial value is set to cover 20 bits of overlay address, then this value is doubled again and again until it exceeds the quantity $10 * 2^{32}/N$, namely, ten times the average size of each peer's address interval. This ensures that C is chosen in such a way that the distance between each peer and its corresponding finger 0 covers 10 consecutive peers on average. In a real-world system, hovever, C should be a fixed parameter known to all peers. In order to

allow C to be constant in a real system with an unpredictable and unknown number of participants, each peer might just evaluate the size of its own address interval and, based on this, decide the minimum magnitude of the first finger to be inserted in the routing table. Setting a minimum magnitude $M > 0$ is tantamount to applying a scale factor 2^M to the cutoff distance C, without forcing other partecipants to do the same (which would be impossible).

By contrast, in the systems with exact routing tables, fingers are computed explicitly rather than incrementally, in order to avoid that higher magnitude fingers could be affected by cumulated inaccuracy arising from fingers of lower magnitude. The cutoff distance is set to 1 as with traditional chordal rings.

After creating a sample ring with N uniformly distributed peers, the simulator fills up each peer's successor list and finger table; fingers can be imprecise or exact, depending on a compile-time flag. The simulator then generates all the routing paths from each peer to the peer owning the overlay address 0 (any destination address is equivalent to 0 modulo rotational transformation of the chordal ring). At this point, the simulator generates a number of sample adversarial configurations over the ring; the fraction f of adversaries over the entire population is specified at runtime, and the simulator obtains each adversarial configuration by marking each peer as adversary with probability f. For each adversarial configuration, the simulator computes statistics of the anonymity sets of all "honest" peers, by processing the set of all routing paths as follows:

1. For each "honest" peer P, scan the routing path from P to address 0 until the first adversarial peer is found. Let us call $last(P)$ the result of the scan. If the path does not meet adversaries, $last(P)$ is assigned the pair $\langle -1, -1 \rangle$; otherwise $last(P)$ is assigned the pair $\langle C, m \rangle$, where C is the first adversarial peer found along the path, and m is the magnitude of the finger which was the last hop in the path up to C, or -1 if such last hop was a successor link. The reason why we take the magnitude of last hop into account shall become clear at the next step.

2. For each adversarial peer C, count all "honest" peers P such that $last(P) = \langle C, m \rangle$ with given m. Let us call $a(C, m)$ such count. $a(C, m)$ is the size of the anonymity set that the colluder C can associate to a generic lookup for address 0 that it could intercept. The reason why this anonymity set depends on m is that C can indeed discriminate among possible originators of a lookup by looking at which incoming link the lookup has came from; intuitively, a lookup coming from the immediate predecessor may have a lot of possible originators, whereas a lookup coming from a link corresponding to the finger of greatest magnitude may only have one originator (namely, the opposite peer on the ring).

3. For each "honest" peer P, if $last(P) = \langle C, m \rangle = \langle -1, -1 \rangle$ then the anonymity set size from P's point of view is equal to the total number of "honest" peers in the ring, namely, $f \cdot N$; otherwise, the anonymity set size is equal to

$a(C, m)$. The case of unintercepted lookup is thus taken into account when estimating the average sender anonymity from the sender viewpoint. [4]

The results have been obtained by running the simulator over 500 sample rings of given size, each with 100 sample adversarial configurations with given percentage of attackers.

Let us first discuss the average sender anonymity as a function of the distance between sender and recipient (this distance is normalized to the size of the complete address space):

– The overall result is that, with imprecise fingers, the average sender anonymity as a function of distance from destination is often lower but always more uniform, compared to traditional fingers. This is displayed by Figure 4, where chordal rings with both kinds of fingers are compared with one another with varying percentage of attackers. By averaging along the whole range of distances, we see that a system with 10000 peers and 30% attackers yields a sender anonymity of 689 when using traditional fingers and 272 with imprecise fingers, a 61% loss. With 50% attackers the loss is 41% but the level of anonymity is however too small (anonymity decreases from 71 to 42). On the other hand, with 10% attackers the loss is just 27% (from 4620 to 3379). Systems with 1000 peers show a lesser impact of imprecise fingers on sender anonymity (maximum loss is 33%, at 30% attackers). To summarize, with imprecise fingers the average sender anonymity becomes less dependent on the target of queries, while the resulting loss of anonymity is not substantial unless the system is large and highly compromised by the adversary. The fundamental reason for this behaviour, is that the routing paths with imprecise fingers become longer (mainly because of the cutoff distance), and more uniform (because of the randomization). Longer paths yield lower sender anonymity, because messages are more likely to get intercepted. However, as we shall see at the end of this Section, randomization leads to a more effective anonymity distibution.
– Another important insight is that, on average, the sender anonymity is in both cases fairly large when the percentage of attackers is not overwhelming. It becomes very poor when this percentage raises 50%, but a system with so many attackers should be considered as highly compromised indeed.

However, the average sender anonymity alone is not informative enough. The variability around the average value must also be considered. We have observed that the variance is always strong, regardless of the fingers being imprecise or not. Figure 5 shows the frequency distribution of sender anonymity in chordal rings with 1000 peers and three different percentages of attackers, averaged along the whole range of distances from destination. The choice between imprecise or traditional fingers leads to deeply different distributions of sender anonymity: with imprecise fingers all distributions span a large interval of pretty good anonymity

---

[4] Actually, the simulator does not require three distinct scans of the entire set of peers in order to accomplish the above three steps.
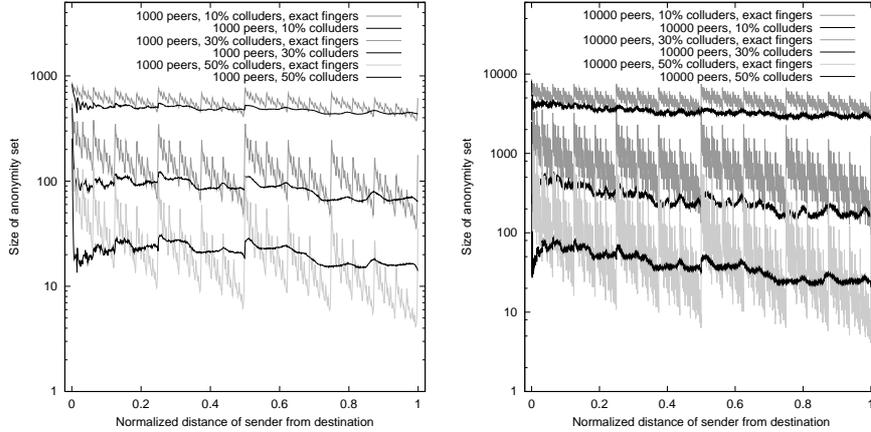
Fig. 4. Average sender anonymity in simulated chordal rings with 1000 and 10000 peers and three different percentages of attackers. Systems with imprecise fingers are compared to systems with exact fingers. The order of captions reflects the order of the curves from top to bottom. Imprecise fingers yield a lower but more regular sender anonymity compared to exact fingers.

levels, as opposed to traditional fingers which only span too small or very large anonymity levels. In other words, imprecise fingers increase the probability of getting a fairly good sender anonymity. The same conclusion can be drawn for systems with 10000 peers (Figure 6). It is the author's opinion that such a better distribution compensates for the lower average level of sender anonymity.

Finally, for a better evaluation we tried to isolate the feature of finger imprecision from other two features that are simultaneously present in the system with imprecise fingers, namely, the cutoff distance, and the incremental finger construction procedure of Section 3.1.

To this end, we first ran the simulator on "hybrid" chordal rings, in which fingers are of traditional kind but the cutoff distance is the same as with imprecise fingers. After simulating overlays with 1000 and 10000 peers, no significant differences emerged with respect to traditional chordal rings of same size.

We then added the feature of incremental finger construction to our "hybrid" rings. This had the consequence of introducing an unavoidable imprecision on fingers of high magnitude, because of the cumulated inaccuracy arising from fingers of lower magnitude. However, the imprecision incurred this way is smaller that the one deliberately implemented with imprecise fingers; with 1000 peers, the highest magnitude finger of the "hybrid" system is affected by an error (measured as the standard deviation w.r.t the exact finger of same magnitude) which is 3.75 times smaller compared to the system with imprecise fingers.

This time some differences appeared in the distribution of sender anonymity (Figure 7), but they only affect the regions of very low and very high anonymity degrees, with no significant differences in between. These are most probably
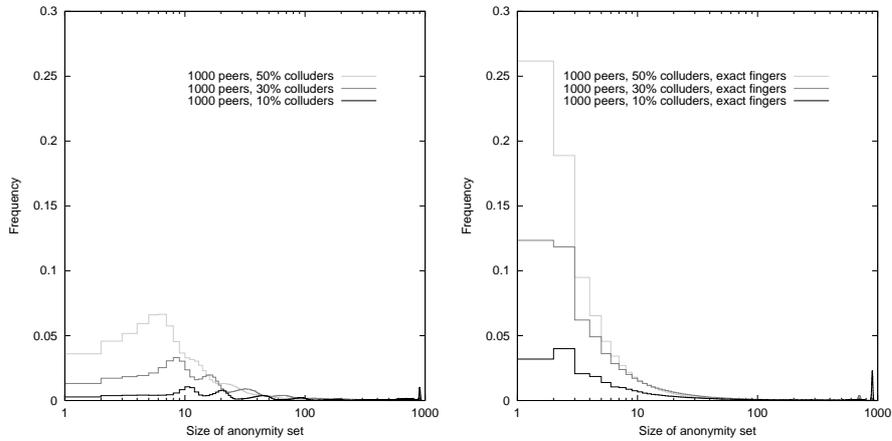
Fig. 5. Average frequency distribution of sender anonymity in simulated chordal rings with 1000 peers and both imprecise and traditional fingers, with three different percentages of attackers. The spikes at right correspond to the cases when messages are not intercepted, yielding the largest possible anonymity set. All curves are heavily affected by the x scale being logarithmic; this must be taken into account when comparing curves from different scenarios.

Fig. 6. Average frequency distribution of sender anonymity in simulated chordal rings with 10000 peers and both imprecise and traditional fingers, with three different percentages of attackers. The spikes at right, corresponding to the cases when messages are not intercepted, are too small to be visible. All curves are heavily affected by the x scale being logarithmic; this must be taken into account when comparing curves from different scenarios.
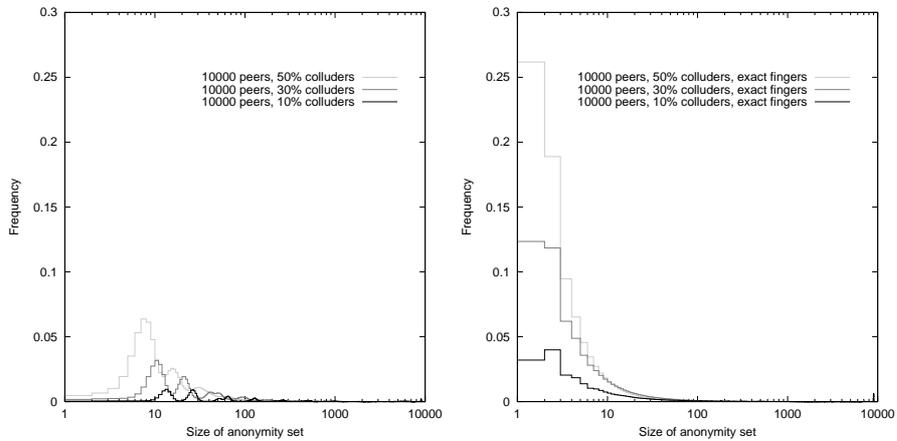
due to the small imprecision affecting high-magnitude fingers in this case. We thus conclude that the improvement in the distribution of sender anonymity, observed in the chordal rings with imprecise routing, is effectively due to finger imprecision.
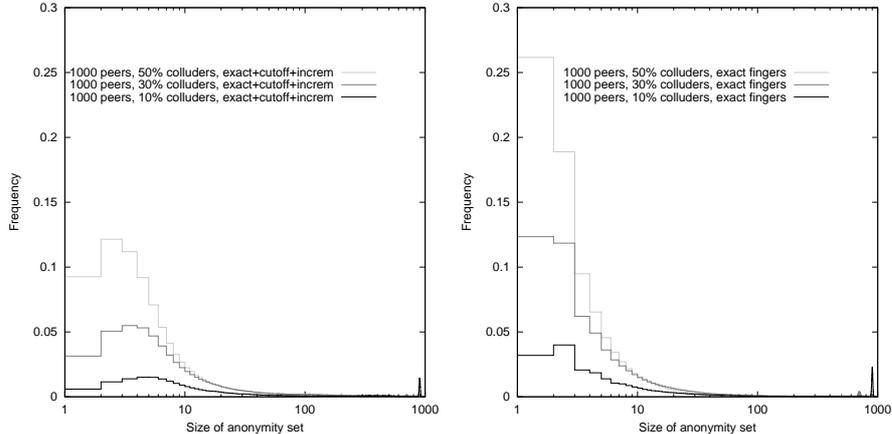

Fig. 7. Average frequency distribution of sender anonymity in simulated chordal rings with 1000 peers and three different percentages of attackers. Left: "hybrid" system with traditional fingers, cutoff distance spanning 10 peers, and incrementally built fingers. Right: traditional Chord system with exact fingers. The differences only affect the regions of very low and very high anonymity.

5 Related work

There are very few attempts to improve anonymity in structured overlays. Achord [25] is an enhancement of Chord [45] with anonymity features. Aiming at enforcing sender anonymity, Achord implements recursive-style [45] routing (because of the indirection) and forces each response to travel back to sender along the same route previously tracked by the corresponding request (so that the sender address need not be disclosed to the receiver; this trick is also cited by Borisov and Waddle [5]).

Other studies [26, 5, 32] focus on measuring sender anonymity in plain Chord. According to O'Donnell and Vaikuntanathan [32], Chord provides a good amount of sender anonymity in terms of size of anonymity set. This is in apparent contradiction with Kannan and Bansal [26]. Apparently, the difference between these two works is that the former considers the anonymity from the attacker point of view, whereas the latter chooses the point of view of the generic "honest" sender. In addition, the latter work shows an analytical mistake, since the event that a lookup is not intercepted by any adversary is overlooked in the anonymity evaluation. Such an event is not so unlikely, and its impact on the average anonymity

set size makes a difference. As we have seen in Section 4, in order to estimate the sender anonymity of our system, we follow the approach suggested by Borisov and Waddle [5] by choosing simulation rather than analytical tools. We too choose the sender viewpoint when estimating sender anonymity, but do not forget about the weight of unintercepted messages, so our results look better that the ones in [26].

Agyaat [43] provides a compromise between anonymity and efficiency by means of a two-level hybrid organization in which the Chord structured overlay works together with the Gnutella unstructured system. Gnutella-like "clouds" are connected with one another by means of a Chord ring. It is an interesting and very effective approach that deserves a deeper anonymity analysis.

Imprecise routing information is at the core of unstructured overlays. With Freenet, for example, a message directed to key A is routed towards a node P if P has previously been able to route back responses from keys "similar" to A [11]. Thus, a routing table entry that points to P does not say anything about the keys actually stored at P, nor does it say much about the placement of P in the overlay topology. GNUnet [2] and MUTE [37] follow a similar approach, with some more randomness. Also SkipNet [24] and Skip Graphs [1], both inspired to the Skip List data structure [34], and Symphony [29], make use of somehow randomized routing entries, although not for anonymity purposes.

The technique of choosing fingers so that they point to sub-optimal distances is also cited by Gummadi et al. [21], as a means of improving routing resilience and neighbour selection while retaining logarithmic-sized routing paths. We have exploited this well known degree of flexibility offered by chordal rings, in order to improve anonymity rather than resilience or neighbour proximity.

6   Conclusions and open issues

The most important result reported in this paper concerns sender anonymity. Previous work has shown that the use of some randomization on long-range connections in structured overlay networks provides better recipient anonymity without sacrificing the nice properties of structured overlays (provable routing convergence and, to some extent, performance). However, we were also concerned with the impact on sender anonymity: the proposed solution would have been impractical, was recipient anonymity obtained at the expenses of sender anonymity. Luckily, the simulations reported in this paper show that the *average* sender anonymity decreases but not so dramatically, and this decrease is compensated by a *better distribution* of the sender anonymity levels: good levels become more likely at the expenses of very low and very high levels.

As an aside, this paper also presents a deep evaluation of sender anonymity of traditional generalized chordal rings.

Our result can be summarized by saying that anonymous routing can be accomplished even in a chordal ring, and can be done in $O(\log(N))$ hops where N is the number of peers in the overlay. If we liked slogans, we would say that anonymity can be asymptotically efficient. The cutoff distance of the chordal

ring is one of the parameters that directly affects the path lengths; future investigations are thus in order, concerning the role of cutoff distance in the trade-offs between anonymity, efficiency, and availability.

The choice of the chordal ring as a reference overlay for our study was not just driven by popularity reasons. In their interesting paper [21], Gummadi et al. have shown that chordal rings provide good resilience to peer failures, a remarkable advantage for real peer-to-peer systems. Although it would be in principle interesting to evaluate the anonymity properties of constant-degree overlays such as Viceroy [28], there is the suspect (a certainty for Viceroy [21]) that constant-degree networks of small degree might have poor resilience against peer failures.

An unexplored security issue is about the algorithm by which a new peer joins the overlay. In order to preserve anonymity, it is crucial that colluding peers be given no control on which position in the overlay they are going to occupy. The obvious, and widely adopted, rule based on the pair $\langle IPaddress, port \rangle$ of the newcomer appears weak as long as the adversary is able to use an IP domain of choice. We are also working at this critical security issue.

A main security concern is about the incremental procedure outlined in Section 3.1, that each peer should follow when building its own finger table. Let us suppose the generic peer P wants to locate its own finger 0. It issues a request which travels along the successor chain until a valid candidate is met. But, if the request meets an adversarial peer, from then on the whole incremental procedure can be diverged towards the adversarial coalition. The finger 0 would be an adversary, and the same would occur with all fingers of greater magnitude, and thus the sender anonymity of P would be entirely compromised. No variant of such a procedure can prevent this kind of opportunistic attack from occurring: any kind of search for fingers may possibly end up in a colluder, and from there on the search can be fully managed within the adversarial coalition. We thus conclude that full sender anonymity is impossible to achieve as long as routing tables are built by running the routing protocol itself, no matter the overlay being structured or not. Yet, one could wonder about algorithms for finger location that decrease the strike probability of this opportunistic attack.

We have managed to embody the mechanism of imprecise routing tables into NEBLO [9], a chordal ring overlay with anonymity features. NEBLO is still in beta development stage, yet it has already been released to the community, under the GNU General Public Licence.

References

1. J. Aspnes and G. Shah. Skip Graphs. In *Proc. of the 14th Annual ACM-SIAM Symp. on Discrete Algorithms (SODA '03)*, January 2003.
2. K. Bennett and C. Grothoff. GAP: Practical Anonymous Networking. In *Proc. of Workshop on Privacy Enhancing Technologies (PET 2003)*, Dresden, Germany, March 2003.
3. K. Bennett, C. Grothoff, T. Horozov, and I. Patrascu. Efficient Sharing of Encrypted Data. In *Proc. of ACISP 2002*, pages 107–120. Springer-Verlag, July 2002.

4. O. Berthold, H. Federrath, and S. Köpsell. Web MIXes: A System for Anonymous and Unobservable Internet Access. In H. Federrath, editor, *Proc. of Designing Privacy Enhancing Technologies: Workshop on Design Issues in Anonymity and Unobservability (PET)*, pages 115–129. Springer-Verlag, LNCS 2009, July 2000.

5. Nikita Borisov and Jason Waddle. Anonymity in Structured Peer-to-Peer Networks. gnunet.org/papers/borisov_waddle.pdf, December 2003.

6. M. Castro, P. Druschel, A. M. Kermarrec, and A. Rowstron. Scribe: A Large-scale and Decentralized Application-level Multicast Infrastructure. *IEEE Journal on Selected Areas in Communications, special issue on Network Support for Multicast Communications*, 20(8), October 2002.

7. D. Chaum. Untraceable electronic mail, return addresses, and digital pseudonyms. *Communications of the ACM*, 4(2), February 1981.

8. D. Chaum. The Dining Cryptographers Problem: Unconditional Sender and Recipient Untraceability. *Journal of Cryptology*, 1(1):65–75, 1988.

9. G. Ciaccio. The NEBLO homepage, http://www.disi.unige.it/project/neblo/.

10. G. Ciaccio. Recipient Anonymity in a Structured Overlay. In *Proc. of the International Conference on Internet and Web Applications and Services (ICIW'06)*, Guadeloupe, French Caribbean, February 2006. IEEE.

11. I. Clarke, O. Sandberg, B. Wiley, and T. W. Hong. Freenet: A Distributed Anonymous Information Storage and Retrieval System. In *Proc. of Designing Privacy Enhancing Technologies: Workshop on Design Issues in Anonymity and Unobservability (PET)*, pages 46–66, July 2000.

12. R. Cox, A. Muthitacharoen, and R. Morris. Serving DNS using a Peer-to-Peer Lookup Service. In *Proc. of the 1st International Peer To Peer Systems Workshop (IPTPS02)*, March 2002.

13. F. Dabek, M. F. Kaashoek, D. Karger, R. Morris, and I. Stoica. Wide-area Cooperative Storage with CFS. In *Proc. of 18th ACM Symp. on Operating Systems Principles*, October 2001.

14. F. Dabek, J. Li, E. Sit, J. Robertson, M. F. Kaashoek, and R. Morris. Designing a DHT for low latency and high throughput. In *Proc. of the 1st USENIX Symposium on Networked Systems Design and Implementation (NSDI '04)*, San Francisco, CA, March 2004.

15. Claudia Díaz, Stefaan Seys, Joris Claessens, and Bart Preneel. Towards measuring anonymity. In Roger Dingledine and Paul Syverson, editors, *Proc. of Privacy Enhancing Technologies Workshop (PET 2002)*. Springer-Verlag, LNCS 2482, April 2002.

16. R. Dingledine, M. J. Freedman, and D. Molnar. The Free Haven Project: Distributed Anonymous Storage Service. In H. Federrath, editor, *Proc. of Designing Privacy Enhancing Technologies: Workshop on Design Issues in Anonymity and Unobservability (PET)*. Springer-Verlag, LNCS 2009, July 2000.

17. J. Eriksson, M. Faloutsos, and S. Krishnamurthy. PeerNet: Pushing Peer-to-Peer Down the Stack. In *Proc. of the 2nd International Workshop on Peer-to-Peer Systems (IPTPS03)*, Berkeley, CA, 2003.

18. J. Kubiatowicz et al. Oceanstore: An Architecture for Global-scale Persistent Storage. In *Proc. of ACM ASPLOS*, November 2000.

19. Michael J. Freedman and Robert Morris. Tarzan: A peer-to-peer anonymizing network layer. In *Proc. of the 9th ACM Conference on Computer and Communications Security (CCS 2002)*, Washington, DC, November 2002.

20. I. Goldberg. *A Pseudonymous Communications Infrastructure for the Internet.* PhD thesis, UC Berkeley, December 2000.

21. K. Gummadi, R. Gummadi, S. Gribble, S. Ratnasamy, S. Shenker, and I. Stoica. The Impact of DHT Routing Geometry on Resilience and Proximity. In *Proc. of ACM SIGCOMM*, August 2003.

22. A. Gupta, B. Liskov, and R. Rodrigues. One Hop Lookups for Peer-to-peer Overlays. In *Proc. of the 9th Workshop on Hot Topics in Operating Systems (HotOS-IX)*, Lihue, Hawaii, May 2003.

23. I. Gupta, K. Birman, P. Linga, A. Demers, and R. van Renesse. Kelips: Building an Efficient and Stable P2P DHT Through Increased Memory and Background Overhead. In *Proc. of the 2nd International Workshop on Peer-to-Peer Systems (IPTPS03)*, Berkeley, CA, 2003.

24. N. Harvey, M. B. Jones, S. Saroiu, M. Theimer, and A. Wolman. Skipnet: A Scalable Overlay Network with Practical Locality Properties. In *Proc. of the 4th USENIX Symposium on Internet Technologies and Systems (USITS '03)*, March 2003.

25. S. Hazel and B. Wiley. Achord: A Variant of the Chord Lookup Service for Use in Censorship Resistant Peer-to-Peer Publishing Systems. In *Proc. of the 1st International Peer To Peer Systems Workshop (IPTPS02)*, March 2002.

26. J. Kannan and M. Bansal. Anonymity in Chord. www.cs.berkeley.edu/ kjk/chord-anon.ps, December 2002.

27. B. Leong and J. Li. Achieving One-Hop DHT Lookup and Strong Stabilization by Passing Tokens. In *Proc. of the 12th International Conference on Networks (ICON)*, November 2004.

28. D. Malkhi, M. Naor, and D. Ratajczak. Viceroy: A Scalable and Dynamic Emulation of the Butterfly. In *Proc. of ACM PODC*, August 2002.

29. G. S. Manku, M. Bawa, and P. Raghavan. Symphony: Distributed Hashing in a Small World. In *Proc. of the fourth USENIX Symposium on Internet Technologies and Systems (USITS'03*, Seattle, WA, March 2003.

30. P. Maymounkov and D. Mazières. Kademlia: A Peer-to-peer Information System Based on the XOR Metric. In *Proc. of the 1st International Peer To Peer Systems Workshop (IPTPS02)*, March 2002.

31. A. Mislove, G. Oberoi, A. Post, C. Reis, P. Druschel, and D. S. Wallach. AP3: Cooperative, Decentralized Anonymous Communication. In *Proc. of 11th ACM SIGOPS European Workshop*, Leuven, Belgium, September 2004.

32. C. O'Donnell and V. Vaikuntanathan. Information Leak in the Chord Lookup Protocol. In *Proc. of the 4th IEEE Int.l Conf. on Peer-to-Peer Computing (P2P2004)*, Zurich, Switzerland, August 2004.

33. P. Perlegos. DoS Defense in Structured Peer-to-Peer Networks. Technical Report UCB-CSD-04-1309, U.C. Berkeley, March 2004.

34. W. Pugh. Skip Lists: a Probabilistic Alternative to Balanced Trees. *Comm. of ACM*, 33(6):668–676, 1990.

35. S. Ratnasamy, P. Francis, M. Handley, R. Karp, and S. Shenker. A Scalable Content-Addressable Network. In *Proc. of ACM SIGCOMM*, August 2001.

36. S. Ratnasamy, M. Handley, R. Karp, and S. Shenker. Application-level Multicast Using Content-addressable Networks. In *Proc. of 3rd Int.l Workshop on Networked Group Communication*, November 2001.

37. J. Rohrer. MUTE: Simple, Anonymous File Sharing. http://mute-net.sourceforge.net/.

38. A. Rowstron and P. Druschel. Pastry: Scalable, Distributed Object Location and Routing for Large-scale Peer-to-peer Systems. In *Proc. of Int.l Conf. on Distributed System Platforms*, November 2001.

39. A. Rowstron and P. Druschel. Storage Management and Caching in PAST, a Large-scale, Persistent Peer-to-peer Storage Utility. In *Proc. of 18th ACM Symp. on Operating Systems Principles*, October 2001.

40. A. Serjantov. Anonymizing Censorship Resistant Systems. In *Proc. of the 1st International Peer To Peer Systems Workshop (IPTPS02)*, March 2002.

41. Andrei Serjantov and George Danezis. Towards an information theoretic metric for anonymity. In Roger Dingledine and Paul Syverson, editors, *Proc. of Privacy Enhancing Technologies Workshop (PET 2002)*. Springer-Verlag, LNCS 2482, April 2002.

42. Rob Sherwood, Bobby Bhattacharjee, and Aravind Srinivasan. P5: A protocol for scalable anonymous communication. In *Proc. of the 2002 IEEE Symposium on Security and Privacy*, May 2002.

43. A. Singh and L. Liu. Agyaat: Providing Mutually Anonymous Services over Structured P2P Networks. Technical Report GIT-CERCS-04-12, Georgia Inst. of Tech. CERCS, 2004.

44. I. Stoica, D. Adkins, S. Zhuang, S. Shenker, and S. Surana. Internet Indirection Infrastructure. In *Proc. of ACM SIGCOMM*, August 2002.

45. I. Stoica, R. Morris, D. Karger, M. F. Kaashoek, and H. Balakrishnan. Chord: a Scalable Peer-to-peer Lookup Service for Internet Applications. In *Proc. of ACM SIGCOMM*, August 2001.

46. M. Waldman, A. Rubin, and L. Cranor. Publius: A Robust, Tamper-evident, Censorship-resistant and Source-anonymous Web Publishing System. In *Proc. of the 9th USENIX Security Symposium*, pages 59–72, August 2000.

47. J. Wang, L. Lu, and A. Chien. Tolerating Denial-of-Service Attacks Using Overlay Networks - Impact of Topology. In *Proc. of ACM Workshop on Survivable and Self-Regenerative Systems*, October 2003.

48. B. Y. Zhao, J. D. Kubiatowicz, and A. D. Joseph. Tapestry: An Infrastructure for Fault-resilient Wide-area Location and Routing. Technical Report UCB-CSD-01-1141, U.C. Berkeley, April 2001.