Recipient Anonymity in a Structured Overlay *

Giuseppe Ciaccio
DISI, Università di Genova
via Dodecaneso 35
16146 Genova, Italy
E-mail: ciaccio@disi.unige.it

Abstract

An open problem in structured overlay networks is related to the anonymity to be provided to recipients, namely, those nodes who respond to request messages. Such a feature is of main concerns when designing censorship-resistant distributed applications. In this paper it is shown that, in a chordal ring overlay, by enforcing a degree of imprecision in each peer's routing table we obtain better recipient anonymity while keeping the length of routing paths within logarithmic length. A suitable metrics for recipient anonymity is established, based on the amount of resources an adversary needs in order to break anonymity of recipients in the overlay. In terms of this metrics, it is shown that imprecise routing tables make it impossible for a "small" coalition of malicious peers to correlate overlay addresses to hosts for censorship or auditing purposes.

1. Introduction and motivation

Peer to peer overlay networks have been receiving a lot of attention by the research community, as flexible and scalable low-level infrastructures for distributed applications of many kinds: network storage, naming, content publication, multicast/anycast, and communication security. They have also been proposed as general networking infrastructures [10, 24, 12, 11], 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 [18, 25, 16, 20, 15, 26] 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, often in a small number of hops. The most known example of a structured overlay is the chordal ring (Figure 1): N peers are arranged in a circle, and each can route messages towards the *successor* on 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 3.2).

On the other hand, unstructured overlays like Freenet [8] 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 scheme. Each host participating to the overlay is responsible for 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 recipient 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 [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; this has to do with the ability, or rather the difficulty, to correlate an overlay recipient address to the Internet address of the destination host. In order to break recipient anonymity, an adversary needs

*This research is supported by the Italian FIRB project *Webminds*.

to build a *map of the overlay*, relating overlay addresses to host addresses. An overlay provides recipient anonymity whenever it makes it difficult or impossible for an adversary to build such a map.

Let us suppose that messages in the overlay do not convey any explicit information about the sender or recipient Internet addresses (anonymity would be hopeless otherwise). All of the information relating overlay addresses to Internet identities is kept into the peers' routing tables. A common instance of such information is the set of pairs $\langle overlay\ address, IP\ address \rangle$ that forms each peer's knowledge of remote peers. In order to map a given part of the overlay, an adversary needs to harvest sufficiently many routing tables in the system, either by attacking many honest nodes or, more probably, by deploying a large enough coalition of malicious nodes. Another precious source of information for the adversary, however, is given by the shape of the overlay graph. In structured overlays, the presence of an arc between two nodes in the graph represents a mathematical relation between the overlay addresses of the two corresponding peers. For instance, arcs in a chordal ring represent distances on the overlay address space which conform to a geometric progression. Thus, in a structured overlay, the routing tables are more informative than in unstructured overlays. This improves the routing performance, but also helps the attacker wishing to map the overlay.

By contrast, the irregularity of unstructured overlays helps recipient anonymity. This is because unstructured networks lack of constraints between arcs in the graph and distances in the overlay address space, and this implies that the map between overlay address and Internet address of a recipient can be hidden to any other peer in the system, including neighbours. This is the reason for the success of unstructured networks as a support for censorship-resistant distributed storages [8], in which storing and retrieving are recipient roles. However, this feature is paid in terms of efficiency and availability.

Thus, there is an obvious interest to find trade offs between the efficiency of structured overlays and the privacy offered by unstructured ones, with the goal of improving recipient anonymity without seriously affecting routing performance. This indeed is the motivation of the work accounted here.

2. Gauging recipient anonymity

A truly censorship-resistant system may only rely on both recipient and sender anonymity. At a first glance these two features are similar, but they actually do not share much with each other and demand different techniques, because of the different roles played by senders in comparison with recipients. Indeed, a sender is an active entity that issues requests, whereas a recipient may or may not issue replies. The fundamental difference is that requests make use of the routing algorithm to propagate, whereas replies usually do not (replies can backtrack the same route of their corresponding requests [19], so that a return address is unneeded and sender anonymity is better preserved). Thus, although possibly observed, replies do not carry useful information related to the place where they have been issued. On the other side, requests issued by senders are labelled by destination addresses, but this information can only be useful to the adversary if the latter has a map of the overlay. In any case, the observable information related to the recipient does not contribute enough to draw a "probability space of recipients", and therefore the existing metrics for sender anonymity [5, 2, 3, 9, 23] cannot be recycled for evaluating recipient anonymity (unless the adversary has some source of information other than intercepted messages).

Communication indirection is a common way to obtain both kinds of anonymity in an overlay network, but, as shown in Section 1, this is only sufficient in unstructured overlays, and is paid in terms of efficiency and availability. Scarlata et al. [21] propose the use of a proxy who publishes a recipient address decoupled from the recipient identity. Recipient anonymity is preserved, and performance too, but the system is not censorship-resistant, as the proxy identity is exposed to the adversary. Serjantov [22] proposes a more sophisticated use of proxies, which can suffer from a similar kind of attack as well.

Thus, it seems that robust censorship resistance is impossible to achieve without sacrificing efficiency. Indirection on unstructured overlays offers no performance or availability guarantees, structured overlays can always be mapped by sufficiently many colluders, and proxies can always be attacked or controlled by sufficiently powerful adversaries.

On the other hand, an adversary striving to break censorship resistance, and recipient anonymity in particular, must deploy some resources: either malicious participants to a structured overlay, or servers for massive denial-of-service attacks.

Because of the above arguments, it is the author's opinion that recipient anonymity in a structured overlay, but also censorship resistance in a generic distributed system, could be gauged in terms of the *amount of resources an adversary must control in the system* in order to achieve his goal. By dividing such amount of resources by the total amount of resources

involved in the whole system, we get a measure that we call the *relative adversarial effort*. The larger such measure, the more resistant the system is to the adversarial actions. If N is the system size (e.g., the number of peers in an overlay) and $K(N)$ is the size of the adversarial coalition (e.g., the number of malicious peers), then the relative effort is $E(N) = K(N)/N$.

On the ground of the above definition we propose the term *pretty good* (recipient) *anonymity* to denote anonymity that cannot be broken with asymptotically small effort. The effort $E(N)$ is termed *asymptotically small*, or "small" for short, when it approaches 0 when N gets large. Our definition of pretty good anonymity, while recognizing that perfect anonymity is unachievable, formalizes the intuition that an anonymous system which could be defeated by a small minority of colluders cannot be regarded as a "good" one.

3. Imprecise routing for anonymity

3.1. Preliminary assumptions

Our discussion on recipient anonymity assumes an adversarial model that, following [9], we term "internal, local, and passive"; that is, the adversary controls 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. An "active" adversary poses much different security challenges, beyond our current scope. 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.

Moreover, we assume that the overlay protocol does not explicitly disclose the identity of any participant.

3.2. Generalized chordal rings

Although we believe that the entire work presented here could be adapted to any structured overlay, for practical purposes we had to choose a reference model of overlay network. In this paper we focus on the most widely used such model, namely, the aforementioned chordal ring. We present here a generalized version of the concepts originally introduced by Stoica et al. [25]. Similar concepts are found in every scalable overlay.

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. 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* [1] If peer P owns the address interval from A_l to A_u, and peer Q is the successor of P, then all addresses owned by Q are greater than A_u (modulo 2^k).

In the ideal scenario in which the map of addresses onto peers is complete, that is, no "hole" is left between each peer and its successor, a message issued towards address A can always reach the *recipient* (the peer which owns A in its own local address interval) by traversing the successor chain starting from the sender. In a realistic scenario, however, a faulty or disconnected peer could break the successor chain and also create a "hole", a discontinuity in the address space. A degree of redundancy (backup locations) can help, but the system must quickly seal the successor chain and restore the address hole, or redundancy would eventually degrade. To this end, an easy solution is to allow each peer to know a *successor list*, rather that just the 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).

Messages issued towards distant (in the overlay address space) recipients would take too long to reach their destinations through the successor chain. For distant recipients we need an alternate routing algorithm, so as to keep the routing path below an acceptable size. A common such mechanism is given by the *fingers*. A finger is an entry in the routing table of a peer, 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 remote address interval. Each peer maintains its own list of fingers. 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 the largest possible 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. Figure 1 illustrates this concept.

The routing algorithm takes advantage of fingers in a so-called "greedy" way (Figure 2). When a peer P

[1] 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 term "overlay network" to the more popular "distributed hash table".

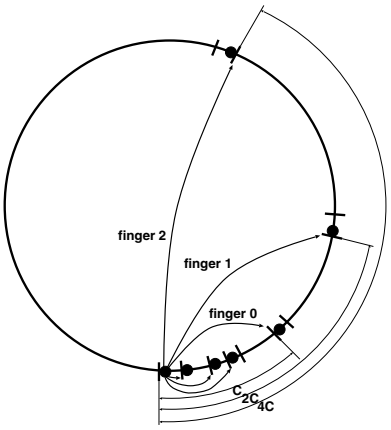

Figure 1. An instance of a chordal ring.

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.

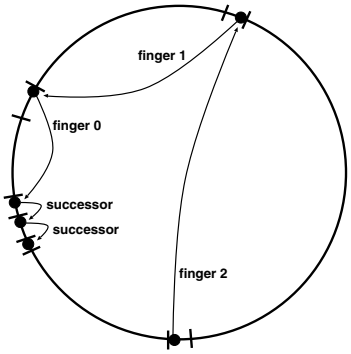

Figure 2. "Greedy" routing in a chordal ring. With N participants and complete routing tables, $O(\log(N))$ hops are sufficient.

In a chordal ring with complete finger tables conforming to a geometric distance rule, a total travel distance of D is covered in two phases, namely:

- $O(\log_2(D/C))$ finger hops until the residual distance falls shorter than the cutoff C; then
- an $O(1)$ number of short-range hops through the successor chain.

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, perhaps in a recursive style [25], 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.

However, the above (traditional, after Chord [25]) 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 map the address intervals of as many other peers. A malicious coalition counting $O(N/\log(N))$ peers, which corresponds to a "small" relative effort for the adversary, can thus map the entire overlay (Section 1).
2. When searching finger 0, peer P exposes its own address interval to the whole successor chain up to the finger; this is bad for recipient anonymity, especially if the successor list is long.

3.3. Improving anonymity with imprecise fingers

The two anonymity flaws outlined at the end of last Section are impossible to fix, because they are inherent to the traditional definition of fingers. The problem get harder if the system allows arbitrarily large routing tables on peers, because in such a case an adversary controlling a smaller number of peers could quickly map large portions of the overlay.

Similar arguments could be put against any structured overlay. Note, however, that we are implicitly

[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.

assuming that routing tables carry the exact overlay addresses of the remote peers they point to, that is, each peers knows the exact distance to any peer listed in its own routing table. Things might become better by relaxing this constraint.

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 idea 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 *imprecise*. Indeed, only exact information can contribute to map the overlay, so if its amount is kept small and fixed in each routing table, a coalition of peer wishing to map the overlay could not be "small".

Having imprecise entries in the routing table raises questions concerning convergence and efficiency of the routing algorithm. The routing paths could become longer, or be unsuccessful. In the following we propose a way to manage imprecise entries in the routing tables so that routing convergence is ensured, routing efficiency is preserved, but the routing tables themselves are of no help for an adversarial coalition to map the overlay.

Let F_P denote a secret random value generated by the generic peer P with uniform probability over $[0, C/4[$ (the cutoff distance C has been introduced in Section 3.2). F_P thus cannot exceed $C/4$.

Let us suppose peer P use the incremental procedure outlined in Section 3.2 to build its own finger table. The first step is to find the finger of magnitude 0. As already pointed out, finding finger 0 potentially exposes the address interval of P to the whole successor chain, up to finger 0, because of the need for all successors to compute their distance from P. To fix such information leak, P acts as follows: rather than sending (the lower bound of) its own address interval to the successor, P alters the information by subtracting F_P, then sends the altered value A.

Let Q be a generic peer in P's successor chain. After receiving an altered value A from its predecessor, it acts as follows:

1. Q estimates (the lower bound of) its successor's address interval, L; this is trivial, as the successor is adjacent in the address space.

2. Q computes the distance between $A - F_Q$ and L, namely, $L - A + F_Q$. This amounts to computing the distance between the original requestor, namely P, and Q's successor, incremented by the sum of the two random secrets F_P and F_Q; the computed distance is thus greater than the real one by an unknown quantity in $[0, C/2[$, since each random secret is in $[0, C/4[$.

3. If such distance is greater than C, then Q contacts the original requestor P and claims to be its finger 0.

4. Otherwise Q's successor is a better candidate, so Q forwards the value A to the successor.

Since the request emitted by P is initially altered by the random quantity F_P, P is not disclosing its own addressing information to the successors. When P is eventually contacted by a peer Q claiming to be its finger of magnitude 0, P can conclude that the distance of Q is the largest possible within an upper bound randomly distributed between $C - C/2 = C/2$ and C. This is a substantial departure from the traditional definition of fingers, according to which the distance of finger 0 is known with far greater precision (the largest distance not exceeding C).

The approximation introduced on finger 0 cumulates over fingers of greater magnitude. Using the incremental procedure, a finger of magnitude 1 is imprecise because its distance is close to a value between $2 \cdot (C/2) = C$ and $2C$, a finger of magnitude 2 is imprecise because its distance is close to a value between $4 \cdot (C/2) = 2C$ and $4C$, and so on. The generic finger m is located at a distance which is the largest possible not exceeding an unknown random value between $C \cdot 2^{m-1}$ and $C \cdot 2^m$ (Figure 3).

Such an amount of finger imprecision is a good device for recipient anonymity. The more distant a finger, the lesser the information about the overlay addresses it actually owns. No matter how large, no adversarial coalition can gather sufficiently exact information from finger tables.

3.4. Imprecise fingers and routing performance

The traditional "greedy" routing algorithm converges even with the imprecise fingers defined in Section 3.3. The number of hops increases, of course, but not dramatically.

Traversing a traditional, *exact* finger of magnitude m reduces the residual distance by no more than $C \cdot 2^m$, as we have seen in Section 3.2. With the imprecise fingers defined as in Section 3.3, however, a hop through a finger of magnitude m accomplishes a distance which is shorter or equal compared to the exact fingers of same magnitude, but never null. We already know from Stoica et al. [25] that the "greedy" routing with exact fingers converges. In our case the fingers fall shorter,

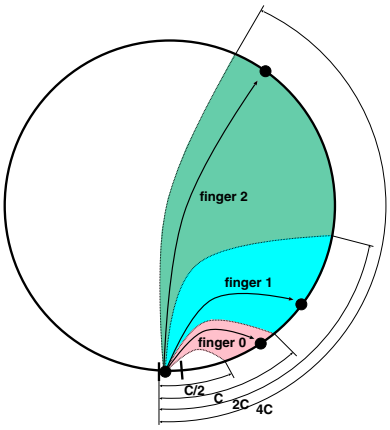

Figure 3. A chordal ring (successors omitted) in which the fingers are affected by an unknown random error. The average error increases proportionally with the distance of the peer. This way, no peer can infer much about the overlay addresses of other peers, and this improves recipient anonymity. Yet, $O(log(N))$ hops are still sufficient to route messages towards an arbitrary destination.

and the successor relation is preserved, so the same convergence arguments of Stoica et al. [25] apply.

At worst, hopping through an imprecise finger of magnitude m decreases the residual distance by no more than $C \cdot 2^{m-1}$. Therefore, to accomplish a distance decrease of $C \cdot 2^m$, it is necessary to traverse no more than two consecutive fingers of same magnitude m. We thus conclude that the use of imprecise fingers at most doubles the worst-case number of traversed fingers. In a system with complete finger tables, the routing paths thus remains $O(\log_2(D/C))$.

4. Simulation results

In order to validate the effectiveness of imprecise fingers, evaluate their impact on various aspects of system performance, and make a comparison with the traditional approach of exact fingers, we have built a simulator for a chordal ring over an address space made of 32-bit addresses.

In the simulated system with imprecise fingers, the cutoff distance C covers at least 20 bits of address and at least 10 consecutive peers on average; fingers are built according to the incremental procedure outlined in Section 3.2. By contrast, the reference system with exact fingers has same number of peers and same address space, but the cutoff distance is set to 1 as with traditional chordal rings; fingers are computed explicitly rather than incrementally, in order to avoid that higher magnitude fingers could be affected by cumulated imprecision arising from fingers of lower magnitude.

All the results have been collected by running the simulator over 500 sample rings of given size.

4.1. Finger imprecision and recipient anonymity

The standard deviation of finger distances is an indicator of the degree of imprecision of fingers themselves, provided the expectation of each finger distance is consistent with the overlay's distance rule (Section 3.2). If, in a population of sample rings, finger distances did not differ much from their mean values, any adversarial peer could deduce the address interval of each of its own fingers with good confidence, and thus a coalition could get a precise map of the overlay with high probability. This however would not be the case, was the standard deviation significant. So a high standard deviation of finger distances in a population of chordal rings is good news for recipient anonymity, and indeed the proposed algorithm for imprecise fingers (Section 3.3) aims at increasing such a statistics. We also expect greater standard deviation with greater finger magnitude, because of the incremental procedure by which imprecise fingers are built (Section 3.2).

Figure 4 represents the standard deviation of finger distances as a function of the magnitude, with different ring sizes. Both imprecise and exact fingers have been evaluated. In all cases, the expectation of each finger distance has been checked against the theoretical value as given by the distance rule (base-2 geometric progression starting with given cutoff distance) and no significant difference was found, so the standard deviation is a correct indicator of finger imprecision.

For an easier comparison among different ring sizes, and in order to help interpreting the results themselves, finger distances and their deviations have been normalized to the interval $[0, N]$ where N is the number of peers in the ring. This amounts to taking the average size of a peer's address interval as distance unit over the ring. With this normalization, a deviation of 100 means that the corresponding finger is likely to point up to 100 peers away from the correct position on the ring.

Due to the different sizes and cutoff distances of the involved chordal rings, fingers on one ring cannot be compared to fingers on another ring, not even if they have the same magnitude. This is the reason why the curves in Figure 4 are not aligned with one another.

Nevertheless, the Figure clearly shows that traditional fingers deviate from their expectations by less than the average size of a peer's address interval, that is, they are practically exact as expected. By contrast, imprecise fingers are affected by a sharply larger deviation (as expected as well).

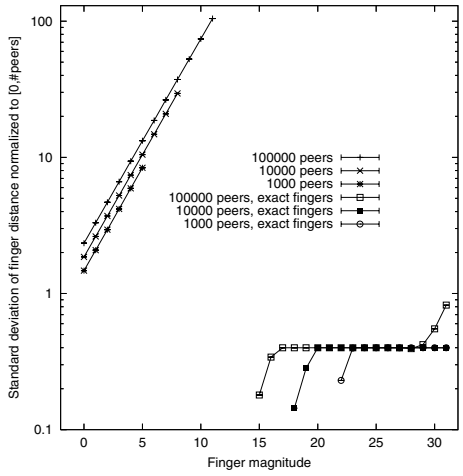

Figure 4. Finger imprecision as represented by the standard deviation of finger distances. The address intervals of the rings are normalized to the number of peers, so that the finger deviations are expressed in terms of (average) peers.

Assuming a deviation of 5 as a threshold between "sufficiently precise fingers" and "fingers too imprecise to be useful for mapping the overlay", we can deduce, for instance, that in a ring with 10000 peers equipped with imprecise fingers, only fingers of magnitude 0, 1, and 2 are useful for the adversary. If we also consider the immediate successor as a further piece of precise information, we deduce that each colluding peer can map at most 4 other peers in the overlay. We therefore conclude that the relative adversarial effort cannot be less than 1/4 in this case, that is, an adversarial coalition must count at least 25% of all participants in order to break anonymity of an arbitrary recipient. The same analysis in the case of exact fingers yields a relative adversarial effort of $1/\log_2(N)$, which roughly amounts to 0.075 with 10000 peers and 0.06 with 100000 peers. So, imprecise fingers do improve recipient anonymity quite a lot.

Another feature of imprecise fingers is that the amount of approximation, and the degree of recipient anonymity thereof, can be tuned by acting on the cutoff distance. This however would be paid with longer routing paths.

4.2. Impact on path lengths

As already envisaged in Section 3.4, the use of imprecise fingers can lead to a greater number of hops along the route from sender to recipient. Indeed, as apparent from Figure 5, chordal rings with imprecise fingers have quite longer routing paths compared to rings with traditional fingers. If we compare the curves of routing path length with imprecise vs. traditional fingers, and focus on the worst cases points, we see the difference never exceed the factor of 2 envisaged in Section 3.4. But, with traditional fingers, the number of hops has a great variability and is often much smaller that the worse cases; this does not happen with imprecise fingers, which exhibit a more regular profile. So the difference between the two families of curves appears more dramatic.

However, the two kinds of simulated chordal rings also have different cutoff distances, besides using different kinds of fingers. In order to isolate the impact of finger imprecision from the impact of cutoff distance, we ran the simulator on a "hybrid" overlay in which the cutoff distance is the same as in the system with imprecise fingers, but fingers are of traditional kind. The results with 1000 peers are reported in Figure 6. The curve yielded by the "hybrid" system has a highly variable shape, similar to the system with traditional fingers, but the span is much larger and reaches the curve of the system with imprecise fingers. We conclude that the two ingredients of our recipe for anonymity, namely, cutoff distance and imprecise fingers, have different costs: the cutoff distance impacts on the worse-case path length, whereas the finger imprecision smooths the profile of path length by driving the whole curve towards the worse case.

5. Related work

Imprecise routing information is at the core of unstructured overlays. With Freenet [8], 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. 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 follows a similar approach, with some more randomness [2].

There are very few attempts to improve anonymity in structured overlays. Achord [14] is an enhancement of Chord with anonymity features. Aiming at enforcing sender anonymity, Achord implements recursive-style routing [25] (because of the indirection) and forces each

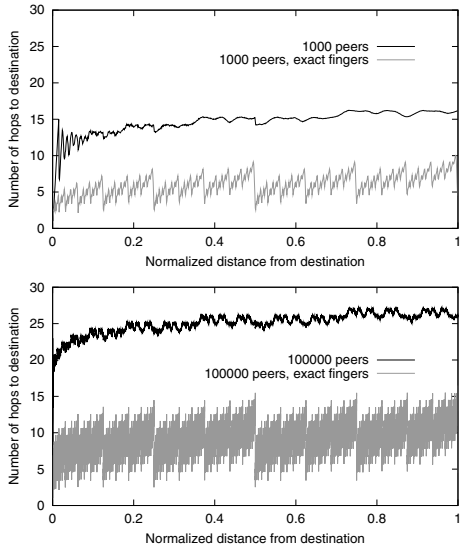

Figure 5. Average length of routing path from each sender to destination 0 in a chordal ring with given number of participants. Both imprecise and traditional fingers have been evaluated.

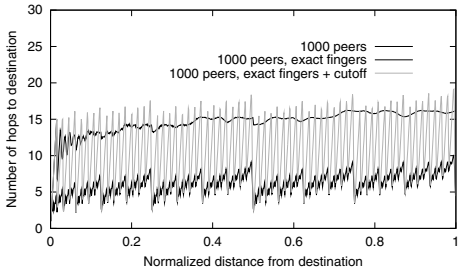

Figure 6. Average length of routing path from each sender to destination 0 in a chordal ring with 1000 peers. In addition to imprecise and traditional fingers, a "hybrid" kind of chordal ring has been evaluated, in which fingers are traditional but the cutoff distance is the same as in the system with imprecise fingers.

response to travel back to sender along the same route previously tracked by the corresponding request. In addition, Achord restricts each peer's ability to know about other peers: each peer is allowed to know the IP addresses of at most $k * \log(N)$ other peers, where k is the finger table size. The goal is similar to ours, namely, to enforce recipient anonymity by making it more difficult for an adversarial coalition to map the overlay. However the proposed solution is weak: the entire Achord ring could be mapped by a coalition of $N/k * \log(N) = O(N/\log(N))$ peers, which is "small" according to the definition given in Section 2 (the relative effort $E(N)$ amounts to $1/\log(N)$, which tends to 0 as N grows). Our work strived to overcome such a limitation.

Other studies [3, 17] only focus on sender anonymity in plain Chord, without considering recipient anonymity.

SkipNet [13], Skip Graphs [1], and Symphony [15], are overlay networks that make use of somehow randomized routing entries. However, randomized routing tables are not enough to obtain imprecise routing. In the aforementioned systems, indeed, each peer is autonomous in choosing the distance each routing entry is to point at. By contrast, imprecise routing requires that no peer has precise knowledge of such distances.

6. Conclusions and open issues

The most important result reported in this paper concerns recipient anonymity. After defining a metrics for this elusive feature, we have proposed the use of some randomization on long-range connections in structured overlay networks as a mean for obtaining better recipient anonymity without sacrificing the nice properties of structured overlays (provable routing convergence and, to some extent, performance). The study has been conducted on chordal rings, where our idea takes the form of what we call *imprecise fingers*. We however believe a similar study can be carried out on other topologies.

The positive impact of imprecise fingers on recipient anonymity, as envisaged by theory, has been confirmed by simulations.

Our result can be summarized by saying that recipients can be anonymous even in a chordal ring, yet the routing 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 efficient.

In practice, the impact of imprecise fingers on routing paths is not negligible. The number of hops increase quite much, although remaining within $O(\log(N))$. A substantial growth of routing path length raises performance and availability concerns: latency-sensitive applications might suffer, and longer routing paths are also more sensitive to the failure probability of individual peers. The cutoff distance of the chordal ring has been found as 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 availabil-

ity.

Another important issue is about the impact of our randomization mechanism on sender anonymity. The proposed solution would be impractical, should recipient anonymity be obtained at the expenses of sender anonymity. However, other simulations [7] have shown that the average sender anonymity decreases of a small amount, and the 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.

An unexplored security concern 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, however, appears weak as long as the adversary is able to use an IP domain of choice.

We have developed a working implementation of the principles accounted in this paper: NEBLO, a NEarly BLind Overlay [6] is a library and runtime system which allows to run a structured overlay network organized as a generalized chordal ring with imprecise fingers. NEBLO is released 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)*, Jan. 2003.

[2] K. Bennett and C. Grothoff. GAP: Practical Anonymous Networking. In *Proc. of PET*, 2003.

[3] N. Borisov and J. Waddle. Anonymity in Structured Peer-to-Peer Networks. gnunet.org/papers/borisov_waddle.pdf, Dec. 2003.

[4] D. Chaum. Untraceable electronic mail, return addresses, and digital pseudonyms. *Comm. of the ACM*, 4(2), Feb. 1981.

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

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

[7] G. Ciaccio. Evaluating Sender and Recipient Anonymity in a Structured Overlay. Technical Report DISI-TR-05-13, DISI, Università di Genova, Oct. 2005.

[8] I. Clarke, O. Sandberg, B. Wiley, and T. W. Hong. Freenet: A Distributed Anonymous Information Storage and Retrieval System. In *Proc. of PET*. Springer-Verlag, LNCS 2009, 2000.

[9] C. Díaz, S. Seys, J. Claessens, and B. Preneel. Towards measuring anonymity. In *Proc. of PET*. Springer-Verlag, LNCS 2482, 2002.

[10] J. Eriksson, M. Faloutsos, and S. Krishnamurthy. PeerNet: Pushing Peer-to-Peer Down the Stack. In *Proc. of IPTPS*, 2003.

[11] M. J. Freedman and R. Morris. Tarzan: A Peer-to-Peer Anonymizing Network Layer. In *Proc. of CCS*, 2002.

[12] I. Goldberg. *A Pseudonymous Communications Infrastructure for the Internet*. PhD thesis, UC Berkeley, Dec. 2000.

[13] 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)*, Mar. 2003.

[14] 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 IPTPS*, 2002.

[15] 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*, 2003.

[16] P. Maymounkov and D. Mazieres. Kademlia: A Peer-to-peer Information System Based on the XOR Metric. In *Proc. of IPTPS*, 2002.

[17] 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)*, 2004.

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

[19] M. Reed, P. Syverson, and D. Goldschlag. Anonymous Connections and Onion Routing. *IEEE Journal on Selected Areas in Communications, special issue on Copyright and Privacy Protection*, 1998.

[20] 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*, 2001.

[21] V. Scarlata, B. N. Levine, and C. Shields. Responder Anonymity and Anonymous Peer-to-Peer File Sharing. In *Proc. of ICNP*, 2001.

[22] A. Serjantov. Anonymizing Censorship Resistant Systems. In *Proc. of IPTPS*, 2002.

[23] A. Serjantov and G. Danezis. Towards an Information Theoretic Metric for Anonymity. In *Proc. of PET*. Springer-Verlag, LNCS 2482, 2002.

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

[25] 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*, 2001.

[26] 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, Apr. 2001.