The Impact of Routing Attacks on Pastry-Based P2P Online Social Networks

Felix A. Eichert[1], Markus Monhof[1], and Kalman Graffi[2]

[1] University of Paderborn, Germany
[2] Technology of Social Networks, Heinrich Heine University Düsseldorf, Germany
graffi@cs.uni-duesseldorf.de
http://tsn.hhu.de

Abstract. Peer-to-Peer (p2p) networks are common in several areas by now. Besides the well-known file sharing platforms, p2p overlays also emerge as a basis for decentralized social networks. In these, the overlay is used as robust storage for several kinds of social information. With gaining relevance, attackers might have an interest in tampering the functionality of the overlay. In this paper we investigate the routing attacks on the distributed hash table Pastry that we use as basis in our p2p social network LifeSocial. We determine through simulations the impact of routing attacks on the performance of the overlay.

1 Introduction

Online social networks (OSNs) are popular nowadays due to the ease of connecting billions of users and allowing them to interact through a set of communication options. Facebook, as most prominent example, connects around 1 billion users worldwide. A limitation of the current centralized approaches is given by the single operator running the social networking site. This operator is able to censor content and opinions, read private and confidential messages, market user data or be shut off in oppressive countries that want to reduce communication on specific topics, like during the Arab spring. Although the majority of the users remain unaware of the risks of using centralized OSNs and ignore the possibility of the communication being overheard, for some users in the world it is crucial or even vital to have the opportunity to communicate and organize with friends in a secure and anonymous way.

Peer-to-Peer (p2p)-based networks [21] can be made indestructible, as no component is considered crucial. Structured p2p overlays, especially distributed hash tables (DHTs) like Chord [27] or Pastry [24] offer a key-based routing [4] interface that allows to implement an efficient and fully retrievable simple storage. Using replication approaches like PAST [6] allow to keep content stored in the DHT remaining, even in the (expected) failure of the initial content hosting node. Distributed social networks propose to alleviate the security and censorship risk of centralized OSN sites.

In previous work, we have presented [16], [11] and [12], which builds a plugin-based p2p framework for hosting online social networks. LifeSocial aims at solving the main challenges in p2p based social networking: reliable and flexible data

L. Lopes et al. (Eds.): Euro-Par 2014 Workshops, Part I, LNCS 8805, pp. 347–358, 2014.

storage, security and controlled quality. Pastry [24] is used as basic substrate, PAST [6] is used for replication, SCRIBE [25] for an integrated publish / subscribe approach. For LifeSocial, we presented a practical access control approach in [15], which introduces a root of trust, deploys a secure key infrastructure and allows to cryptographically provide access control for single data elements. For the control of the quality of the overlay, we created a tree-based monitoring approach in [13] which is used to capture the current state of the running network. A further extension to a large-scale distributed control loop has been sketched in [17] and [18].

Security and especially the proper functionality of the overlay is vital for a p2p based online social network. As routing is solely performed by participating peers, routing might be tampered by these due to several reasons, such as free riding [26] or malicious aims. In this paper, we focus on routing attacks in Pastry [24] that are relevant in the context of social network. For that we introduce in Section 2 briefly in the underlying structure of the p2p overlay that we examine as well as the potential attacks in this overlay. We discuss approaches presented in literature that address routing attacks in Pastry in Section 3. We then present the evaluation setup and our simulation results on the impact of malicious nodes in Pastry in Section 4. Finally we conclude with an summary of our results and a look-out for future work in Section 5.

2 Background

In this section introduce shortly in Pastry, which has been implemented in FreePastry. FreePastry is popular in the academic community and has been used for several prototypes. After that we present attacks that can be conducted on a p2p network using the Pastry overlay. As stated in Section 1, we will limit the selection to routing attacks.

2.1 Pastry Overlay

P2P networks typically consist of an immense number of participants which often are referred to as nodes or peers. Those nodes have no hierarchical order and each have the same role in the network, so there is no centralized organization like in common client-server architectures. Such networks need certain algorithms to locate nodes or respectively nodes that are responsible for desired data and for routing messages to those nodes. Today most p2p networks are implemented as so called *overlay networks* which are using some sort of IP based network for communication

Pastry is an overlay for p2p networks with good performance and high reliability according to [24]. Each node in the Pastry overlay is assigned a unique identifier, the so-called "nodeId", that is uniformly distributed over a 128bit range. The state of a single node contains a *Leaf set*, a *Routing table* and a *Neighborhood set*. The *Routing table* contains a selection of nodes, and its corresponding IP addresses, which pose alternatives for routing of messages. For the

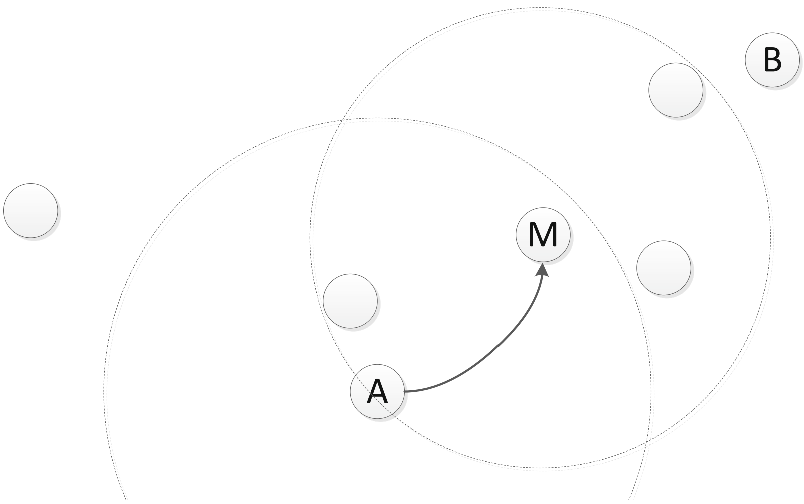


Fig. 1. Sample trajectory of a lookup-message initiated by the node A with the encounter of a malicious node using the *DROP* strategy. This malicious node M drops the message that comes from A instead of forwarding it to another node so the message will eventually reach the destination B.

purpose of routing the nodes in this table are ordered by the length of the shared prefix of their identifier with the identifier of node that owns this *Routing table*. The *Leaf set* contains nodes, whose identifier are closest to the owning node's identifier while the *Neighborhood set* contains nodes that are closest measured by a proximity metric. This proximity metric might for example be the number of IP routing hops or geographic distance. These sets allow the nodes to organize their connection between each other and to route messages throughout the network.

There are a few parameters that can be defined by the originator of the network which affect its overall performance and reliability of the network. One parameter is b defines the basis of how many bits are considered as one character in the nodeId. b is typically 4 or 16. The nodeId is represented as a String of characters and the prefixes are calculated based on characters.

The Pastry overlay implements a DHT. Therefore each node is responsible for holding a certain range of keys denoted by the distance of the nodes identifier to the next smaller one. To lookup a value of a certain key that is not hold by the requesting node, this node sends a request message to a node which has an identifier closer to the target node. This message is forwarded from one node to another until it reaches the node that holds the requested key. For a normal operating network, when N is the number of node in the network, the destination is reached in $\lceil \log_{2^b} N \rceil$ steps. In the case a certain node denies service, the node with the next higher identifier takes over the responsibility for that node's keys. The nodes use their states to find nodes that are close, determined by the proximity metric, to the destination node.

Since p2p networks consist of many widely distributed nodes and their participation is voluntarily, they pose heterogeneous unpredictable environments [1]. Single nodes may fail to forward messages or break down completely. In this case, near nodes notice these failures since they are no longer able to communicate with those nodes and propagate the break-down of them to other nodes in the network. This way the nodes react on turnover in the network and reorganize themselves to ensure the operation of the network.

2.2 Routing Attacks

In contrary to other types of attacks, routing attacks aim on compromising the ability of a p2p network to forward messages to designated nodes. That means that the attacker tries to stall the forwarding of messages so that they reach their recipients with a maximum delay or not at all.

Routing attacks focus on the forwarding of messages on behalf of other nodes. In this paper we focus on the attacks *Blocking any Request* (DROP), *Largest Distance First* (LDF) and *Approach at Minimum Pace* (AMP). The *DROP* strategy simply drops all messages to forward, but answers those directed at the current node. Thus we focus on forwarding routing attacks. In the case of a node joining, the node picks bootstrapping nodes until it is served. This strategy can easily be transferred to the Pastry overlay. It is illustrated in Figure 1. When it comes to countermeasures, this strategy, while it is easy to implement, might easily be detected and then be ignored. Detected nodes are not able to harm the network anymore.

The second attack strategy is termed *Largest Distance First* strategy (LDF). This strategy is illustrated in Figure 2. Here a malicious node forwards incoming lookup messages. The malicious approach is that the node does not forward the message in the direction of the targeted node like a ordinary node would do. This node estimates the node of those he knows whose identifier has the largest distance to the targeted node. This has the effect that the message moves away from instead of moving towards its destination. An implication of this attack is an increase of the routing hops a message needs to reach its destination. That increases the number of nodes that are involved in the processing of one message and therefore increases the load of those nodes. The increase of routing hops might additionally lead to longer physical distances that the message has to travel what results in longer transmission delays. While the message is not dropped and might eventually reach its destination, its processing is being stalled. This lowers the efficiency of the network.

The last attack strategy that we consider is the *Approach at Minimum Pace* strategy (AMP). As shown by Figure 3, a malicious node following this strategy, also forwards lookup messages. In contrary to *LDF*, in this strategy as next hop for a message to be forwarded a node is chosen which is closer to the destination of the message, such as the neighboring node of forwarding node. While the distance to the destination does not increase, it decreases at a minimal pace.

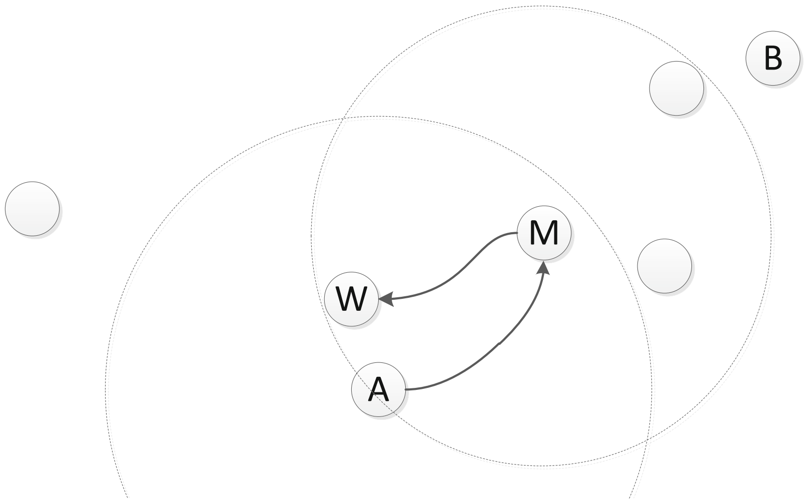


Fig. 2. Sample trajectory of a lookup message with the encounter of a malicious node using the *LDF* strategy. The message from *A* is being forwarded by the malicious node *M*. Instead of forwarding the message into the direction of the destination *B*, *M* forwards it to node *W*, which is the node with the largest distance to *B* that *M* knows.

This means that the direct effect of this strategy is not as striking as the *LDF*s. Waiving this part of significance might circumvent certain types of countermeasures. The node still forwards messages in the correct direction. Therefore, we suspect the detection of those malicious nodes to be much harder than e.g. those following the *LDF* strategy.

In both the *LDF*- and the *AMP* strategy we use the distance of the nodeIds. Pastry however with the Neighborhood set provides another term of distance, which relies on arbitrary proximity metrics. Because a node can not necessarily know the distance in terms of the Neighborhood set between to other nodes, it can not decide if a node has a higher or lower distance to the destination node of a message than itself. So it is not a reliable basis for the behavior of a malicious routing strategy.

3 Related Work

Security is a crucial issue of p2p networks and is in focus of several scientific publications. [5] indicates some open problems including security problems of today's p2p networks. They state that the p2p networks suffer from the unreliable nature of single peers. Secure Routing in organic networks has been addressed in [22], [23] and [14]. In these papers, networks are considered with multi-hop routing, in which the forwarding actions of the nodes cannot be directly observed. The authors propose to maintain a compressed history about the forwarded data by each peer, which might be occasionally verified publicly.

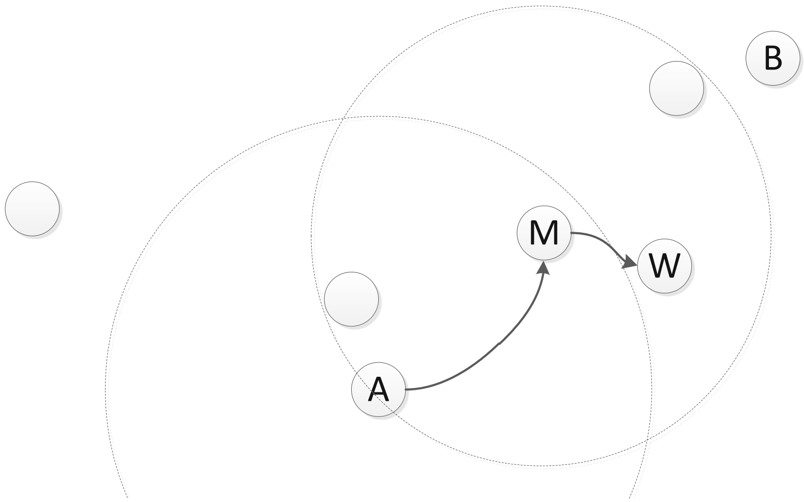


Fig. 3. Sample trajectory of a lookup message initiated by the node A. The message is being sent to the malicious node M that uses the AMP strategy in this example. M forwards the message into the direction of the destination node B. M chooses the node W, which is a bit closer to B, but as close to M as possible.

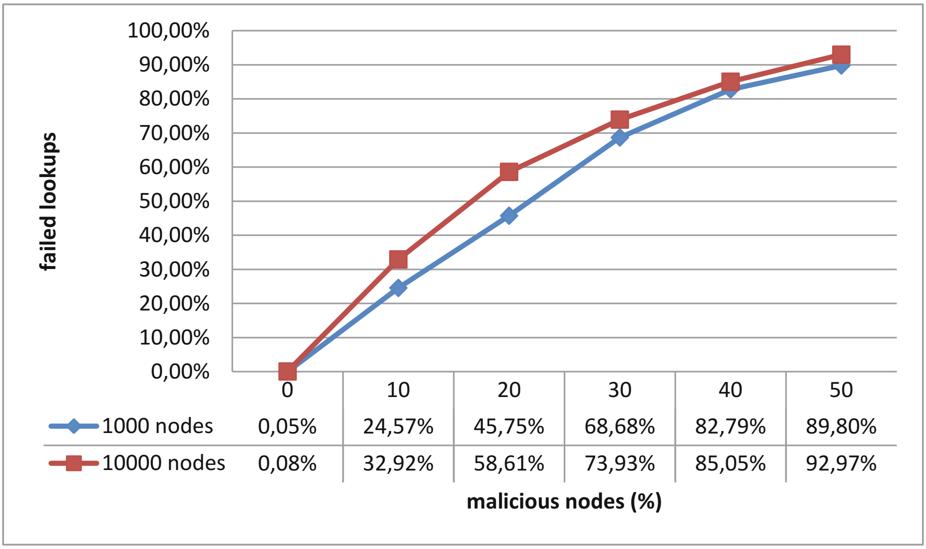
[3] proposes an approach for a reputation-based selection of peers that are used for routing operations. This enhances the reliability of the network by avoiding unreliable or faulty nodes. This countermeasure could dampen the effect of the routing attacks proposed in this paper. Mechanisms that ensure cooperation of peers in a p2p network are developed by [7]. These prevent peers from taking advantage of services of a network without contributing into the infrastructure. Some attacks are being presented by [2] which aim to prevent the attacked p2p network to fail to route messages properly. The paper also presents countermeasures against those attacks. An overview on further countermeasures is given in [28]. The authors provide a taxonomy on possible solutions, but do not give an evaluation on their effectiveness in various DHTs, such as Pastry.

A different approach of harming a p2p network, called Index Poisoning, is presented by [20]. It is stated that p2p file sharing systems are very vulnerable to this kind of attack. Index Poisoning harms the network by flooding it with wrong data that makes it difficult to retrieve correct data from it.

In [9], the authors researched the effect of one type of a routing attack on p2p networks. The paper defines a certain testbed for a simulation and gives the results of it. The scenario of the simulation is limited to a attack where malicious peers in a network do not forward any messages which equates to the *DROP* strategy that we simulate. Since this simulation is similar to ours, we will take the experimentation results from this paper in comparison with our results for the specific kind of attack.

Table 1. Malicious strategies and corresponding measurements of network performance/reliability

Strategy	Measurement
DROP	ratio of failed lookups
LDF	average number of hops per lookup
AMP	average number of hops per lookup



	0	10	20	30	40	50
1000 nodes	0,05%	24,57%	45,75%	68,68%	82,79%	89,80%
10000 nodes	0,08%	32,92%	58,61%	73,93%	85,05%	92,97%

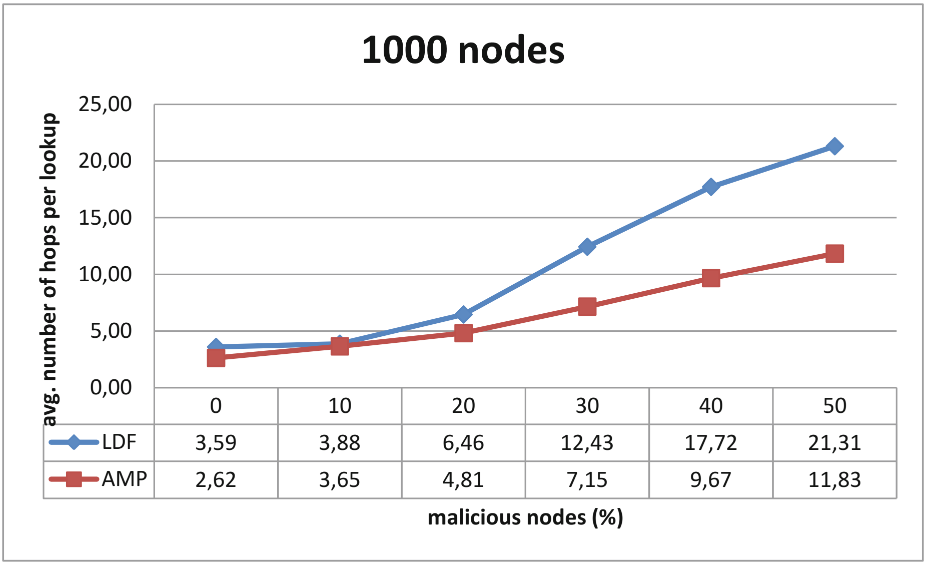
Fig. 4. Failed lookups in addiction to malicious nodes

4 Evaluation

For the evaluation of impact of the various attack strategies we used the the event based p2p network simulator PeerfactSim.KOM[1] [19], whose layered architecture allows easy implementation of different simulation scenarios. An introduction into the structure and functionality of PeerfactSim.KOM is given in [10] and [8]. PeerfactSim.KOM offers the advantage that it simulates multiple layers of a communication system, e.g. the transport layer, the network layer and the application layer. For each layer it contains multiple standard implementations, which can be interchanged, when needed.

For each network size we examine the influence of the three different kinds of attack strategies stated in Section 2.2 on the performance and robustness of the overlay. For the experiments we use the following scenario: We have Pastry of either 1000 or 10000 nodes and for both numbers of nodes we evaluate the effect of

[1] http://www.peerfact.org



	0	10	20	30	40	50
LDF	3,59	3,88	6,46	12,43	17,72	21,31
AMP	2,62	3,65	4,81	7,15	9,67	11,83

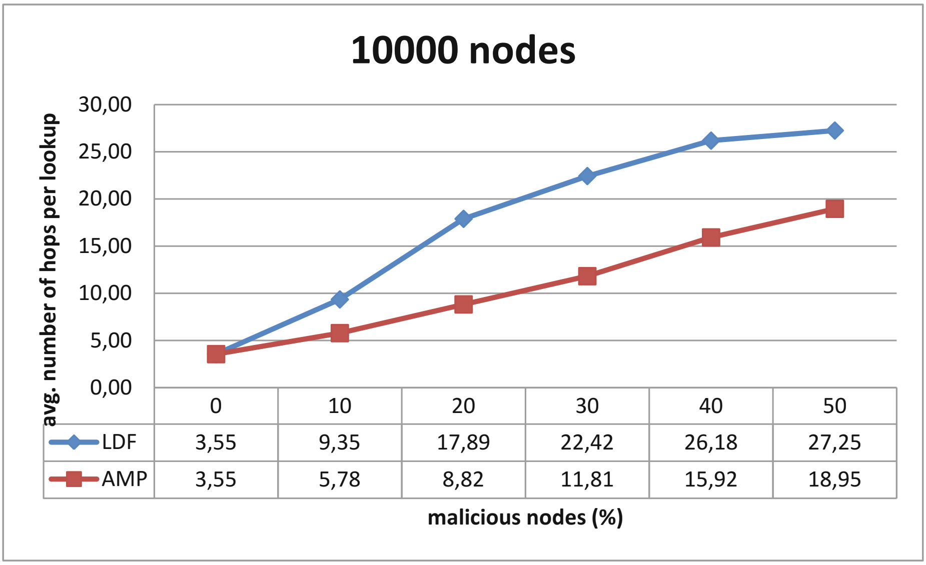
Fig. 5. Average number of hops per lookup in a network with 1000 nodes

0%, 10%, 20%, 30%, 40% and 50% malicious nodes in the network. A higher percentage of malicious nodes does not seem reasonable, because we expect that the Pastry will barely work with less than 50% of good nodes. The main metrics we focus on are the *average number of hops per lookup* and the *ratio of failed lookups* which give us insights on the performance and reliability of the overlay under attack. Table 1 shows the used strategies and corresponding measurements.

The process of the simulation is divided in two phases. In the *construction phase*, whose duration of 40 simulation minutes, all nodes act normally and build up the network. The second phase is the *operation phase* in which the nodes start lookup operations for keys in the DHT. We set its duration to 80 minutes in simulation time. For all three strategies we used a preferably realistic Net-Layer, which is capable of empirical determined latency, bandwidth and packet loss. The value of b, which is in a way the step size per hop, for the Pastry overlay is set to 4, as it is proposed in [24].

4.1 Experimental Results

First we present the results of the simulations with the *DROP* strategy. Figure 4 shows the failed lookups according to the percentage of malicious nodes. As we can see even a small percentage of malicious nodes leads to a alarming ratio of failed lookups. For larger networks the impact of this strategy is higher. Especially this is noticeable for a malicious node ratio between 10% and 30%. This is as with a larger amount of peers the average number of hops per lookups has to increase, so there is a higher chance that a node on the route is behaving maliciously. With 50% malicious nodes nearly all lookups fail.



	0	10	20	30	40	50
LDF	3,55	9,35	17,89	22,42	26,18	27,25
AMP	3,55	5,78	8,82	11,81	15,92	18,95

Fig. 6. Average number of hops per lookup in a network with 10000 nodes

In contrast to this, the impact of the *Approach at Minimum Pace*- and the *Largest Distance First* strategy on the failed lookups is negligible. Even for 10000 nodes and 50% malicious nodes the percentage of failed lookups lays around 1% for the *AMP* strategy, for the *LDF* strategy it is even less. This is due to the fact that these strategies only prolong the delivery of the messages in order to be harder to detect.

For these strategies we have a look at the *average number of hops per lookup*. The results for a network with 1000 nodes are illustrated in Figure 5. As we can see, the red line, which shows the impact of malicious nodes with the *AMP* strategy, has a nearly linear growth, the *LDF* strategy has a much higher influence on the average number of hops. This is due to the *AMP* strategy at least routes the lookup message in the right direction towards its destination, the *LDF* strategy however sends the message to that node (which is known by the current node) that has the largest distance, in terms of the node identifier, to the destination node. So with higher probability of malicious nodes, the probability that the next hop behaves badly is increased. In the worst case for the *LDF* strategy the lookup message can get stuck in an infinite loop, while the *AMP* strategy always routes the message to its destination. The results for 10000 nodes are shown in Figure 6 and they are conform to that for 1000 nodes. For both strategies the impact is just slightly higher.

Another result is, that the *average hops per lookup* are increasing in time, at least for the beginning of the *construction phase*. A graphical representation of this is given in Figure 7 for a network with 10000 from which 50% are malicious. The normal development is showed as a reference, too. Why that is the case might be interesting for future work.

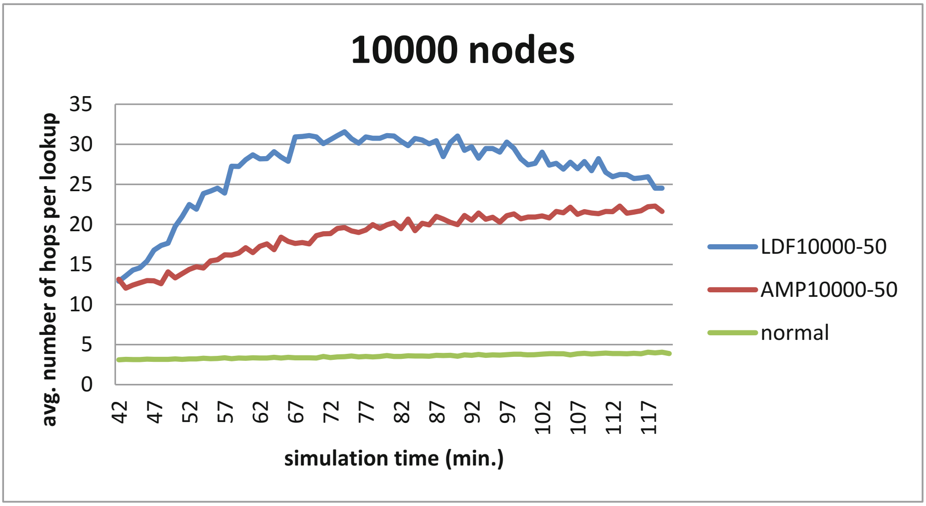


Fig. 7. Average number of hops per lookup in a network with 10000 nodes over simulation time

5 Conclusion

We simulated malicious nodes with three different strategies in Pastry overlay networks with 1000 and 10000 nodes. As a result we can state that Pastry is vulnerable for routing attacks. Nodes that simply drop all messages to forward effect large damage on the network. Solution are to be found to detect and isolate these nodes. Assuming that solutions for detecting misbehavior are found, malicious nodes might still try to modify the route length by sending the message only a bit forward or as far from the target away as they can. Through our simulations we show, that especially for big networks a small ratio of malicious nodes is sufficient to highly decrease the performance of the network. One aspect that could be interesting to be examined in the future could be why the number of hops is increasing over time or first in- and then decreasing. Also for these attacks, solutions need to be found and implemented. On the long term, secure p2p overlays, such as Pastry, are suitable to implement novel applications on top, such as online social networks.

References

1. Buragohain, C., Agrawal, D., Suri, S.: A Game Theoretic Framework for Incentives in P2P Systems. In: IEEE P2P 2003: Proc of the Int. Conf. on Peer-to-Peer Computing, pp. 48–56 (September 2003)
2. Castro, M., Druschel, P., Ganesh, A., Rowstron, A., Wallach, D.S.: Secure routing for structured peer-to-peer overlay networks. SIGOPS Oper. Syst. Rev. 36(SI), 299–314 (2002)

3. Cornelli, F., Damiani, E., di Vimercati, S., Paraboschi, S., Samarati, P.: Choosing reputable servents in a p2p network. In: ACM WWW 2002: Proc. of the Int. Conf. on World Wide Web, pp. 376–386 (2002)
4. Dabek, F., Zhao, B.Y., Druschel, P., Kubiatowicz, J., Stoica, I.: Towards a Common API for Structured Peer-to-Peer Overlays. In: Kaashoek, M.F., Stoica, I. (eds.) IPTPS 2003. LNCS, vol. 2735, Springer, Heidelberg (2003)
5. Daswani, N., Garcia-Molina, H., Yang, B.: Open problems in data-sharing peer-to-peer systems. In: Calvanese, D., Lenzerini, M., Motwani, R. (eds.) ICDT 2003. LNCS, vol. 2572, pp. 1–15. Springer, Heidelberg (2002)
6. Rowstron, A.I.T., Druschel, P.: Storage Management and Caching in PAST, A Large-scale, Persistent Peer-to-peer Storage Utility. In: IEEE HotOS 2001: Proc. of the Workshop on Hot Topics in Operating Systems (2001)
7. Feldman, M., Lai, K., Stoica, I., Chuang, J.: Robust Incentive Techniques for Peer-to-Peer Networks. In: ACM EC 2004: Proc. of the ACM Conf. on Electronic Commerce, pp. 102–111 (2004)
8. Feldotto, M., Graffi, K.: Comparative Evaluation of Peer-to-Peer Systems Using PeerfactSim.KOM. In: IEEE HPCS 2013: Proc. of the Int. Conf. on High Performance Computing and Simulation (2013)
9. Gottron, C., König, A., Steinmetz, R.: A Survey on Security in Mobile Peer-to-Peer Architectures - Overlay-Based vs. Underlay-Based Approaches. Future Internet 2(4), 505–532 (2010)
10. Graffi, K.: PeerfactSim.KOM: A P2P System Simulator Experiences and Lessons Learned. In: IEEE P2P 2011: Proc. of the Int. Conf. on Peer-to-Peer Computing (2011)
11. Graffi, K., Groß, C., Mukherjee, P., Kovacevic, A., Steinmetz, R.: LifeSocial.KOM: A P2P-based Platform for Secure Online Social Networks. In: IEEE P2P 2010: Proceedings of the International Conference on Peer-to-Peer Computing (2010)
12. Graffi, K., Groß, C., Stingl, D., Hartung, D., Kovacevic, A., Steinmetz, R.: LifeSocial.KOM: A Secure and P2P-based Solution for Online Social Networks. In: Proc. of IEEE CCNC (2011)
13. Graffi, K., Kovacevic, A., Xiao, S., Steinmetz, R.: SkyEye.KOM: An Information Management Over-Overlay for Getting the Oracle View on Structured P2P Systems. In: IEEE ICPADS 2008: Proc. of the Int. Conf. on Parallel and Distributed Systems, IEEE (2008)
14. Graffi, K., Mogre, P.S., Hollick, M., Steinmetz, R.: Detection of Colluding Misbehaving Nodes in Mobile Ad Hoc and Wireless Mesh Networks. In: IEEE Global Telecommunications Conference (GLOBECOM). IEEE (2007)
15. Graffi, K., Mukherjee, P., Menges, B., Hartung, D., Kovacevic, A., Steinmetz, R.: Practical Security in P2P-based Social Networks. In: IEEE LCN 2009: Proceedings of the International Conference on Local Computer Networks (2009)
16. Graffi, K., Podrajanski, S., Mukherjee, P., Kovacevic, A., Steinmetz, R.: A Distributed Platform for Multimedia Communities. In: IEEE ISM 2008: Proceedings of the International Symposium on Multimedia (2008)
17. Graffi, K., Stingl, D., Rückert, J., others: Monitoring and Management of Structured Peer-to-Peer Systems. In: IEEE P2P 2009: Proceedings of the International Conference on Peer-to-Peer Computing (2009)
18. Klerx, T., Graffi, K.: Bootstrapping Skynet: Calibration and Autonomic Self-Control of Structured Peer-to-Peer Networks. In: IEEE P2P 2013: Proceedings of the International Conference on Peer-to-Peer Computing (2013)

19. Kovacevic, A., Kaune, S., Heckel, H., Mink, A., Graffi, K., Heckmann, O., Steinmetz, R.: PeerfactSim.KOM - A Simulator for Large-Scale Peer-to-Peer Networks. Tech. Rep. Tr-2006-06, Technische Universität Darmstadt, Germany (2006)
20. Liang, J., Naoumov, N., Ross, K.W.: The Index Poisoning Attack in P2P File Sharing Systems. In: IEEE INFOCOM 2006: Proceedings of the International Conference on Computer Communications, vol. 6 (2006)
21. Liebau, N., Pussep, K., Graffi, K., Kaune, S., Jahn, E., Beyer, A., Steinmetz, R.: The Impact Of The P2P Paradigm. In: AMCIS 2007: Proceedings of Americas Conference on Information Systems (2007)
22. Mogre, P., Graffi, K., Hollick, M., Steinmetz, R.: A Security Framework for Wireless Mesh Networks. Wireless Communications and Mobile Computing Special Issue: Architectures and Protocols for Wireless Mesh, Ad Hoc, and Sensor Networks (2010)
23. Mogre, P.S., Graffi, K., Hollick, M., Steinmetz, R.: AntSec, WatchAnt, and AntRep: Innovative Security Mechanisms for Wireless Mesh Networks. In: IEEE LCN 2007: Proc. of the Annual Conf. on Local Computer Networks. IEEE (2007)
24. Rowstron, A., Druschel, P.: Pastry: Scalable, Decentralized Object Location, and Routing for Large-Scale Peer-to-Peer Systems. In: Guerraoui, R. (ed.) Middleware 2001. LNCS, vol. 2218, p. 329. Springer, Heidelberg (2001)
25. Rowstron, A., Kermarrec, A.-M., Druschel, P.: SCRIBE: The Design of a Large-Scale Event Notification Infrastructure. In: Crowcroft, J., Hofmann, M. (eds.) NGC 2001. LNCS, vol. 2233, p. 30. Springer, Heidelberg (2001)
26. Schoder, D., Fischbach, K.: Peer-to-Peer Prospects. Communications of the ACM 46(2), 27–29 (2003)
27. Stoica, I., Morris, R., Karger, D., Kaashoek, M.F., Balakrishnan, H.: Chord: A Scalable Peer-to-Peer Lookup Service for Internet Applications. In: SIGCOMM 2001: Proceedings of the International Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications, ACM (2001)
28. Villanueva, R., Villamil, M.-D.-P., Arnedo, M.: Secure routing strategies in dht-based systems. In: Globe 2010: Int. Conf. on Data Management in Grid and Peer-to-Peer Systmes (Globe), pp. 62–74 (2010)