Privacy Preserving Distributed Hash Tables - Reality and Future

Gonçalo Pestana (goncalo@hashmatter.com)

Dec 2018

DRAFT: This document is work in progress. Please send your comments, suggestions and corrections to gpestana@hashmatter.com or join the conversation at https://github.com/gpestana/notes/issues/8

Abstract

Distributed Hash Tables (DHT) are overlay networks that enable peers to store and request data in a peer-to-peer (P2P) network. Peers are responsible for storing data and for participating in the lookup and routing protocol. A DHT does not require peers to have a complete view of the network and does not rely on central authorities, resulting on resilient, scalable a decentralized networks. DHTs are an important building block for the decentralized web and many P2P systems. However, the decentralized nature of DHTs introduces privacy vulnerabilities to the services built on top of it: in naive implementations of decentralized networks users potentially disclose private data and metadata to many untrusted parties. In this paper we review the literature of mechanisms and protocols to achieve privacy preserving DHTs and their vulnerabilities. Finally we outline open problems and directions for future research.

Introduction

Distributed Hash Tables (DHT) are an important building block for the decentralized web and P2P systems. DHTs are overlay networks that allow peers to store and lookup information stored in a decentralized network, where participant nodes are responsible for routing requests and storing data.

As a result of its P2P nature, naive design and implementation of DHTs leak user behavior information to other participants. Information about "who is participating in the network?"; "who is requesting a particular set of data?"; and "who is storing and providing a particular set of data?" and "what are the relations between peer in the network?" can be collected by adversaries with relatively low resources. Data leaked by DHTs can be used to target and censor individuals, communities and services.

Privacy preserving networks are designed and implemented to achieve user privacy as a goal. We argue that 1) the comeback of decentralized networks is a crucial opportunity for building privacy preserving networks from the ground-up and; 2) that DHTs are an important primitive in the decentralization landscape and thus implementations of the DHTs should take user privacy in consideration.

Given its decentralized nature, naive DHTs designs are vulnerable to multiple attacks that can directly affect privacy (Wang, Mittal, and Borisov 2010), (O'Donnell and Vaikuntanathan 2004). The underlying reasons for the privacy vulnerabilities in DHTs are twofold. First, consistent hashing conceptually maps data blocks to a peer (or set of peers). Although this property allows the network to scale its capacity, it also constitutes a vector for metadata leaks. Secondly, peers need to cooperate with each other openly through a key-based routing protocol in order to resolve data lookups, since each peer have only a partial view of the network. Scalability and peer cooperation comes at a privacy price. Since the requests issued by the lookup initiator converge towards the lookup destination over time, a passive adversary can link peers with a particular content request. Intuitively, this means that potentially a passive adversary participating in the network can easily collect enough information disclosing what data is being requested and by whom.

This work aims at building the foundations to research and implement privacy preserving DHTs. In the first section, we review protocols and techniques to design and build secure and privacy preserving DHTs and their current implementations. After, we outline vulnerabilities that researchers have found in those mechanisms. We finally discuss future directions for the design and implementation of privacy preserving DHTs taking into considerations how it impacts properties such as decentralization, scalability and performance of the network.

Privacy preserving DHTs: protocols and mechanisms

This section covers the literature on privacy preserving P2P routing mechanisms, with special focus on DHTs.

Octopus (Wang and Borisov 2012) is a DHT lookup mechanism that aims at providing security and anonymous lookups. The threat model assumes partial adversaries that control up to a fraction of the network (max. 20%) and discards the possibility of global adversaries with the ability to monitor all the traffic in the network. It shows that anonymity in DHTs as defined by (Pfitzmann and Hansen 2009) can only be achieved if the DHT is secure against active attacks. Otherwise, an adversary can use its influence in the network to perform lookup bias attack, lookup misdirection attack and finger pollution attack to deanonymize a lookup initiator. Thus, the authors start by describing how active attacks can be identified and flagged. Since all the active attacks are performed by manipulating adversary finger tables, the strategy for flagging honest adversaries in the network consists of honest peers to check the correctness of other node's routing tables. This checking mechanism is coupled with a reputation system that maintains a list of adversaries and the proof they have cheated. For this purpose, Octopus defines an additional overlay network for peers to check the correctness of its neighbor's routing table. To punish malicious peers, a Certificate Authority (CA) is used to revoke rights for identified malicious nodes to participate in the network. Thus, active adversaries are flagged and not allowed to participate in the network. Octopus lookup protocol ensures anonymity by using onion routing routing protocol and using several anonymous paths to send lookup requests without revealing who is the lookup initiator. Moreover, it splits queries into multiple paths and introduces dummy queries to achieve lookup anonymity. The authors proceed to demonstrate that Octopus has reasonable lookup latency and communication overhead.

Nisan (Panchenko, Richter, and Rache 2009) is a protocol for peers to select uniformly a set of nodes in a P2P network while having limited knowledge of the network. It also aims at providing a privacy layer so that no other peer in the network can guess which set of nodes were selected by a particular peer. Nisan considers active and passive attacks. Active attacks are performed by peers by … Passive attacks consist of… As part defense against active adversaries which may drop or bias requests, In order to protect against active attacks, Nisan proposes an improved redundancy lookup mechanism - *aggregated greedy search* The goal is to make sure that adversary nodes are not able to influence the lookup, while ensuring that the redundant lookup paths do not converge. However, the authors showed that if the requested peers know what is the lookup ID, adversary peers can perform eclipse attacks which defeat the aggregates greedy search. To avoid this, the queried peer must not know what is the ID the lookup initiator is interested in. Thus, the lookup initiator requests the whole finger table from the queried peer, rather than the peer closest to a particular ID. Both mechanisms, however, do not protect the lookup initiator against passive attacks. As a matter of fact, redundant lookups increase the attack surface for passive attackers, since there is more information flowing in the network about the lookup initiator goals and simple packet correlation can easily ...

Torsk (McLachlan et al. 2009) is a structured peer-to-peer low-latency anonymity protocol ..

Bifrost (Kondo et al. 2009) is an anonymous communication protocol that uses Chord DHT as an overlay network for managing node participation and routing tables in the network. The anonymity requirements set by the authors are sender anonymity, receiver anonymity and data flow untraceability. Bifrost maintains a node management layer (NML) and an anonymous routing layer (ARL). The NML is responsible for keeping meta information about the overlay network, more specifically network information about the nodes connected to in the Chord overlay. The ARL is responsible for the actual communication which is similar to onion routing. It including route construction and message encryption and depends on a Public Key Server infrastructure. Bifrost encrypts the onion packets so that the final destination is in the middle of the onion path, and subsequent relayed messages in the circuit are dummy messages.

(Backes et al. 2011) presents a solution for hiding who is relaying key messages in the network, which is an approach often neglected in the literature (usually only provider and sender are protected).

Privacy vulnerabilities

This section describes the privacy vulnerabilities in DHTs and their protocols.

(Evans, GauthierDickey, and Grothoff 2007) Freenet "darknet" (Clarke et al., n.d.) implements a decentralized routing algorithm under a restricted-routing topology (friend-2-friend). The routing algorithm relies on location swapping for improving lookup performance of such topologies to `Olog(n)`, under the conditions that trusted node connections form a small-world network. Evans, GauthierDickey, and Grothoff (2007)] shows how relatively low-resourced attackers can subvert the network topology and content availability by lying about their links during the location swapping protocol.

(Mittal and Borisov 2008)focus on showing

(Wang, Mittal, and Borisov 2010) proposes active attacks on secret buddy mechanism which compromises user anonymity.

(Wang, Mittal, and Borisov 2010) has shown that information leaks in the NISAN lookup lead to a significant reduction in user anonymity.

(O'Donnell and Vaikuntanathan 2004) shows that even though Chord DHT does not have anonymity as main design goal, it does offer some degree of anonymity against passive adversaries and under specific network conditions.

Open questions and future research

Decentralized reputation and trust mechanisms: Some of the literature reviewed rely on reputation systems and trust mechanisms, coupled with central authority infrastructure for flagging network adversaries (Wang and Borisov 2012). A potential research direction is to study hsow to maintain a decentralized and secure data structure storing trust information. Upon detecting adversary attacks, a peer should be able to update the data structure with a proof of the attack (e.g. signed poisoned routing table by an active adversary) and any peer should have easy access to the relevant data structure to check if a peers is trusted in the network. The data structure does not need a network consensus a can be split across the network. The trust information data store should also have privacy preserving characteristics, similarly to ClaimChain (Kulynych et al. 2017).

Incentives: Privacy has costs and peers who reliably offer resources to the network should be rewarded in order for the net incentives to accrue to positive network outcomes.

Scalable and secure DHT: malicious peers that control a fraction of the network are able to perform attacks that compromise privacy even when privacy mechanisms are in place.

Measure anonymity and decentralization: *reference to 'why I'm not an entropist'* and '15 years systematization': 1) it must be possible to measure anonymity - and privacy - of a network and a single user at a given time; 2) create a framework to categorize a) anonymity and b) decentralization level. Only after, we can reason about private DHTs across the board.

- **Implementation and benchmarks in production**:

Conclusion

References

Backes, Michael, Ian Goldberg, Aniket Kate, and Tomas Toft. 2011. "Adding Query Privacy to Robust Dhts." *CoRR* abs/1107.1072. http://arxiv.org/abs/1107.1072.

Clarke, Ian, Oskar S, Matthew Tosel, and Vilhelm Verendel. n.d. "Private Communication Through a Network of Trusted Connections: The Dark Freenet."

Evans, N. S., C. GauthierDickey, and C. Grothoff. 2007. "Routing in the Dark: Pitch Black." In *Twenty-Third Annual Computer Security Applications Conference (Acsac 2007)*, 305–14. https://doi.org/10.1109/ACSAC.2007.7.

Kondo, M., S. Saito, K. Ishiguro, H. Tanaka, and H. Matsuo. 2009. "Bifrost : A Novel Anonymous Communication System with Dht." In *2009 International Conference on Parallel and Distributed Computing, Applications and Technologies*, 324–29. https://doi.org/10.1109/PDCAT.2009.35.

Kulynych, Bogdan, Marios Isaakidis, Carmela Troncoso, and George Danezis. 2017. "ClaimChain: Decentralized Public Key Infrastructure." *CoRR* abs/1707.06279. http://arxiv.org/abs/1707.06279.

McLachlan, Jon, Andrew Tran, Nicholas Hopper, and Yongdae Kim. 2009. "Scalable Onion Routing with Torsk." In *Proceedings of the 16th Acm Conference on Computer and Communications Security*, 590–99. CCS '09. New York, NY, USA: ACM. https://doi.org/10.1145/1653662.1653733.

Mittal, Prateek, and Nikita Borisov. 2008. "Information Leaks in Structured Peer-to-Peer Anonymous Communication Systems." Edited by Paul Syverson, Somesh Jha, and Xiaolan Zhang. Alexandria, Virginia, USA: ACM Press; ACM Press. https://doi.org/10.1145/1455770.1455805.

O'Donnell, C. W., and V. Vaikuntanathan. 2004. "Information Leak in the Chord Lookup Protocol." In *Proceedings. Fourth International Conference on Peer-to-Peer Computing, 2004. Proceedings.*, 28–35. https://doi.org/10.1109/PTP.2004.1334928.

Panchenko, Andriy, Stefan Richter, and Arne Rache. 2009. "NISAN: Network Information Service for Anonymization Networks." In *Proceedings of the 16th Acm Conference on Computer and Communications Security*, 141–50. CCS '09. New York, NY, USA: ACM. https://doi.org/10.1145/1653662.1653681.

Pfitzmann, Andreas, and Marit Hansen. 2009. "A Terminology for Talking About Privacy by Data Minimization: Anonymity, Unlinkability, Undetectability, Unobservability, Pseudonymity, and Identity Management."

Wang, Qiyan, and Nikita Borisov. 2012. "Octopus: A Secure and Anonymous DHT Lookup." *CoRR* abs/1203.2668. http://arxiv.org/abs/1203.2668.

Wang, Qiyan, Prateek Mittal, and Nikita Borisov. 2010. "In Search of an Anonymous and Secure Lookup: Attacks on Structured Peer-to-Peer Anonymous Communication Systems." In *Proceedings of the 17th Acm Conference on Computer and Communications Security*, 308–18. CCS '10. New York, NY, USA: ACM. https://doi.org/10.1145/1866307.1866343.