Privacy Preserving Distributed Hash Table [DRAFT]

Gonçalo Pestana (goncalo@hashmatter.com)

Dec 2018

Abstract

Distributed Hash Tables (DHT) are overlay networks that allow peers to store and lookup data in a decentralized network. Peers are responsible for routing requests and storing data without the need for of external control. DHTs are an important building block for the decentralized web and many P2P systems. However, its decentralized nature introduces privacy vulnerabilities to the services built on top of it. We discuss privacy threat model of DHTs, define the requirements for a privacy preserving DHT and review the literature for mechanisms and protocols to achieve it. We also 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. Questions such as "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?" can be answered by anyone with relatively low resources. The data leaked by DHTs can be used to target and censor individuals, communities and services.

Privacy preserving networks are designed and implemented in such ways that user privacy is protected against targeted and mass surveillance and censorship. 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.

This work aims at building the foundations to research and implement privacy preserving DHTs. We review protocols and techniques used to design and build secure and privacy preserving DHTs and their current implementations. We also discuss future directions for design and implementation of privacy preserving DHTs, taking into considerations how it impacts properties such as decentralization, scalability and performance of the network. We conclude that ...

The remaining paper is structured as follows. First, we review literature on privacy preserving protocols and techniques on P2P networks. Secondly, we describe how a general purpose DHT works and outline the most popular protocols used nowadays and their respective applications. Thirdly, we describe the threat model of DHTs from a security and privacy perspectives. We then proceed with by reviewing protocols and techniques that aim at preserving complete or a subset of privacy properties in DHTs. Finally, we outline open questions and future research directions.

Distributed Hash Tables

- any peer can participate in the network
- recursive vs iterative routing protocol
- constrained overlay vs non-constrained

Applications

In this section we outline current and potential applications of DHTs and motivate where privacy is important in each case.

- **Distributed file systems and content caches** (e.g. IPFS)
- **Private directory** (e.g. Tor)
- **Distributed service discovery** (e.g. Ethereum)

Problem statement: privacy vulnerabilities

Given its decentralized nature, naive DHTs are vulnerable to multiple attacks that can directly affect privacy. The underlying reason for the privacy vulnerabilities in DHTs is that peers need to cooperate with each other by routing requests within the network, since each peer has only a partial view of the network. Anyone can passively track the requests done by a particular peer. Since the requests issued by the lookup initiator converge towards the lookup destination over time, a passive adversary can link lookup initiator with a particular content request. Intuitively, this means that potentially every peer participating in the network can easily learn which content a particular peer is requesting. According to (Pfitzmann and Hansen 2009) and our threat model (Section X), such system design does not provide user privacy.

Literature Review

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. (Wang and Borisov 2012) remarks that anonymity in DHTs as defined by (Pfitzmann and Hansen 2009) can only be achieved if the DHT is secure against active attacks. An adversary can use its influence in the network to perform lookup bias attack, lookup misdirection attack and finger pollution attack to de-anonymize a lookup initiator. Thus, (Wang and Borisov 2012) starts by describing how active attacks can be identified and flagged. Since all the active attacks are performed by manipulating adversary finger tables, the strategy to flag those attacks consists of honest nodes checking the correctness of other node's routing tables coupled to a reputation system. For this purpose, Octopus defines another 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. Once the active attacks are identified and flagged, Octopus ensures anonymity by using an anonymous path to send lookup requests without revealing who is the initiator. Moreover, it splits queries into multiple paths and introduces dummy queries to achieve lookup anonymity. (Wang and Borisov 2012) showed 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 …

Bifrost (Kondo et al. 2009) is an anonymous communication protocol that uses a DHT (Chord) as an overlay network for managing node participation and peer routing tables. The anonymity requirements 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.

Threat model

We assume that an attacker can control up to a fraction `f` of the network nodes, and suggest that `f` should be 0.2. We assume the worst case in which the adversary nodes are can passively and actively attack the network and may cooperate. Adversaries that can control more than a relatively small fraction of the network and monitor the whole network activity are not considered. (Mittal and Borisov 2009),(Wang and Borisov 2012),(Panchenko, Richter, and Rache 2009) claim that such adversary is unlikely to exist in large P2P networks.

We consider a DHT to be fully privacy preserving if a lookup initiator is anonymous. Based on (Pfitzmann and Hansen 2009) definition, a lookup initiator peer is anonymous if honest or adversary peers can not tell from the set of possible lookup initiators - all nodes in the DHT - who initiated a particular lookup. Additionally, given a lookup initiator, a network peer should not be able to determine what is the lookup target.

We also consider message unlinkability (Pfitzmann and Hansen 2009) as a requirement for a privacy preserving DHT. An adversary must not be able to link multiple messages as belonging to an anonymous initiator.

Attacks on DHT

Most of the privacy preserving mechanisms and protocols [shadowwalker], (Wang and Borisov 2012), … need ways to enforce security in DHT before providing privacy. According to (Wang and Borisov 2012) DHTs are vulnerable to the following attacks:

- *Lookup Bias Attack:*:
- *Lookup Misdirection Attack*: malicious nodes may provide manipulated finger tables so that the lookup initiator only uses malicious nodes as part of the lookup path. This attack potentially deanonymises the lookup initiator since the attacker acquires a lot of information about the lookup along the path.
- *Finger Pollution Attack*: when honest nodes attempts to update its finger table by requesting other peer's routing tables, malicious node may attempt to pollute its routing table with malicious nodes in order to perform lookup bias

attacks and lookup misdirection attacks.

Thus, we conclude that in order to achieve a privacy preserving DHTs, we need to ensure that malicious nodes cannot deceit honest nodes with manipulated finger tables. This is a hard precondition to our goal of achieving privacy preserving DHTs.

Approaches for privacy preserving DHTs

Friend to friend restricted network topology

as in Freenet "darknet"

In-DHT garlic routing

as in I2P

In-DHT onion routing

Onion routing [(Goldschlag, Reed, and Syverson 1999), (Reed, Syverson, and Goldschlag 2006)] has been researched and used (Dingledine, Mathewson, and Syverson 2004) to protect network level metadata leakage

Open questions and future research

- **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.
- **Reputation on a DHT**: How to maintain reputation (e.g. CA) without central authorities.
- **Measure anonymity**: *reference to 'why I'm not an entropist'*
- **Implementation and benchmarks in production**:

Conclusion

References

Dingledine, Roger, Nick Mathewson, and Paul Syverson. 2004. "Tor: The Second-Generation Onion Router." In *Proceedings of the 13th Conference on Usenix Security Symposium - Volume 13*, 21–21. SSYM'04. Berkeley, CA, USA: USENIX Association. http://dl.acm.org/citation.cfm?id=1251375.1251396.

Goldschlag, David, Michael Reed, and Paul Syverson. 1999. "Onion Routing." *Commun. ACM* 42 (2). New York, NY, USA: ACM: 39–41. https://doi.org/10.1145/293411.293443.

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.

Mittal, Prateek, and Nikita Borisov. 2009. "ShadowWalker: Peer-to-Peer Anonymous Communication Using Redundant Structured Topologies." In *Proceedings of the 16th Acm Conference on Computer and Communications Security*, 161–72. CCS '09. New York, NY, USA: ACM. https://doi.org/10.1145/1653662.1653683.

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."

Reed, M. G., P. F. Syverson, and D. M. Goldschlag. 2006. "Anonymous Connections and Onion Routing." *IEEE J.Sel. A. Commun.* 16 (4). Piscataway, NJ, USA: IEEE Press: 482–94. https://doi.org/10.1109/49.668972.

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