Privacy Threat Model for Distributed Hash Tables

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 lookup arbitrary data in a peer-to-peer (P2P) network. Peers are responsible for storing data and for participating in the lookup and routing mechanism. A DHT does not require a peer to have a complete view of the network and does not rely on central authorities, allowing for resilient, scalable a decentralized networks to form. 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: while users in centralized services may disclose private data to one party, in naive implementations of decentralized networks users disclose private data to many untrusted parties. In this work, we discuss the privacy threat model of DHTs and define the requirements for a privacy preserving DHT.

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. 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 proceed to outline a privacy treat model for DHTs based on previous research work.

Threat Model

Assumptions

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.

Requirements

When studying a DHT protocol from a privacy preserving perspective, we consider two dimensions - it should not leak any information the lookup initiator and it should provide storage anonymity (O'Donnell and Vaikuntanathan 2004) . Based on (Pfitzmann and Hansen 2009) definition, a lookup initiator 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.

Active attacks

Security is a precondition for privacy. Designing a privacy preserving DHT requires the network to actively identify and ban malicious nodes which may actively provide erroneous routing information that can lead data leaks. In this section we outline the active attack surface on DHTs that must be addressed as a requirement for privacy.

According to (Wang and Borisov 2012) DHTs are vulnerable to the following active 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.

Passive attacks

References

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.

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.