Privacy Preserving Distributed Hash Table [DRAFT]

Gonçalo Pestana (goncalo@hashmatter.com)

Dec 2018

Abstract

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.

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 …

- decentralized systems and user privacy
  - decentralized systems pose new threats to anonymity (Greschbach, Kreitz, and Buchegger 2012)
- DHTs as building block to the decentralized web
- what are DHTs
- examples applications
- However, since nodes need to collaborate to route requests and store data, a lot of data is leaked.
- How to design and build privacy preserving DHTs?
- measure performance and privacy (threat model)

The remaining paper is structured as follows. First, we describe how a general purpose DHT works and outline the most popular protocols used nowadays and their respective applications. Secondly, we describe the threat model of DHTs from a security and privacy perspectives. Thirdly, we review 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.

X. Distributed Hash Tables

- any peer can participate in the network
- constrained overlay vs non-constrained

X. Threat model

- high level properties of a privacy preserving DHT (use well known treat models e.g. used by cwitch)
- hide not only the content but also the context of the information.

1) hide initiator information;
2) hide source information;
3) safe against security attacks;
4) decentralized - not relying on centralized services/registries;

Friend to friend restricted network topology

as in Freenet "darknet"

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

In the context of DHTs, …

To the author knowledge, in-DHT onion routing has not been deployed yet.

Provably secure onion circuit building

A node must be able to hide from internal and external peers which nodes were selected when constructing the onion circuit. Formally that is translated by saying that the probability for an adversary node to successfully guess the IDs of the set of nodes which form a circuit to be uniformly distributed in the network, that is:

- given a set `N = {n, n_1, ..., n_i-1}` of `i` nodes that exist in the network in a given time;
- a set of nodes `R = {n, n_1, ..., n_l-1}` which form an onion circuit with a set of nodes of length `l`,
- the length of the circuit `l`;
- and `Pr(n)`, which represents the probability that an adversary can successfully guess that a node `r` is a relay chosen to be part of a circuit, must be:

  `Pr(n) = l/sum(N)`

The probability must hold even though:

1) the node `n` does cannot have a full picture of the netowork at a given time and;
2) the adversary node knows the overlay fingertable of `n` at any time and;
3) relay nodes enter and leave the network in unpredictable ways (churn).

Two approaches to agains route capturing and bias attacks when getting information to construct the onion circuit are:

1) Identify and isolate nodes performing attacks; This is not trivial to achieve in a completely decentralized way (i.e. without relying on central certificate authorities);
2) Use mechanisms to find node IDs that do not rely on direct finger table sharing from other nodes. One naive example could be to passively listen to network activity for a while and select the node IDs that are more likely to be honest. This approach has the problem of publicly narrowing down too much the subset of nodes that are eligible to be part of the circuit, since possible malicious nodes can learn which nodes are more likely to be picked by the circuit constructor when using this passive mechanism.

There a

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.

Greschbach, Benjamin, Gunnar Kreitz, and Sonja Buchegger. 2012. "The Devil Is in the Metadata—New Privacy Challenges in Decentralised Online Social Networks," March.

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.