Privacy Vulnerabilities on Distributed Hash Tables

Gonçalo Pestana (goncalo@hashmatter.com)

Abstract In this paper we enumerate the privacy vulnerabilities of Distributed Hash Tables (DHT) and show how those vulnerabilities affect current decentralized applications that rely on such networks.

Introduction

Distributed Hash Tables (DHT) are overlay networks that enable distributed nodes to store and request data in a peer-to-peer (P2P) network. Data storage and data lookups are resolved collectively through a collaborative routing protocol in a deterministic way. The collaborative nature of DHTs results on resilient, scalable and decentralized networks where nodes are not required to maintain a complete view of the network while, simultaneously, not relying on central authorities. These properties make DHTs an important building block for the decentralized web and P2P systems. However, the decentralized nature of DHTs introduces privacy vulnerabilities due to the metadata leaked during the routing and lookup protocols. In naïve DHT implementations, it is trivial for any adversary to gather information about which nodes are requesting what data and which nodes are providing what data, without being detected and with relative low resources. Nowadays, the decentralized web ecosystem is relying heavily on DHTs as networks such as IPFS and Dat are steadily reaching mainstream adoption. Thus, it is important to address privacy preserving vulnerabilities of DHTs. Failing to deliver on user privacy will render decentralized systems unattractive as a viable alternative to centralized systems. In this paper we enumerate the privacy vulnerabilities of the most common DHT protocols and show how those vulnerabilities affect current applications that rely on distributed hash tables.

Threat model

First, we define privacy and the adversary model considered throughout the paper. We adopt a similar threat model as by (Wang, Mittal, and Borisov 2010), where the adversary is local and controls a fraction f of the network nodes (where f is never larger than 20% of the total nodes in the network). The adversarial nodes perform passive attacks by logging and correlating network requests to infer private and behavioral information of network users. In this paper, we do not consider active attacks or global passive adversaries.

Previous work

(Wang, Mittal, and Borisov 2010)

(Memon et al. 2009) presents *Montra*, a monitoring system to capture traffic in a DHT networks with relatively few resources in a Kademlia DHT (Maymounkov and Mazières 2002), while making sure the monitoring does not bias the network traffic. The monitoring is performed by DHT nodes called Minimal Visible Monitors (MVM). MVM nodes do not route requests or store content. Instead, MVM nodes are placed in multiple resource locations in order to capture lookup requests and thus link lookup originator with requested content ID. The underlying reason that makes it possible to accurately monitor the request traffic is because the Kademlia protocol (and generally all DHT protocols) replicate content to multiple nodes. Thus, the MVM nodes will eventually receive and log lookup requests for content that maps with their peer ID. *Montra* shows how trivial it is to leverage the collaborative nature of DHTs to effectively monitor network requests and link peer

identities with requested content. This can be accomplished with minimal network impact and requires relatively few computational resources. The authors show that their monitoring system can accurately capture around 90% the query traffic in a network with around 32000 peers using a relatively cheap computer (Intel Core 2 Duo with 1 GB RAM). A monitoring system such as *Montra* can be used to record and correlate user behavior over time based on the lookup requests issued by their nodes.

(Mittal and Borisov 2008) has shown several that

So while (Memon et al. 2009), [ref], [ref], show that vanilla DHT protocols leak metadata that can be correlated to , (Mittal and Borisov 2008), [ref], [ref], show how hard it is to achieve lookup and storing privacy in DHTs.

Privacy vulnerabilities of DHTs

We now enumerate four privacy attacks in DHT networks and how they affect the privacy of users.

Caching stalker: This attack allows an adversary to identify an user based on the content she is storing and to learn more information about the user over time, such as her physical location.

Range estimation attack: Wang et al. [14] have shown that based on the positions of a few queried malicious nodes in the lookup, the adversary can narrow the range of the lookup target into a small set of nodes. This attack allows adversaries with relatively few resources pair IP addresses with content lookup

Conclusion

References

Maymounkov, Petar, and David Mazières. 2002. "Kademlia: A Peer-to-Peer Information System Based on the Xor Metric." In *Revised Papers from the First International Workshop on Peer-to-Peer Systems*, 53–65. IPTPS '01. London, UK, UK: Springer-Verlag. http://dl.acm.org/citation.cfm?id=646334.687801.

Memon, Ghulam, Reza Rejaie, Yang Guo, and Daniel Stutzbach. 2009. "Large-Scale Monitoring of Dht Traffic." In *Proceedings of the 8th International Conference on Peer-to-Peer Systems*, 11–11. IPTPS'09. Berkeley, CA, USA: USENIX Association. http://dl.acm.org/citation.cfm?id=1855663.1855674.

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.

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.