Working Title: Simulation based Analysis of Kadcast Privacy Properties

John Smith
The Thørväld Group
jsmith@affiliation.org

Julius P. Kumquat
The Kumquat Consortium
jpkumquat@consortium.net

ABSTRACT

When discussing privacy properties of blockchains or Distributed Ledger Technologies, the focus of the analysis is often on the consensus layer of the blockchain. Another major factor for a comprehensive privacy analysis is the network layer, which defines how messages are passed between peers. On this layer many privacy sensitive informations like public IP addresses of users and connections between peers are used and can potentially be recorded by malicious parties. In this report we will conduct a privacy analysis of the structured P2P overlay Kadcast [Rohrer&Tschorsch], which can be used for efficient broadcasting in blockchain based technologies. The focus of our analysis is the usage of Kadcast for transaction broadcasting and the evaluation of an adversary's capability of linking observed transactions to IP addresses. We show that Kadcast is susceptible to network wide deanonymization attacks by botnets, corroborating the hypothesis that its efficient overlay comes at the cost of privacy. To mitigate the impact of such an attack, we propose to preface the broadcasting phase with an anonymity phase, using the Dandelion spreading algorithm [Fanti].

1 INTRODUCTION

So called "Cryptocurrencies" and other blockchain based technologies are gaining increasing attention since the emergence of Bitcoin, both in the academic community and in media [].

These cryptocurrencies, like Bitcoin can be used analogous to fiat money to transfer "coins" from a sender to a receiver. In Bitcoin, new coins can be "mined" by investing computing power. Unlike a fiat money transaction, which is generally handled by a bank, there is no central instance that authorizes Bitcoin transactions. Instead Bitcoin is a peer-to-peer based system that is run publicly on the internet and allows all participants to achieve consensus on a single valid transaction history without the need for a trusted third party. Users can participate in the network via pseudonymous identities and there is no inherent link between a pseudonymous public identifier and a natural person, albeit deanonymization attacks can be performed to potentially link the public key of a user to an IP address or even a real name []. To store the transaction history, Bitcoin uses an append-only, distributed ledger (the so called "blockchain") and consensus on the state of the ledger is periodically reached by the participants of the network via a distributed consensus algorithm. [for detailed explanation refer to nakamoto paper] Since payments and money flow comprise of very privacy sensitive information, and everyone can join the Bitcoin peer-to-peer network and access the transaction history, the privacy of Blockchains has been an active research area since ...[]. There is a large amount of scientific literature [specifically] on the privacy of the consensus layer of blockchains [], describing weaknesses [e.g using heuristics to construct entity graph, ...] and algorithms to mitigate privacy issues [mixing, coinjoin, zk-snarks, ...] [bib].

Yet another important factor for a comprehensive privacy analysis is the network layer, which defines how messages are passed between participants of the network. Because of the open nature of public blockchains, any adversery can join the network and potentially gain access to privacy sensitive information on the networking layer, like the IP addresses of peers they are connected to. The gathered information can then be used to perform deanonymization attacks, e.g. to link transactions to public IP addresses [first described in Koshy].

One of the most important jobs of the network layer in blockchains is to handle broadcast messages, since both new blocks and new transactions are sent as broadcasts through the network. While Bitcoin started with a relatively naive approach of broadcasting new blocks and transactions in the network [TODO maybe shortly explain gossiping], the network overlay of Bitcoin[/blockchains] has since been revisited multiple times to optimize the message passing with regards to efficiency and privacy [Koshy, Bojja, Fanti, Rohrer, ...] and is continued to be researched and optimized [].

One [semi-]recently devoloped peer-to-peer overlay that can be used for efficient broadcasting in P2P networks is Kadcast. Kadcast is a Kademlia based network overlay that diverges from the naive and redundant[/overhead heavy] approach of "gossiping", to a structured message propagation, leveraging the bucket logic introduced by Kademlia, and achieving a complete network coverage with minimal messages [excluding drops/churn/purposefully introduced redundancy for stability reasons]. [TODO vllt infografik einfügen wie kadcast/bucket logic funktioniert?] The initial application for Kadcast was to broadcast newly mined blocks. Because the direct link between IP addresses and transactions is already broken when a transaction is included in a block, the block propagation can be optimized for efficiency only, without any concerns for privacy issues [on the network layer] [TODO check again].

However, when using Kadcast to broadcast transactions instead, an adversary that is connected to the network can potentially perform deanonymization attacks and link transactions to the public IP addresses that issued them.

Since the travel path of messages in Kadcast is structured, and message redundancy is reduced to a minimum, the question was raised, what sorts of privacy implications using Kadcast as a broadcasting algorithm for transactions would have [R.].

To start answering this question, we want to use a quantitative approach and perform a deanonymization attack as it could happen in a real world scenario. Because of the lack of historical data, we use a [highly] abstracted network simulation to generate data which then can be used for the attack. The simulation setup, including constraints of our simulation and potential resulting implications for the attack are described in [Section Simulation Setup].

We further model our adversary and describe the means of our attacker, explaining which simplifications we made to the attacker model, and why we can use such a simplified model without sacrificing much effectiveness of the attack [Ch. atk.model].
We will then shortly elaborate which metrics we applied to evaluate the impact of the attack [Ch. metrics].
In our evalution we will take a closer look at the results of our simulated attacks, and review if the privacy impacts of using Kadcast for transaction broadcasting align with our research hypothesis, that the structured overlay has [negative] effects on the privacy of the network.
In the last step of our evalution, we will try to mitigate the privacy issues we have observed previously by prefacing the broadcasts with an anonymity phase as described in [Dandelion].
We conclude our analysis by briefly summarizing our results and giving an outlook on where future research on this topic could be headed from this point on.
TODO Results + Outline