Tracing Bitcoins Across Privacy Enhancing Technologies

Master-Thesis von Andreas Rothenhäuser aus Miltenberg
Tag der Einreichung:

1. Gutachten: Prof. Dr. techn. Stefan Katzenbeisser
2. Gutachten: Nikolaos Alexopoulos

TECHNISCHE
UNIVERSITÄT
DARMSTADT

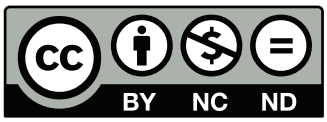
Computer Science
Security Engineering Group

Tracing Bitcoins Across Privacy Enhancing Technologies

Vorgelegte Master-Thesis von Andreas Rothenhäuser aus Miltenberg

1. Gutachten: Prof. Dr. techn. Stefan Katzenbeisser
2. Gutachten: Nikolaos Alexopoulos

Tag der Einreichung:



Erklärung zur Master-Thesis

Hiermit versichere ich, die vorliegende Master-Thesis ohne Hilfe Dritter nur mit den angegebenen Quellen und Hilfsmitteln angefertigt zu haben. Alle Stellen, die aus Quellen entnommen wurden, sind als solche kenntlich gemacht. Diese Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.

Darmstadt, den 7.11 .2018

(Andreas Rothenhäuser)

Abstract

Test

Contents

1. Introduction **6**

2. Related Work **7**
2.1. Blockchain Analysis 7
2.2. Network Analysis 7
2.3. Privacy Enhancing Technologies 8

3. Background **9**
3.1. Blockchains 9
3.2. The Bitcoin Protocol 9
3.3. Authorship Attribution 9

4. Problem Statement + Methodology **10**

5. Experiments + Evaluation **11**

6. Glossary **13**

A. Some Appendix **14**

List of Figures

List of Tables

1 Introduction

The importance of looking at money flows in criminal prosecution was already known back in 80 BC, when Marcus Tullius Cicero phrased the question "Cui bono?" - "who benefits?" in his speech "Pro Roscio Amerino". Since then this idea has evolved to the wide field of financial investigation. Today following the money trail to identify suspects, find backers or explore the hierarchy of a criminal organization is considered a basic task by law enforcement authorities around the world. The ongoing Trump-Russia investigation of Robert Mueller is just one example of how the money trail can convict criminals of all kinds [1, 2].

In 1994 the Internal Revenue Service (IRS) published a whole book dedicated to the art of financial investigation. The book was written as a guide for investigators, summarizing techniques that can be applied to almost any criminal case. Not only does it showcase how certain financial information can be interpreted and leveraged to build a successful case against a suspect, but it also lists potential sources of such information. Looking at chapter 5 of this book, one can learn, that financial institutions "are probably the single most important source of information available to the financial investigator" [3].

This is the reason why the recent rise of the so called cryptocurrencies like Bitcoin poses a major problem to law enforcement. In 2012 the Federal Bureau of Investigation (FBI) assesses, that "law enforcement faces difficulties detecting suspicious activity, identifying users, and obtaining transaction records" [4]. Financial institutions like the European Central Bank (ECB) follow along by stating problems with the enforcement of Anti-Money Laundering/Combating the Financing of Terrorism (AML/CFT) regulations [5]. Unlike traditional financial systems the Bitcoin protocol is designed to work in a decentralized manner without any kind of financial institution to manage it. The protocol itself is the financial institution and investigators can not just walk in to it with a search warrant and receive all the information they need. The traditional techniques of financial investigation as described in the aforementioned book can not be applied in the world of Bitcoin & Co. Law enforcement agencies are confronted with a whole new challenge that demands for new investigative techniques.

In this paper I want to propose such a technique and discuss how it can be applied in the context of criminal investigations. My approach borrows ideas from athorship attribution to improve upon state of the art blockchain analysis. I will use Hierarchical Agglomerative Clustering (HAC) to group Bitcoin addresses based on feature vectors. Opposing many of the previous suggestions on blockchain analysis the main goal is not to deanonymize *most of the Bitcoin users* or *the average Bitcoin user*, but the deanonymization of an individual within a predefined section of the blockchain. In addition I consider the use of Bitcoin mixing and other privacy enhancing technologies not only possible, but due to the context of criminal prosecution as even likely.

The paper will be organized as follows: In chapter 2 I will discuss previous work, that has been done in the field of blockchain analysis and the assesment of anonymity properties in the bitcoin ecosystem. In chapter 3 I will briefly introduce some background aspects like the workings of blockchains and the Bitcoin protocol, special demands in the context of criminal investigations and some basic techniques of athorship attribution. In chapter 4 I will develop the precise problem statement of this thesis and explain the methodology of my approach. In chapter 5 I will describe the experiments I set up to assess the performance of my approach and provide the evaluation results.

2 Related Work

A lot of work has already been done regarding the anonymity properties of Bitcoins. Most of the papers that have been written on this topic can be roughly divided into three different categories based on their perspective and approach. The first category engages in the so-called blockchain analysis. As the name implies, blockchain analysis tries to leverage the information contained in the Bitcoin blockchain itself. Opposing to that, the second category applies techniques of network analysis. These approaches draw information from the analysis of the Bitcoin Peer-to-Peer (P2P) network traffic. The last category of papers deals with a variety of privacy enhancing technologies that have been proposed to improve the Bitcoin protocol over the years.

2.1 Blockchain Analysis

A very prominent approach based on blockchain analysis was already mentioned by the inventor of Bitcoin [**?**] and later used by Reid and Harrigan in [**?**]. It utilizes the fact that Bitcoins, similar to real coins, need to be combined with other coins, if the amount that needs to be payed exceeds the value of any single coin the payee possesses. This combination of Bitcoins is done by creating so called multi-input transactions which have two or more coins as an input and at least one output. [**?**] as well as others ([**?**, **?**, **?**, **?**]) apply Bitcoin address clustering based on this observation in order to reduce the anonymity set of the Bitcoin system. The authors conclude, that multi-input transactions testify a common owner of all inputs of the transaction and thus map the corresponding Bitcoin addresses to the same superordinate entity. If an attacker now manages to identify one single Bitcoin address, the other addresses belonging to the same entity are compromised as well.
Another approach building on top of the aforementioned multi-input heuristic is to identify so called change addresses. This heuristic again relies on an inherent property of the Bitcoin protocol. Since unspent Bitcoins need to be spent in total or not at all, a common use case is to add a special output to the transaction that receives the change (the money you get back, when you did not pay the exact amount). If the change address of a transaction is known, it can be mapped to the same entity as the inputs, reducing the anonymity set of the whole system even further. The change address heuristic was first explored in [**?**]. Lateron it was refined in [**?**] to account for changes in the usage of Bitcoins.

2.2 Network Analysis

The second category discusses approaches that are a bit more invasive, since they require an attacker to participate in the Bitcoin P2P network other than in the originally intended way. Opposing the first category, the target is not to reduce the anonymity set, but to link Bitcoin addresses directly to IP addresses. In [**?**] the authors describe an apporoach based on detection of unusual relay patterns in the Bitcoin P2P network. They propose two patterns that, whenever observed, reveal the true originator of a Bitcoin transaction to everybody listening. Firstly, if one peer is the only one to relay a certain transaction to the network, it must be the true originator and its IP address can be linked to the corresponding Bitcoin address. Secondly, if one peer is the only one to relay a single transaction multiple times the same assertion holds. In order to be able to observe such relay patterns across the network the attacker needs to have a connection to quasi every participating peer.
An even more sophisticated approach was proposed by Biryukov []. Again the attacker needs to connect to as much peers as possible in order to succeed. Conforming the protocol a newly connected peer will advertize its own IP address to the eight random peers it connects to. These in turn will relay the new IP address to their own connections and so on. An completely connected attacker can now log the first

eight peers he receives a new IP address from having a good chance of retreiving the eight random peers of the newly connected peer. Using this eight peers as a fingerprint for following transactions received over the network can unmask the true originator of the transaction. Again a link between the transaction and an distinct IP address is established.

2.3 Privacy Enhancing Technologies

The last category of papers does not deal with the Bitcoin protocol itself, but with privacy enhancing technologies that can be found in the Bitcoin ecosystem. Since they are meant to improve the anonymity of Bitcoin users, any work on assessing or attacking these also adds to the discussion on anonymity in the Bitcoin system. There are two important kinds of privacy enhancing technologies I want to discuss. The first are the so called mixing services and the second is the so called CoinJoin protocol.

3 Background

3.1 Blockchains

3.2 The Bitcoin Protocol

3.3 Authorship Attribution

4 Problem Statement + Methodology

5 Experiments + Evaluation

Bibliography

[1] T. McCarthy, "Trump and 'collusion': what we know so far about mueller's russia investigation," 2018.

[2] J. Rubin, "Mueller follows the money trail," 2018.

[3] D. Vogel *et al.*, *Financial Investigations: A Financial Approach to Detecting & Resolving Crimes*. DIANE Publishing, 1994.

[4] FBI, "Bitcoin virtual currency: Intelligence unique features present distinct challenges for deterring illicit activity," 2012.

[5] ECB, "Virtual currency schemes – a further analysis," 2015.

6 Glossary

AML/CFT Anti-Money Laundering/Combating the Financing of Terrorism

ECB European Central Bank

HAC Hierarchical Agglomerative Clustering

IRS Internal Revenue Service

FBI Federal Bureau of Investigation

P2P Peer-to-Peer

A Some Appendix