2017 IEEE 37th International Conference on Distributed Computing Systems

Consensus Robustness and Transaction De-Anonymization in the Ripple Currency Exchange System

Adriano Di Luzio
Computer Science
Sapienza University of Rome, Italy
Email: diluzio@di.uniroma1.it

Alessandro Mei
Computer Science
Sapienza University of Rome, Italy
Email: mei@di.uniroma1.it

Julinda Stefa
Computer Science
Sapienza University of Rome, Italy
Email: stefa@di.uniroma1.it

***Abstract*—Distributed financial systems are radically changing the way we do business and spend our money. Ripple, in particular, is unique in its kind. It is built on consensus and trust among its users and it allows to exchange both *fiat* currencies and goods over its network. It does so by storing the accounts of its users, their balances, and all the transactions in a distributed *ledger*, publicly accessible.**

In this paper we perform an in-depth study of the Ripple exchange system and its public distributed ledger. We analyze payments, the structure of payment paths, and the role of the entities in the system such as Gateways (the equivalent of banks) and Market Makers. We also analyze the internal stream of events and show that Ripple relies on a surprisingly small number of active validators, raising concerns on the actual robustness and fairness of the system. Moreover, we consider the degree of anonymity that Ripple is able to guarantee. By examining the first three years of Ripple history (more than $500\,\mathrm{GB}$ worth of data), we show that even approximate information on a single payment can uncover, with incredible accuracy, the entire financial life of the user. For example, anyone who overhears our order of a Latte at our favourite bar can easily get complete and unlimited access to our balance, our previous and future payments, our monthly income, as well as critical information about the places where we shop and the people we trust.

***Index Terms*—Ripple, credit networks, distributed systems, privacy, anonymity, payments, transactions, distributed ledger, consensus.**

I. Introduction

Distributed financial systems allow users to exchange money (or assets) without any intermediate central authority. They provide (quasi) real-time transactions, ubiquity, and fairness. These systems have the potential to reshape the financial world and to create entirely new scenarios, fascinating and alarming at the same time. They guarantee some degree of anonymity, thus freedom! However, they also raise new, unprecedented concerns, like the unavoidable upsurge in illicit activities.

Ripple is one of the most popular distributed financial systems. It is based on consensus and trust between users. It allows to trade XRPs, the Ripple crypto-currency, existing *fiat* currencies like the US dollar and the euro, crypto-coins, and any other goods. In addition, it can also act as a *bridge* (unique in its kind) between traditional banking and electronic financial systems. Ripple stores all the transactions, the accounts, and the balances of its users in a distributed *ledger* (*i.e.* a book for recording financial transactions). The consistency of the system is guaranteed by a consensus protocol, used to validate and commit transactions in the ledger.

In this work we question two important features of Ripple. First, the robustness of the consensus protocol and of the validation process; second, the actual degree of anonymity provided to the users. Validation is carried out by special servers called *Validators*. By looking at the public ledger as well as additional information extracted from internal events in Ripple, we monitor the execution of the validation process and get important information on the Validators that take part in the consensus protocol. The analysis, presented in Section IV, shows that the number of active Validators is small, and that the large majority of the transactions are validated by a surprisingly small number of entities. These findings raise some concerns about the robustness of the Ripple system against denial of service attacks and about its fairness and democracy, since the claim that there is no central authority would be better supported by a larger and more diverse set of Validators with respect of the actual set in the system.

Then, we consider anonymity of the users. In Section V we uncover unprecedented privacy issues that affect Ripple. We consider more than 23M unique payments and show how to uncover the real identity behind virtually all the transactions in the system by using a small amount of side channel information. Even very approximate information about a transaction carried out by any Ripple user, like the information that Bob had a Latte this morning in a known cafeteria and payed using Ripple, is enough to reveal the entire financial history of Bob on the Ripple system with surprisingly high probability.

An additional contribution of this work is a thorough and in-depth analysis of the Ripple ecosystem. We consider the public ledger and perform a detailed exploration of its $500\,\mathrm{GB}$ worth of data. We study the currencies that are most used in the system, discovering that a few of them where probably crafted just to launch denial of service attacks. We investigate the structure of payments and payments paths on the trust network of Ripple. Moreover, we look into the role of Market Makers, Gateways, and common users in Ripple, thus discovering, for

1063-6927/17 $31.00 © 2017 IEEE
DOI 10.1109/ICDCS.2017.52

example, that Market Makers are responsible not only for allowing cross-currency payments, but for enabling a large part of the standard transactions in the system. To keep focus on consensus robustness and anonymity, we present this contribution in the appendix.

II. Related Works

Distributed financial systems allow (quasi) real-time payments without intermediate authority. These systems often store the history of transactions in a distributed and publicly accessible database, usually called ledger (or block-chain in the case of Bitcoin). This is useful for transparency and consistency, but it can be a problem for user privacy. Bitcoin, for example, provides some degree of privacy by using anonymous public keys for transactions; still it has some privacy issues that have been extensively analyzed [1], [2]. For example, it is possible to recover half of Bitcoin user profiles [3] and link their IPs to public Bitcoin accounts and to their transactions [4].

Distributed financial systems have also a dark side, since they can be used to facilitate illicit activities, like selling illegal goods, money laundering, and unlawful gambling, to name a few. These concerns have been considered in the literature for Bitcoin [5], for example.

Ripple, on the other hand, started to attract the attention of the scientific community only recently. Its consensus protocol [6] has been analyzed [7], [8]. This analysis has resulted in a modification of the protocol consisting in an increase of the agreement majority required to approve transactions.

On the privacy side, several concerns related to Ripple have also been raised. The users of the system themselves, in a forum online, proposed new methods for *proxying* payments [9]. Moreno-Sanchez *et al.* [10] proposed two novel heuristics to cluster the users of distributed financial systems. With the first heuristic, they link Bitcoin and Ripple accounts. With the second heuristic, they cluster different, apparently non-correlated, Ripple accounts that are actually owned by the same entity and describe a de-anonymization methodology to uncover the owner.

However, none of these works investigates whether, and how accurately, it is possible to de-anonymize a Ripple user by looking at the details of one of her payments. A similar problem has indeed been considered for credit card payments. For example, Montjoye *et al.* [11] analyze a dataset of anonymized credit card payments. They show that, by considering 4 different spatio-temporal points related to credit card payments in the dataset, one can de-anonymize the individual that carried out the payments and reveal his entire history of transactions in the dataset itself. To the best of our knowledge, this kind of analysis has never been performed with a world-scale financial distributed system. Ripple, with its public distributed ledger, gives the opportunity to fill this gap.

III. Background

Ripple is a decentralized financial system. Its main goal is to enable people to exchange assets—money, goods, and so on—everywhere in the world and almost in real-time. It does so by maintaining a *distributed ledger*, literally, a "book for recording transactions". At a high level of description, the ledger maintains information on all accounts in the system—balance, currency, transactions that have changed the balance in the past, and so on. Ripple creates a credit/debit network, similarly to what traditional banking systems do. Say, for example, that Alice opens a new account with her bank. She, then, deposits $5\,\mathrm{USD}$ in the account and she is issued a debit card linked to it. At that moment the bank gets into debt with Alice for $5\,\mathrm{USD}$. When Alice makes a purchase of $5\,\mathrm{USD}$ with her debit card (*e.g.* a beer in her favorite pub) the bank zeroes out the debit towards Alice. In addition, the bank's debit is moved from being towards Alice's account, to being towards the owner of the pub's account. All transactions and debit movements are recorded in the bank's database.

Ripple works in a similar way. Debt moves from user to user. Users can deposit money to another Ripple user (e.g. a *Gateway*, the Ripple's equivalent of a traditional bank) or purchase goods (*e.g.* a beer). These operations can be done through Ripple transactions and each transaction modifies the Ripple's distributed ledger. Gateways are particular users that may be the Ripple's interface to an existing bank or an online platform that acts like one. Their purpose in Ripple is manifold. They provide users an entry or exit point to the system (by converting real-world assets to Ripple issuances, and vice-versa), they can act as intermediaries between any two users (like banks in real life), and they can serve as points-of-exchange in the Ripple's network.

For each pair of users in the system, including individuals and Gateways, Ripple keeps track of the balance (*credit* or *debit*) between the parties. Transactions modify the balance in one way or the other. The safety and the consistency of the system are guaranteed by a consensus protocol [6] that approves the transactions to be recorded in the ledger. In the case of a deposit, the ledger is modified so as to keep track of the Gateways's debit towards the user. In the case of a payment, it keeps track of the *movement* of debit from one party to the other (*e.g.* towards the pub owner, in the above example).

A. Ripple IOUs, and differences with other banking systems

Ripple fundamentally differs from traditional banking and from distributed electronic currency systems like Bitcoin. The difference comes from the fact that, in Ripple, transactions between users can be ephemeral, not corresponding to real-life amounts deposited from one user to another, like, *e.g.*, in traditional banking. Indeed, the Ripple ledger records only debits and credits between users and the transactions that modify the balance, and is unaware on how and when the actual money or goods are deposited or exchanged. This means that, if Alice owes Bob $10\,\mathrm{USD}$ in Ripple, nothing prevents her from running away and never actually paying her debt. For this reason, Ripple payments are often simply referred to as "*I Owe You*" (IOU) payments.

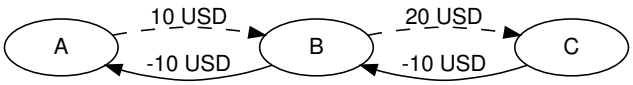


Fig. 1. A trusts B for 10 USD, and B trusts C for 20 USD (trust-lines are dashed in the figure). Therefore, C can potentially issue a IOU payment of up to 10 USD to A, through B (the solid lines represent the payment path).

Another fundamental difference between Ripple and traditional banking or electronic money exchange systems is that, in Ripple, transactions happen in quasi real time. This is due to the (fast) distributed agreement protocol through which transactions are validated. The protocol validates transactions within a fistful of seconds. As a consequence, paying someone on the other side of the world through Ripple takes, on average and with the current system workload, from 5 to 10 seconds. This is significantly faster than Bitcoin transactions (that take 10 minutes for the validation and about 1 hour for the actual confirmation), and extremely faster than traditional bank wire transfers that might take days or a week to go through.

Ripple also defines its own crypto currency, the XRP (the *ripple*). XRPs are pre-computed by the system's creators [12], [7] and are distributed either by them for research or other purposes, or can be purchased by users after they join the system. Ripple users can directly exchange XRPs or trade them with any other currency (they act as a *bridge*). In addition, XRPs are also used to prevent *ledger spamming*. A small XRP fee is indeed collected for each transaction submitted to the system. The aim is to mitigate denial of service (DoS) attacks. At the same time, requiring XRPs for every transaction restricts the exchanges only between users that own XRPs. The fees collected during transactions are not destined to other Ripple users, or validators, like in Bitcoin or other platforms. They are destroyed after the corresponding transaction is confirmed. Nevertheless, Ripple's designers made sure that there will be enough XRP liquidity to allow the system to last for thousands of years. The effectiveness of XRP fees to prevent spamming has been thoroughly investigated in [12] and [7].

B. XRPs, trust-lines, validators, and balances

Ripple allows two types of exchanges: direct XRPs payments and other currencies/goods payments (IOU payments). Direct XRP payments are straightforward. Ripple keeps track of the amount of XRPs that each user owns; any user willing to transfer XRPs to someone else is required to *fill* a transaction, cryptographically *sign* it, and *submit* it to the system. Once the transaction is successfully included in the ledger, it is considered final, complete, and immutable. The XRP balance of the two parties is accordingly updated. The amount of XRPs is subtracted from the sender's balance and added to the receiver's. The XRP is the only currency that cannot be owed to other users—it it effectively transferred from balance to balance and this does not require any cooperation from the receiving party. On the other hand, IOU payments work in a different way. First, users willing to receive IOU payments are required to create so called *trust-lines* with other users in the system by declaring the amount of trust towards each of them. In the system, if user Alice trusts Bob for $10\,\mathrm{USD}$, this means that Alice is willing to give Bob credit for up to $10\,\mathrm{USD}$. The amount of trust is specific to a currency and bounds the amount of IOU payments in that currency that can be transferred over the trust-line. Clearly, trust is uni-directional. In the above example of Alice and Bob, the trust-line of $10\,\mathrm{USD}$ from Alice to Bob limits IOU transactions in the opposite direction (from Bob to Alice) to $10\,\mathrm{USD}$.

Trust-lines are the components of transaction paths in Ripple. In fact, a transaction path is a sequence of trust-lines, along which IOU payments travel in the opposite direction. Therefore, if A trusts B for $10\,\mathrm{USD}$, and B trusts C for $20\,\mathrm{USD}$, then C can potentially send a IOU payment of up to $10\,\mathrm{USD}$ to A through B (see Figure 1) even in the case when C has no direct trust from A. The actual IOU payment that can be delivered, however, depends on the current debit on the trust-lines of the path. Ripple keeps tracks of debit between users. Indeed, for every user and every currency (except XRP) Ripple keeps the balance of the debit with a record consisting of three fields: amount, currency, and issuers.

Every time that a user needs to make a IOU payment to another user, a route is created that can potentially serve as a payment path of the given amount. The payment path is then submitted to the system for a validity check of the trust-lines in the path—amount of trust and current debit. This is done by the *Validators*—servers whose purpose is to check that validity of the transactions and to execute the consensus protocol. When agreement is reached, the transactions in the agreement are permanently added to the distributed ledger as a new page.

C. Currencies, markets and exchanges

Ripple allows users not only to make IOU payments in a currency, but also to trade currencies. These types of transactions, also called currency exchange offers or orders, create lots of possibilities for the users of the system. For example, a user A, that has incoming trust-lines in USD only, can pay a user B in Euro. Transactions of this kind are called "cross-currency" IOUs and they require a "bridge" between the two currencies at some point of the transaction path. The bridging is done by *Market Makers*, a particular type of Ripple users that create the exchange offers in the system. Ripple's path-finding algorithm exploits Market Makers to deliver cross-currency payments and it does so by selecting the path with the best exchange rate available.

Thanks to Market Makers and to exchange orders, Ripple network is flexible and robust. Same-currency payments can use one or more exchange offers to make up for the lack of direct trust on a particular currency between the parties; multiple offers can be chained, *e.g.* from USD to EUR, then from EUR to BTC, if there is no Market Maker that trades USD with BTC directly or if it is just cheaper; XRPs can be used as a universal bridge between markets—any currency to XRP, then from XRP to any other currency. For the

same reasons, Ripple users can also try to take advantage of the exchange offers, exploiting the price skew between two or more markets. This process, called arbitrage, consists in buying assets at a competitive exchange rate and then selling them immediately at a higher price. Arbitrage is allowed by design in the Ripple exchange system and can also be performed automatically, for example by a financial bot.

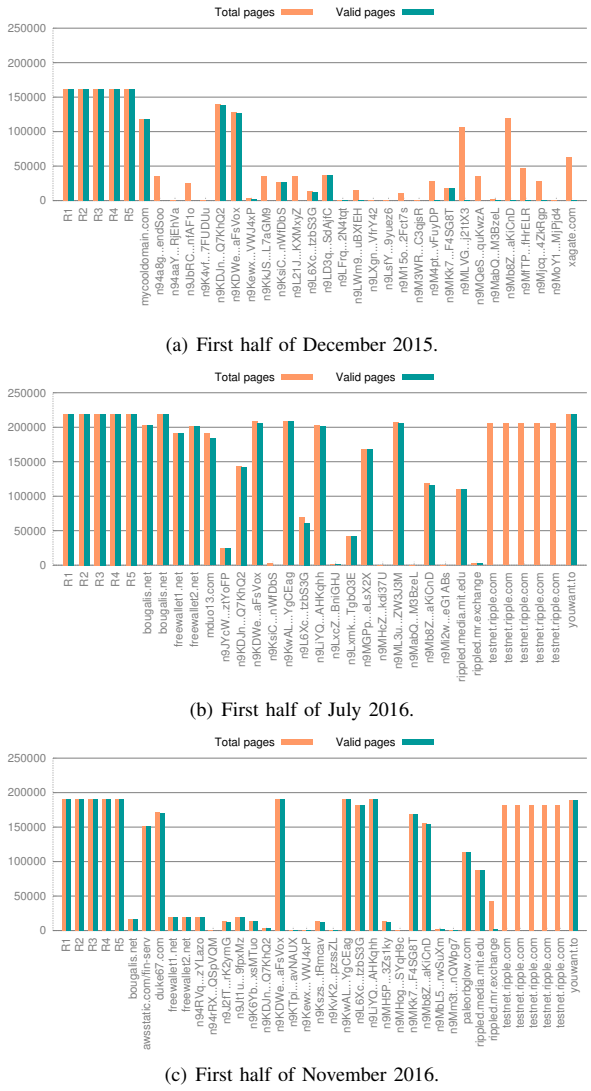
IV. Robustness and Trustworthiness of the Ripple Consensus Protocol

Validators play a crucial role within Ripple! They run the consensus protocol to decide the transactions that go through and are sealed in the next ledger page. Note that, by design, each Ripple validator can choose which transactions to sign and support. This means that validators can authorize invalid or fraudulent transactions and vote for their inclusion in the distributed ledger or, conversely, block specific transactions without justification. However, in both cases, unless all validators collude, the disagreement would be noticeable to any of the "correct" validators that participate in the process. Therefore, the trustworthiness of the agreement process is correlated to the number of the validators and their geographical distribution worldwide. Indeed, the more widespread they are and the larger their number, the more difficult for an adversary to take over the validation process and potentially control the whole system [6]. For this reason, we deemed important to investigate the validation process and servers in Ripple. In particular, we focused on the number of validators, on the transactions that they process daily, and on discovering, where possible, the institutions behind the validation servers. In our belief, this analysis sheds light on the robustness and trustworthiness of the consensus within Ripple.

Ripple's distributed ledger does not store any information regarding the validators or the consensus protocol. For the purpose of our analysis, we needed to collect real-time information on the consensus rounds and the validation process in the system. We did so by setting up a Ripple server that made use of the Ripple's validation stream to capture and store the aforementioned data. The collection spanned over three different periods of two consecutive weeks each—the first two weeks of December 2015, July 2016, and November 2016. The purpose was to investigate a possible evolution in the Ripple validation protocol[1].

From the streams we were able to infer the validators operating during the collection periods under investigation, their public keys, and the pages signed by each of them on the several consensus rounds run in the system. Then, we performed a more in-depth investigation including the information within the ledger corresponding to the data collection period. The goal was to analyse the behavior of the validators in terms of number of pages they agreed (or refused) to sign and the contribution of each of them to the validation of the pages

[1]We note that Ripple Labs Inc. provides statistics on the validators at a daily level (`https://validators.ripple.com`). Our analysis, not only spans a larger period (2-weeks), but it also sheds light on how the validation process within Ripple is evolving.



(a) First half of December 2015.

(b) First half of July 2016.

(c) First half of November 2016.

Fig. 2. Number of total and valid (ledger) pages signed by validators in the corresponding timeframe.

that were finally included in the ledger. Recall that, only those pages that are signed by at least 80% of the validators end up in the distributed ledger. The results of our study for the three periods under investigation are presented in Figure 2. In the graphs, we plot, for each validator, the number of pages it has signed (total) and the number of the signed pages that ended up in the ledger (valid). Validators are labeled with the internet domain they associate with or with their public key ID. An exception is made for the validators labeled with R1–R5 known to be managed by Ripple Labs Inc., the company that originally created Ripple.

We first observe that the Ripple Lab servers are the ones who contribute the most to the validation process. They present the overall highest number of both total and valid signed ledger pages. However, this was somehow expected since R1–R5 belong to Ripple Labs. Since the genesis of Ripple the

company has invested time and resources for their availability, computational power and low-latency responses, necessary for the efficiency of the consensus and the overall system.

Let us focus on the analysis of the December 2015 collection period. Excluding R1–R5 (Ripple Labs), among the 29 validators observed, just a handful of 3 of them were actively contributing to the consensus protocol (see Figure 2(a)). We note that they are currently unidentified, i.e., they are not publicly associated with any financial institution. In the same period, 5 additional validators display a very small fraction of valid pages. We believe that, at the time, they were struggling to stay in sync with the rest of the system, due to limited hardware or network performance. For this reason, the pages they signed do not match those of the rest of the system. The remaining 21 validators exhibit an odd behavior. None of the pages signed by them is valid. Therefore, we believe that, they either were contributing to a different, private Ripple ledger, or their latency made it almost impossible to participate in the distributed protocol.

When we consider the July 2016 collection period, we notice that the number of the actively contributing validators is considerably increased. Indeed, 10 among the 28 observed during the period present a number of valid pages close to or comparable to those of R1–R5 (see Figure 2(b)). In addition, it is worth mentioning that 4 of them had a publicly associated domain—`bougalis.net`, `modulo13.com` and `youwant.to`. Even so, at the time of writing, all these domains displayed empty or inaccessible web-pages. The other 6 active validators are currently unidentified. This is quite surprising. Running a validator is an expensive task, as it requires powerful machines with broadband internet connection in order to guarantee a timely reaction during the consensus protocol. In addition, the validation process does not raise any revenue to the machine owner. So, a regular Ripple peer would have almost no reasons to manage a validator. Even though we cannot be certain, we believe that the 6 actively operating anonymous validators might be controlled by as many financial institutions raising revenue within Ripple (*e.g.* associated with Ripple Gateways or Banks). These institutions could therefore be motivated to contribute to the robustness and consistency the system overall. At the same time, they might not want to publicly announce their support to the system in terms of transaction validation.

Finally, let us focus on the third timeframe considered, that of November 2016 (see Figure 2(c)). We observe that the number of the validators observed in the period has grown w.r.t the previous timeframes. There are now 34 of them (see Figure 2(c)). However, the fraction of the very active ones has decreased in comparison to the period of July 2016. Indeed, only 8 of the 34 observed present a number of valid pages comparable to those of R1–R5. While, in July 2016, the fraction was $^{10}/_{28}$. As an example, in July 2016, the two validators labeled with `freewallet1.net` and `freewallet2.net` participated to $+200\,000$ consensus rounds. While, in November of the same year, they contributed to less than $20\,000$ ledger pages (*i.e.* an order of magnitude less). Similarly, in July the two validators associated with `bougalis.net` succeeding the Ripple validators in the figure were available as much as R1–R5. Whereas, in November, one of them completely disappeared, while the other has been present for only $15\,000$ rounds.

Both in July and November 2016, there are 5 validators which exhibit around 200K pages signed, with none of them being part of the main ledger. A further investigation showed that these validators are running the consensus protocol for the Ripple's test-net (`testnet.ripple.com`), a parallel instance of the ledger in which developers and users can freely experiment with the system. Additionally, the three periods share only 9 (over a total of 70 validators seen) that appear in each of them as active contributors to the consensus protocol. This points out that the set of Ripple validators is still dynamically changing.

Lastly, we observe that the number of peers that effectively contribute to the consensus process has largely increased in time since December 2015. But, as of November 2016 it is still relatively small. We believe that the 5 Ripple Labs validators will continue to be available anytime in the future, guaranteeing the successful conclusion of every consensus round. However, the overall small number of validators raises concerns regarding the general availability of the system. E.g, a malicious party hijacking or compromising the majority of these validators could endanger the whole Ripple system. A solution could be introducing a carefully crafted reward system that would stimulate the entry of new validation servers in Ripple. For example, the reward could be defined as an added tax value to the transactions that go through in each validation round. A larger number of validators would lead to a better distributed validation process that in turn would improve the reliability of the entire system.

V. De-anonymizing Ripple Transactions

Imagine that Bob is about to buy a latte in his favorite bar, that accepts Ripple payments. Alice, a stranger to Bob, is waiting in line right behind him. So, when Bob transfers through Ripple the sum of 4.5 USD, the price for a latte, Alice is able to gather the following information on the transaction: the Ripple address of the bar (the receiver), the currency and amount of the transaction, and the time at which the transaction occurred. We would like to investigate whether, with this information, Alice can de-anonymize the address of Bob within Ripple and thus link all past and future transactions to his accounts. More generally, we want to investigate the protection of user and transaction privacy within Ripple.

Transactions and accounts in Ripple are supposed to be anonymous by design. Ripple accounts are unambiguously identified by a 160 bits string, typically displayed in a human-readable form by using the Base58 encoding. These identifiers are randomly generated and contain no semantic information on the real-world entity that created the account[2], which, in

[2]Users may choose to set up a meaningful Ripple username for their account, but these atypical cases are out of the scope of our study.

TABLE I
DETAILS ON THE ROUNDING PROCESS FOR THE DIFFERENT CURRENCY STRENGTH GROUPS. FOR EACH CURRENCY, A GIVEN RESOLUTION LEVEL ROUNDS THE ORIGINAL VALUE TO THE CORRESPONDING CLOSEST 10^x VALUE.

Strength	Currency	Max (m)	Average (a)	Low (l)
Powerful	BTC, XAG XAU, XPT	10^{-3}	10^{-2}	10^{-1}
Medium	CNY, EUR, USD AUD, GBP, JPY	10^{1}	10^{2}	10^{3}
Weak	XRP, CCK, STR KRW, MTL	10^{5}	10^{6}	10^{7}

principle, makes Ripple accounts and payments intrinsically anonymous. However, the lack of semantic information is not enough to guarantee the anonymity of Ripple payments, in view of a situation like the above regarding Alice and Bob. In the remainder of this section we show how, by exploiting publicly available information stored in the Ripple's distributed ledger, we are able to link the almost entirety of transactions to the corresponding Ripple users, severely jeopardizing their privacy.

A. Experimental Setup

We carried out a study over the $+23$M Ripple transactions from the timeframe January 2013 (system genesis) – September 2015. For each transaction, we extract the following features: i) the sender account S that submitted the payment; ii) the amount A delivered; iii) the timestamp T of the transaction defined as the moment in which the corresponding ledger page passed the consensus protocol; iv) the currency C delivered; v) the destination account D that received the payment. The goal is to de-anonymize the senders S starting from any subset of the other transaction fields $A, T, C,$ and D, as we have seen in the bar example of Section V. For completeness, we investigate the impact that each field has on the de-anonymization process by considering scenarios where the information on the field value is course-grained or of "lower resolution" (e.g., the timestamp being at the level of hour or day instead of seconds).

More formally, for each transaction we consider a list of type $\langle A_{res}, T_{res}, C_{res}, D_{res}\rangle$, where the subscript res denotes the resolution granularity of the corresponding field. We define a resolution factor for each payment feature so that lower resolutions coarsen the quality of the data. The destination account D and the currency C are composed of nominal values. As a result, their resolution is binary: either we consider them as part of the fingerprint or they are completely ignored in the analysis. On the other hand, amount A and timestamp T represent discrete numerical values. Ripple's ledger provides a precision to the level of seconds for the payment timestamps T and to the level of 1×10^{-6} (one millionth) for the amount exchanged A. In the case of timestamps, we coarsen their quality by reducing their temporal resolution: from seconds (T_{sc}) to minutes (T_{mn}), hours (T_{hr}) and days (T_{dy}). As an example, the worst resolution of the timestamp will modify the value 2015-08-24 15:41:03 to 2015-08-24 00:00:00.

Regarding the transaction amount A we apply a rounding process on the corresponding value. We consider three levels of resolution denoted with A_m (maximum), A_a (average) and A_l (low), that correspond to three different rounding processes. Currencies have different market strengths. For this reason, the rounding process depends also on the strength of a given currency in the market. Let us explain this better with an example. For the EUR currency we consider the following three different resolutions: maximum (A_m), achieved by rounding to the closest tens (10^1), average (A_a), achieved by rounding to the closest hundreds (10^2), and low (A_l), achieved by rounding to the closest thousands (10^3). Now, a Bitcoin (BTC) is worth hundreds of EUR. As a result, the payments in BTC involve values which are considerably lower (several orders of magnitude) w.r.t the payments in EUR. We take this into account in the rounding process and, for BTC, we consider the following three resolutions: A_m, rounding to the closest thousandth (10^{-3}), A_a, rounding to the closest cent (10^{-2}), and A_l, rounding to the closest tenth (10^{-1}). We group currencies with similar market strength together and we apply the same rounding process to members of the same strength group. Table I presents the specifics for the rounding process of each resolution level for the different currencies considered.

B. Results

In our analysis, we are interested in discovering whether a subset of the transaction fields excluding the sender produce a unique fingerprint throughout Ripple's history. More formally, given a list of Ripple transaction fields with corresponding resolutions $LT = \langle A_{res}, T_{res}, C_{res}, D_{res}\rangle$, we define the information gain *IG* of LT as the percentage of Ripple transactions whose sender address field S can be uniquely identified. Intuitively, the $IG(LT)$ measures the accuracy with which we are able to de-anonymize the sender of a Ripple transaction starting from the fields and the corresponding resolutions in LT. For the purpose of our study, we examine the trend of IG varying the number of transaction subfields considered, as well as their resolution. The results are presented in Figure 3.

We start by noting that, the IG value is the highest when we consider all fields in lists LT at highest resolution (case $\langle A_m, T_{sc}, C, D\rangle$). Indeed, in this case we are able to determine uniquely more than 99.83% of transaction sender addresses. This means that Alice, from the previous bar example was likely to uncover Bob's Ripple identifier with almost certain probability. In other words, by picking any random Ripple payment, one can deterministically and reliably uncover the account that submitted it to Ripple, which, in turn, will also uncover all past and successive transactions performed by the same account.

Then, we investigate the impact that the various resolution levels have on the IG. To start, we remove the $0/1$ subfields, i.e. currency C and destination address D, separately. The lack of information on currency C does not particularly impact the de-anonymization result (see the results on $\langle A_m, T_{sc}, -, D\rangle$

in Figure 3). In fact, we can still de-anonymize the sender account for 99.83% of the transactions. On the other hand, by removing the destination of each payment D (the case of $\langle A_m, T_{sc}, C, -\rangle$), the percentage of unique payments slightly decreases to 93.78%. Then, we investigate the impact of the resolution for the fields A (transaction amount) and T (timestamp). Completely removing A does impact the IG in a non negligible way. The detection accuracy of the sender drops to 89.86% (see $\langle -, T_{sc}, C, D\rangle$). However, the lack of the timestamp information T has even a greater impact. The IG for the case $\langle A_m, -, C, D\rangle$ drops to less than what you get, on average, with a coin toss (48.84%, see Figure 3). This indicates that T's information gain not only is higher than A's, but is also the highest among all the features.

Finally, we notice that, when we reduce the resolutions of A and T to the intermediate values discussed in the previous section, the IG decreases even further. The detection accuracy is particularly impacted if both A and T are more coarse grained at the same time (*e.g.*, slightly more than 50% for $\langle A_l, T_{dy}, C, D\rangle$ in Figure 3). If, in addition, we also remove the information within C (currency) and D (destination), the IG drops down to 1.28% (see the last quadruple's result in Figure 3). This is expected. A low approximation of the transaction amount coupled with the lack of currency and destination makes it very difficult to identify a transaction in Ripple.

That said, our results raise several new, unprecedented concerns related to the privacy of Ripple users. We show how anyone is able to de-anonymize, with high accuracy, any Ripple sender address, by analyzing the (possibly approximated) details of just a single payment. Similar issues have been raised before for other distributed financial systems. Bitcoin, as an example, though fundamentally different from Ripple, yet provides a public, distributed ledger of transactions and also suffers from de-anonymization attacks [2], [13]. A possible solution is to create multiple Bitcoin wallets unique to every single transaction, in order to make it difficult to link them to a single user. However, a similar approach is difficult to achieve in Ripple due to its underlying trust backbone—every new wallet would need to create enough new trustlines in order to perform transactions. This makes the bootstrapping very complex and expensive. In addition, each wallet would require to be trusted by the receiver of the payment, decreasing the usability of the system and possibly allowing the different wallets to be linked back together.

VI. Conclusions

This work aimed at shedding light on the robustness and user privacy of the Ripple currency exchange system. We started by studying the trustworthiness of the validation process—one of the system's main pillars. We collected and analyzed real-time data from three separate periods of 2-weeks each with the purpose to investigate also possible evolutions in the transaction validation ecosystem. Our results show how the consensus in Ripple depends only on a very small number of validation servers, and that the set of Ripple

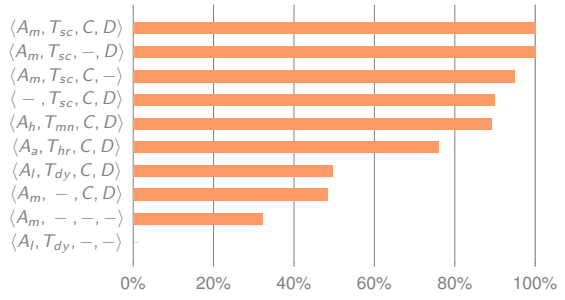


Fig. 3. Percentage of Ripple payments producing a unique fingerprint, per feature and resolution. A, T, C and D denote respectively the transaction amount, timestamp, currency and destination, whereas their subscripts indicate their resolution as follows: m, h, a and l indicate the max, high, average and low resolutions for A; sc, mn, hr, dy refer to seconds, minutes, hours and days for T.

validators dynamically changes in time with a very high pace. Additionally, the small number of reliable validators indicates that the strength and the robustness of the Ripple agreement protocol, and thus, the entire system, has still plenty of room for future improvements.

Then, we focused on investigating user privacy issues within the system. We showed how, by using just sidechannel information on a single transaction, one is able to uncover with very high accuracy (more than 99%) the unique identifier of the user that originally submitted the transaction. As a consequence, it is possible to discover all the past transactions of a user within the system, and it becomes straightforward to keep track of all his future payments as well.

Last, we analyzed, for the first time, one of the largest distributed and peer-to-peer money exchange systems, Ripple. By examining its distributed ledger we studied, in details, the first three years of life of the system. We uncovered the trends of its payments, to which extent users and order exchanges contribute to delivering payments, what differentiates Gateways, Market Makers and other Ripple peers, and how balances and trust are distributed within the system.

Appendix
Ripple: An in-depth analysis of the system

To perform our analysis of Ripple, we collected, stored, and then processed all the data of the first three years of its history—from the *genesis* to September 2015. We did so by building an ad-hoc Ripple client that downloaded more than $500\,\mathrm{GB}$ worth of data from the Ripple's distributed ledger. The ledger is publicly accessible and stores the whole history of Ripple including every single transaction submitted and validated since its inception in January 2013—-payments, balances, trust relationships between users, offer exchanges, and so on.

A. Most-used currencies

We start off our study of the Ripple ecosystem by investigating what currencies are exchanged by the users. In Figure 4

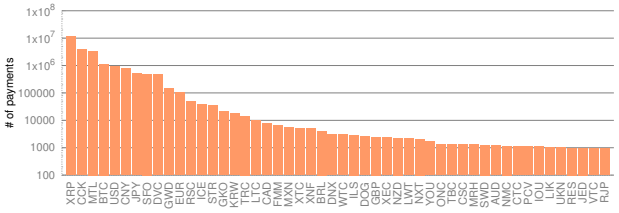


Fig. 4. Ripple's most used currencies, since January 2013 till September 2015. The y-axis is logarithmic.

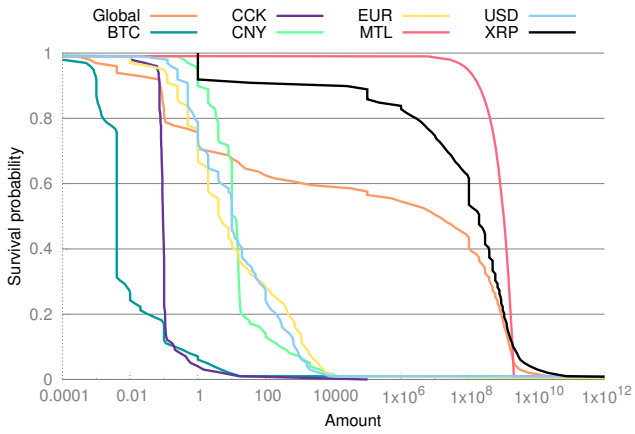


Fig. 5. Survival function of amounts exchanged through Ripple, since January 2013 till September 2015. The term "Global" refers to the currency-unaware distribution of amounts.

we list the currencies that are used the most in the entire life of Ripple in terms of number of transactions. First, note that the XRP is on the top of the list. Indeed, more than 10^7 payments (49% of payments) exchanged XRPs, for an average of 10 000 per day. This is surprising, being XRP's main purpose to prevent ledger spamming. As our analysis reveals, Ripple users also actively use it to make payments. In particular, more than 700K transactions (almost 10% of XRP payments) are sent to the ∼`Ripple Spin` account, owned by a gambling website launched in 2015 that let users bet XRP. Over 1M payments are sent to Ripple's `ACCOUNT_ZERO`, the special account that initially owns all the available XRPs. This number of transactions has a different explaination. After the system is bootstrapped, all the funds in `ACCOUNT_ZERO` are distributed to the other users, and its balance is zeroed. The secret key associated with this account is publicly known (*i.e.* it is hard-coded in Ripple's protocol definition), so anyone can sign transactions on its behalf. For this reason, Ripple spammers used this account to repeatedly send back-and-forth to their accounts small amounts of XRPs with the intent of increasing the system load.

Among the well-known currencies, Bitcoin is the first to appear on the list (fourth with 4.7% of the transactions), followed by the USD (3.8% of the transactions), the Chinese Yuan CNY (3.3% the transactions), and the Japanese Yuan (2.1% of the transactions). EUR is only 11th with 0.4% of the transactions. What is surprising is that the second and the third mostly used currencies are the CCK and the MTL, not among the currencies officially recognized by the currency codes standard [14]. To further investigate the matter we plot, in Figure 5, what we call the *survival* function of payment amounts for CCK, MTL, and some other of the leading currencies. The survival function for a given currency is defined as the percentage of payments in that currency exchanging an amount larger than a certain value.

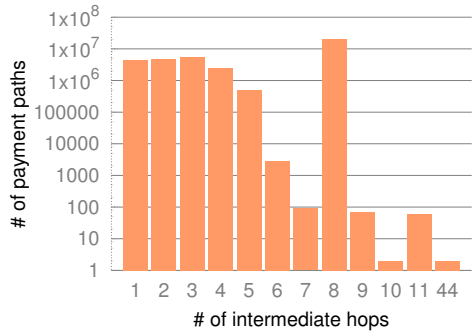
The trend of the MTL shows that all the payments deliver incredibly high amounts, of the order of 1×10^9. In our analysis (we omit the detailed results for brevity) we discovered that a single account submitted quite a large number of MTL payments specifically "crafted" to increase the load of the system. This was probably an attempt of denial of service to Ripple that did not succeed (the ledger does not show any discontinuities related to the event). The only visible effect is that the attacker's account piled up a significant MTL debit, of the order of 1×10^{22}.

Figure 5 shows interesting trends for other important currencies. The EUR and the USD, for instance, have similar market values and their survival functions are remarkably similar. The BTC, instead, is significantly stronger than the other currencies. Therefore a large number of transactions exchange a small amount of BTCs as we can see from its survival function. The CCK has a survival function similar to the BTC, consisting of a large number of micro-transactions. Therefore, we believe that the CCK refers to either something quite valuable in the real world (similarly to the BTC), or, possibly, to another example of DoS attack to the Ripple system.

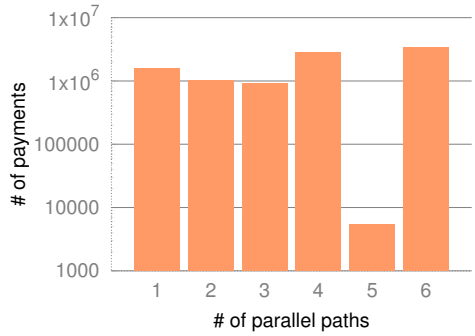
B. Payments and payment paths

Transactions in Ripple travel along trust-paths made of trust-lines between users. Thus, we deemed important to investigate the structure of the paths used by the Ripple transactions. We did so by examining the length of the paths and the number of parallel paths in which payments are split to successfully reach destination. Among the 23M transactions in the period under investigation, 13M of them are direct XRP payments. This analysis, thus, focuses on the 10M transactions that require more than one hop on the trust-lines to reach destination.

In Figure 6(a) we plot the number of payments against the number of intermediate hops in the trust-path. Generally speaking, the majority of payments are delivered through less than 5 intermediate hops and the trend is, as one would expect, decreasing. However, there is an exception. A considerable amount of payments is routed through exactly 8 intermediate nodes. A further investigation revealed that this is due to 3.3M transactions involving the exchange of MTL, a currency that is notoriously known for ledger spam within Ripple, as we have seen in Section A. We believe that these transactions were intentionally forced to be routed through exactly 8 intermediate hops in order to purposely increase the load on the system and attack its availability.



(a) Intermediate hops.



(b) Parallel paths.

Fig. 6. Number of intermediate hops per payment path and number of paraller paths per payment. Both the y-axes are logarithmic.

Figure 6(b) plots the number of payments vs the number of parallel paths in which they are split to reach destination. We observe that there is a high number of payments that are either not split at all (16.3%) or split in two, three, or four parallel paths (respectively 10.4%, 9.3%, 28.9%). Of the remaining, more than 1M (34.8% of the total) refer to the same 3.3M MTL transactions that, we believe, are ledger spam and that were forceded to make use of exactly 6 parallel paths.

C. The impact of Market Makers in Ripple transactions

Market Makers, as any other user in Ripple, often contribute as hops in single-currency transaction paths. But, most importantly, without them and their exchange offers it would be impossible to make cross-currency payments. Therefore, we study to which extent the lack of Market Makers, in view of a possible attack, would impact the functionalty of Ripple.

Let us start off with the offers. The 16M ledger pages under investigation in this work contain around 90M offers. What is surprising is that a large part of these offers is placed by just a handful of Market Makers. In fact, 44M (50%) are generated by 10 Market Makers only; 67M (75%) by 50 Market Makers only, and 87% by 100 Market Makers only. A direct consequence is that, by taking over or thwarting the functionality of a very small number of users (*e.g.* the 10, 50, or 100 more active Market Makers) an attacker could control

TABLE II
NUMBER OF TRANSACTIONS SUBMITTED AND DELIVERED BY RIPPLE IN ABSENCE OF MARKET MAKERS.

Category	Submitted	Delivered	Delivery rate
Cross-currency	1,185,521	0	0%
Single-currency	538,169	194,300	36.10%
Total	1,723,690	194,300	11.2%

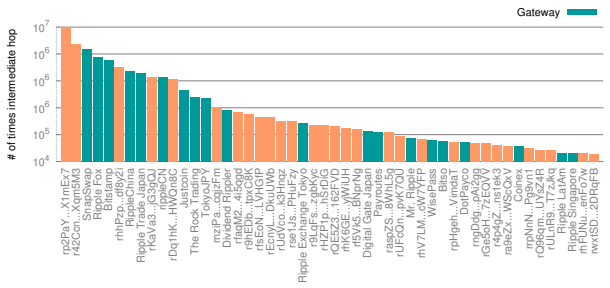
or block a massive number (respectively 50%, 75%, and 87%) of the offers placed in the system.

The result regarding the offers is, in our belief, an indication on the importance of Market Makers in Ripple. However, to further investigate the matter we focus on the impact that Market Makers have in successfully delivering transactions. To do so we proceeded in the following way. We started from a stable snapshot, in terms of joins and leaves, of the Ripple network. We took the status of Ripple in February 2015 as the snapshot. Then, we extracted all payments submitted after the snapshot and successfully delivered until August 2015. The goal is to investigate how many of these payments would have been still delivered in the absence of Market Makers. So, we remove them and the exchange orders from the system and replay the extracted payments on the modified trust network. During this simulation we carefully handled the user balances by updating them after each successful payment. In addition, we also reflected in the modified trust network the updates happening on the real system to trust-lines (modification or creation of new ones). Overall, we analyzed more than 1.7M payments, 68.7% of which are cross-currency payments. The results are shown in Table II.

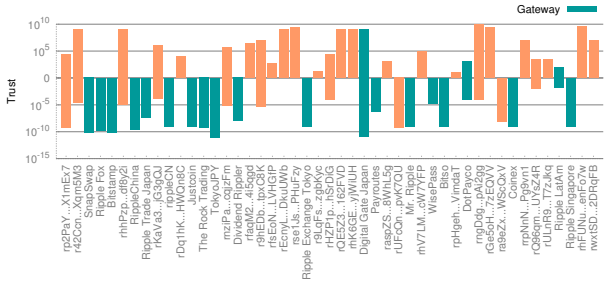
A first and expected observation is that, without Market Makers, all the cross-currency payments (68.7% of the total) fail. However, what is more striking is the impact of Market Makers on the delivery of single-currency transactions. In fact, as can be seen in Table II, almost 63% of single-currency transactions fail to be delivered without Market Makers. As a consequence, only 11.2% of the 1.7M payments submitted to the system make it to destination. This result shows that Market Makers are crucial for the Ripple exchange infrastructure. The offers they make are fundamental and their lack drastically impacts the successful delivery of both cross-currency and single-currency transactions. This result also suggests that Market Makers availability is critical and that attacking it has the potential to harm the entire system.

D. Gateways and users

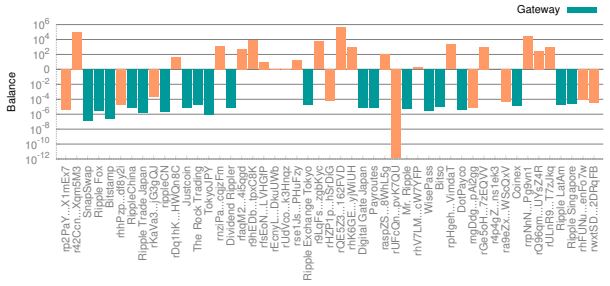
As of August 2015, Ripple counted more than 165K users, +55K of which were actively participating (*i.e.* by submitting transactions, creating offers, etc.). We continue our investigation by finding which are the most influential users of Ripple, *i.e.*, those that appear more frequently in payment paths. What we found is surprising. A handful of 50 peers contributed in about 86% of all the 10M multi-hop transactions in the system since its genesis. Our analysis shows that these users are either



(a) The 50 users that appear more frequently as intermediate hops in transaction paths. The green color hihglihts users that have publicly announced to be Gateways.



(b) Trust of the users in Figure 7(b) in August 2015. Positive trust is received by other nodes, negative trust is given to other nodes.



(c) The balance of the users in Figure 7(b) in August 2015 as the difference between credit and debit. The balance aggregated and shown in EUR.

Fig. 7. The 50 users that appear more frequently as intermediate hops in payment paths, their trust-relationships, and their balances. The X-axis shows a short version of the unique identifier of the user (or the name of the organization that endorse the user, in case of Gateways). The y-axis is logarithmic.

Gateways or tremendously active users (i.e. central exchange points of Ripple's backbone of trust). Figure 7(a) shows the number of times each of them serve as intermediate hop in the system. In the figure, we indicate the users that are known to be Gateways. At the time of the writing, these users are either publicly endorsed by a financial institution, or appeared in the crowd-sourced list of well-known Ripple Gateways[3].

This first analysis reveals an unexpected phenomenon. Among the most influential users, there are several of them that are not publicly announced Gateways—usually trustworthy as capable of providing evidence of their association with existing financial institutions. In fact, the very 2 most active nodes (`rp2PaY` and `r42Cc`) are not publicly announced Gateways. Yet, they appear in a number of paths that is almost an order of magnitude higher than that of the other users. A further investigation on the ledger showed that both users have been "activated" (i.e. received their first XRP payment) by a third Ripple user known as ∼`akhavr` in December 2013 and January 2014. This suggests a possible connection between ∼`akhavr` and the 2 most active nodes.

The next 3 most influential users are 3 well-known Gateways (SnapSwap[4], Ripple Fox, and Bitstamp) that trade the BTC, the EUR, the USD and the CNY. However, just 20 of the 50 Ripple most active users are Gateways. This means that there are several common users (users that are not Gateways) that bring a fundamental contribution to the routing of Ripple payments. For this reason, it is interesting to consider what does differentiate simple users from Gateways. A useful insight can be learned by looking at the trust relationships of each user. More in details, we considered the outgoing and incoming trust-lines of the 50 most active Ripple users (see Figure 7(b)).

A first observation is that Gateways are the ones with the highest amount of trust from other users. Most of them ($^{17}/_{20}$) do not declare trust towards the other users. These results support the intuition that Gateways are trusted parties of the system. They have proven to be associated with real-world financial institutions and they can provide financial guarantees about their assets. For this reason, they are trusted by common users of the system and, sometimes, by the other Gateways as well. Those who are not publicly declared as Gateways, conversely, are rarely trusted by the others. However, they need to trust at least one Gateway to join the payment network, creating a connection between them and the rest of the system.

To further investigate this topic we studied the balances of Ripple top users (see Figure 7(c)). Once again, Gateways and common users have completely different traits. Gateways exclusively display negative balances (*i.e.* debt), while most of the common users show positive balances (*i.e.* credit). This is reasonable, as Gateways are the Ripple equivalent of traditional banks, which collect money (and the associated trust) by the other users and give money (and the associated trust) to a smaller set of users.

[3]https://ripple.com/build/data-api-v2/#get-all-gateways

[4]SnapSwap announced the cessation of its activities in September 2015.

References

[1] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," May 2009. [Online]. Available: http://www.bitcoin.org/bitcoin.pdf

[2] S. Meiklejohn, M. Pomarole, G. Jordan, K. Levchenko, D. McCoy, G. M. Voelker, and S. Savage, "A fistful of bitcoins: Characterizing payments among men with no names," in *Proceedings of the 2013 Conference on Internet Measurement Conference*, ser. IMC '13, 2013.

[3] E. Androulaki, G. O. Karame, M. Roeschlin, T. Scherer, and S. Capkun, *Evaluating User Privacy in Bitcoin*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 34–51. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-39884-1_4

[4] A. Biryukov, D. Khovratovich, and I. Pustogarov, "Deanonymisation of clients in bitcoin p2p network," in *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*, ser. CCS '14. New York, NY, USA: ACM, 2014, pp. 15–29. [Online]. Available: http://doi.acm.org/10.1145/2660267.2660379

[5] G. F. Hurlburt and I. Bojanova, "Bitcoin: Benefit or curse?" *IT Professional*, vol. 16, no. 3, pp. 10–15, May 2014.
[6] D. Schwartz, N. Youngs, and A. Britto, "The ripple protocol consensus algorithm," *Ripple Labs Inc White Paper*, p. 5, 2014. [Online]. Available: https://ripple.com/files/ripple_consensus_whitepaper.pdf
[7] P. Todd, "Ripple Protocol Consensus Algorithm Review," 2015. [Online]. Available: https://raw.githubusercontent.com/petertodd/ripple-consensus-analysis-paper/master/paper.pdf
[8] F. Armknecht, G. O. Karame, A. Mandal, F. Youssef, and E. Zenner, *Ripple: Overview and Outlook*. Cham: Springer International Publishing, 2015, pp. 163–180. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-22846-4_10
[9] R. Forums, "Ripple Privacy - details on proxy payments or alternative?" [Online]. Available: https://forum.ripple.com/viewtopic.php?f=1&t=8304&p=57936
[10] P. Moreno-Sanchez, M. B. Zafar, and A. Kate, "Listening to Whispers of Ripple: Linking Wallets and Deanonymizing Transactions in the Ripple Network," *Proceedings on Privacy Enhancing Technologies*, vol. 2016, no. 4, pp. 436–453, July 2016.
[11] Y. de Montjoye, L. Radaelli, V. K. Singh, and A. S. Pentland, "Unique in the shopping mall: On the reidentifiability of credit card metadata," *Science*, vol. 347, no. 6221, pp. 536–539, Jan. 2015. [Online]. Available: http://dx.doi.org/10.1126/science.1256297
[12] D. Hide, "Ledger spam," 2014. [Online]. Available: https://web.archive.org/web/20160327021005/http://availableimagination.com/ledger-spam/
[13] A. Biryukov, D. Khovratovich, and I. Pustogarov, "Deanonymisation of clients in bitcoin p2p network," in *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*, ser. CCS '14, 2014.
[14] "Currency codes," Secretariat of the Maintenance Agency, Standard ISO 4217:2015, 2015.