On the Costs of Bitcoin Connectivity

By Ashish Gehani

Highly connected information technology systems typically have substantial economic value, making them attractive to resource-rich adversaries. Bitcoin is an example of such a system. The effect of high connectivity manifests in a number of orthogonal dimensions, each of which creates a different kind of security concern. We discuss each of them and possible mitigations.

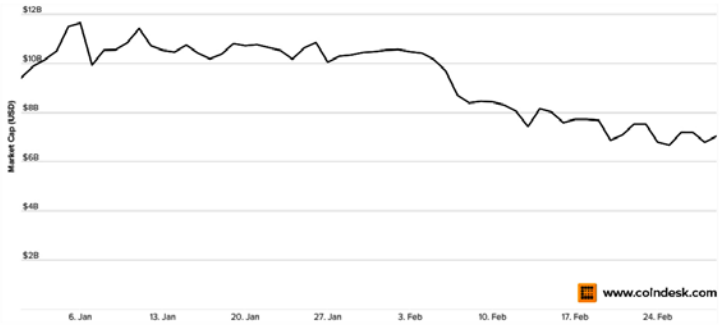

Figure 1 – Bitcoin market capitalization during the time frame when the largest exchange declared bankruptcy. Source: CoinDesk

Abstract

Highly connected information technology systems typically have substantial economic value. This makes them attractive to resource-rich adversaries, such as nation-states that may seek to exert an influence on particular participants or even the entire system. Bitcoin is an example of such a system. The effect of high connectivity manifests in a number of orthogonal dimensions, each of which creates a different kind of security concern. We discuss each of them and possible mitigations.

The cumulative value of existing Bitcoins crossed $10 billion in November 2013 [3], as illustrated in figure 1. Bitcoin can be used at numerous vendors, ranging from 75,000 mainstream companies [14] to an online black market [22]. In February 2014, a Bitcoin exchange declared bankruptcy, claiming $480 million in Bitcoin deposits had been stolen. While attacks by individuals have been studied both in academic literature and by software developers, the system's resilience to state-level adversaries has received less scrutiny. We consider problems that can arise when significant resources are brought to bear on attacking the highly connected Bitcoin network.

Background

David Chaum proposed the idea of digital cash over three decades ago [6], back in 1982. It used novel cryptographic constructs to imbue electronic transactions with the anonymity of physical cash, to ensure that digital cash could only be created by banks [7], and to prevent individuals from spending the same digital cash more than once [8]. Over the years, thousands of academic papers have been written on the topic, and hundreds of startups have been created to translate the ideas into practice [13].

The use of digital cash remained limited to a small set of technology enthusiasts as recently as 2008. In October of that year, Satoshi Nakamoto proposed Bitcoin [18], the first digital-cash scheme that was completely decentralized. It supports anonymous users, has no central mint, and distributes the effort of preventing double spending. In the wake of the financial recession and uncertainty in global economies, Bitcoin has grown rapidly. The Bank of England estimates that over 40 million Bitcoin accounts have been created worldwide [14].

By early 2013, the size of the Bitcoin market had crossed $1 billion [12]. Since late 2013, the size of the market has fluctuated in the range of $3-14 billion [3]. Bitcoin is now an acceptable form of payment at more than 75,000 mainstream companies. It can be used to buy computers from Dell, book hotels through Expedia, and pay for service from Dish TV. The users are more mainstream as well, with only 22 percent professing to be anarchists in 2014, down from 42 percent in 2013 [14].

Goal

While financial transactions between banks are regulated and monitored, little oversight exists for pseudonymous digital-cash systems such as Bitcoin [16]. Academics and developers have studied the system's vulnerability in the face of malicious participants. However, the analyses typically assume that a majority of the participants are cooperative. Our study relaxes the assumption to examine risks that arise when significant adversaries are involved, such as large multi-national organizations or state-level actors. We consider attacks where adversaries are afforded large amounts of computation and storage, control over the significant portions of or locations in the communications networks, and enough funding to employ numerous skilled developers for extended periods.

Understanding the resilience and vulnerabilities of the highly connected Bitcoin system to large-scale attacks has at least three benefits:

- It enables us to understand what adversaries, ranging from criminals or terrorists to hostile nations, may be able to accomplish through extant Bitcoin infrastructure.
- It allows us to understand how government agencies can manage the system to enforce the law.
- It lets us prioritize the research needed to understand the complex dynamics that result from the interplay between the technical, economic, and legal aspects of the Bitcoin ecosystem.

Bitcoin basics

We now describe the essence of Bitcoin. This will allow us to explain the types of risks that result when adversaries are sufficiently powerful. Every participant is identified by a public key (and knowledge of the corresponding private key). When a participant wishes to make a payment, he signs and broadcasts a transaction. This points to past transactions through which the payer has received sufficient funds to make the payment, and lists the payer, the payment amount, one or more payees, and the amounts to be paid to each of them. The payment is only considered valid after it has been added to the public ledger, which is a chronological list of all the transactions that have occurred since the beginning of Bitcoin in early 2009.

Any participant can aggregate a number of transactions into a block, compute the hash of the catenation of the public ledger and the block, and then search for a hash preimage in a predefined range (which serves as a proof of work). The resulting block and proof are appended to the public ledger, which is referred to as a blockchain. This process is known as mining Bitcoins because the first participant to successfully add the block receives a fee, as well any differences between the payments made by payers and the total received by the payees.

Consequences of connectivity

We describe the implications of high connectivity for the Bitcoin network. These occur in multiple dimensions—communication, computational, privacy, and logical control—as we explain below.

Communication dependence

Bitcoin was designed to work in distributed environments, where network connectivity is not always reliable.

Consequently, when participants are disconnected from each other, they continue to operate with miners in each partition extending their blockchain. When connectivity is restored, the blockchain that took the most work to create is accepted by all participants. This introduces a vulnerability in the case that an adversary controls the network infrastructure.

The more obvious vulnerability occurs when data transmissions are interfered with. For example, the Chinese government's Golden Shield project (also known as the Great Firewall of China) supports IP blocking, DNS redirection, packet and URL filtering, and resetting connections [24]. This would allow it to selectively filter transactions from particular participants, preventing them from reaching miners who should incorporate them in the blockchain. The result would be that the victims would be unable to receive or send payments, effectively having their assets frozen.

The less apparent weakness manifests when control information in the network is manipulated. Bitcoin allows any host to join the network. In order to bootstrap its connectivity to the network, a new node connects to a seed set of hosts, asks them for lists of their neighbors, and can then recursively contact them to enlarge its set of lists. However, an adversary that controls the network infrastructure can manipulate the reachable set of hosts. This results in a less subtle weakness—the ability for the adversary to scope the set of neighbors of a victim, and thereby determine which transactions are forwarded.

The more subtle concern results from the ability of the adversary to shape the topology of the network. Bitcoin relies on random neighbor selection to yield a low-degree, low-diameter network for efficient propagation of information [10]. This is consistent with well-understood properties of expander graphs. By filtering bootstrapping messages, the adversary can alter the shape of the network connectivity graph, as illustrated in figure 2. While this has limited impact for small networks, this results in severe scaling challenges for large networks. In particular, it can result in high latency for information propagation [1], reducing the usability of the system. This is of particular concern to unstructured peer-to-peer networks, such as Bitcoin, which can only handle seven transactions per second on average. (In contrast, Visa typically processes 2,000 transactions per second [19].)

Trust aggregation

While highly connected systems contain many nodes that are online at any given point in time, they must accommo-

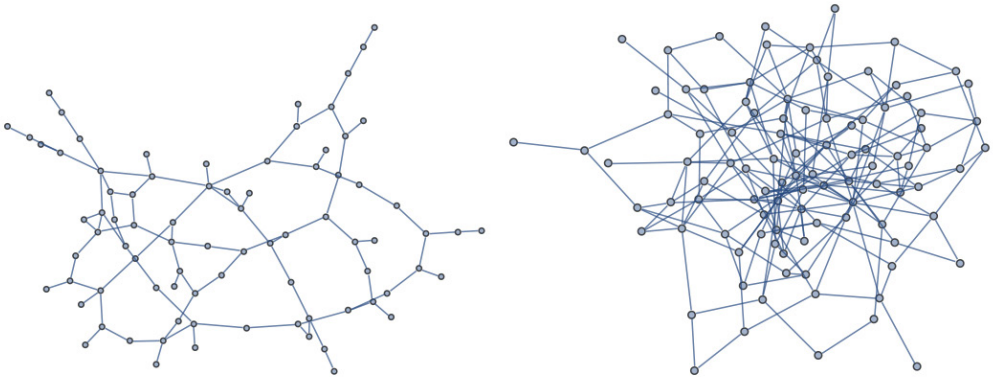

Figure 2 – Depictions of small Bitcoin networks, each constructed as a random graph. The second network contains about the same number of nodes as the first but with roughly double the number of links between neighbors. This results in lower susceptibility to node disconnection and shorter average diameter (and hence communication latency).

date nodes that may be temporarily unreachable. Often this results in trust being delegated to proxies that act on behalf of nodes, either in their absence or when the nodes themselves do not have sufficient resources (such as power, memory, or computation capability). However, over time trust starts to aggregate in a limited subset of the network, making the system vulnerable to targeted attacks.

In the context of Bitcoin this manifests in two forms. First, users may choose to use a wallet service where they authenticate to the service that then acts on their behalf. This is achieved via the user sharing his Bitcoin credentials with the wallet provider. Further aggregation of trust may occur if the wallet provider also serves as an exchange (for converting between bitcoin and national currencies). Users may opt for this arrangement due to its convenience. The bankruptcy of Mt. Gox [11] demonstrated the weakness of such pooling of trust.

Devices with resource constraints can utilize a light version of the Bitcoin software. This has the advantage that it limits its activity to checking transactions related to the user, rather than validating all activity in the system. For example, instead of retaining the entire blockchain (that is currently over 26 GB), it can maintain just the headers (that currently take a little over 23 MB) [19]. However, this has the effect of becoming a second avenue through which trust becomes more centralized. This is because users of light Bitcoin clients rely on others to validate the integrity of the rest of the blockchain.

If an adversary is able to command more computational power than the rest of the participants together, he can exert control over which variant of the blockchain is accepted by the Bitcoin network. In this case, the adversary can select a particular blockchain and extend it faster than others can extend what was previously the legitimate blockchain. The consequence is that all participants will accept the adversary's blockchain. To see that this is a real threat, note that the Bitcoin community recently objected to the pooling of resources by GHash.IO members due to this concern [5].

Privacy violations

The presumed anonymity of participants in the Bitcoin network is based on the fact that they are only identified by their public keys. However, large amounts of data are collected by programs such as Upstream and PRISM [23]. This auxiliary information can be used to de-anonymize some public keys. Further, all transactions in the entire history of the Bitcoin network are recorded in a single public ledger that is necessarily available to every member. It is has been shown that this can be leveraged to further de-anonymize other participants' public keys [17].

Metcalfe's law [21] observes that the value of a network grows as the square of its size, measured in nodes. A corollary to this is that an adversary seeking to de-anonymize participants gets quadratically increasing opportunity to do so as the size of the network grows. More concretely, when the number of Bitcoin users increases by a factor of φ, the number of transactions between different users is expected to grow by a factor of $O(\varphi 2)$. The chance that the adversary is able to subvert the privacy of any transaction grows commensurately.

In practice, the Bitcoin network consists of nodes that run the full protocol while others run the light version. We model the total number of nodes as N. The fraction of the nodes running the full protocol is denoted by f. Therefore, on average there are $(f\,N\,)^2$ edges between nodes running the full protocol. Since $(1 - f\,)N$ must then be running the light protocol, the ratio of light nodes to full nodes is $\frac{(1-f)N}{fN} = \frac{1-f}{f}$. Thus, there will be $\frac{1-f}{f}(fN) = (1-f)N$ edges between nodes running the light and full versions. In the above model, we can consider the fraction of transactions whose privacy can be violated. This can be viewed as three cases: first, when the adversary subverts a fraction s of the light software—that is, $\frac{s(1-f)N}{(1-f)N+(fN)^2}$ of all transactions will be de-anonymized; second, when the adversary subverts the same fraction of the software running the full protocol, that is, $\frac{s(fN)^2}{(1-f)N+(fN)^2}$ of the transactions will be affected; and third, when the adversary subverts the same fraction of both types, that is $\frac{s(1-f)N}{(1-f)N+(fN)^2} + \frac{s(fN)^2)}{(1-f)N+(fN)^2} = s$. Asymptotically, (as $N \to \infty$) the fraction of transactions with privacy violations goes to 0, $O(s)$, and $O(s)$. The cost of the third option is bounded below by the cost of the second option, while the benefits are comparable. We can therefore conclude that the nodes running the full protocol are most attractive as targets of a well-provisioned adversary.

Logical coupling

The reference code is developed jointly in open fora by members of the Bitcoin community. However, sufficiently motivated adversaries could infiltrate the group over time. The potential for damage can be gauged by studying past incidents resulting from software flaws. In August 2010, an integer overflow vulnerability was exploited to bypass a verification

check [9]. Failure to discover it would have allowed two accounts to fraudulently be credited with 184 billion Bitcoins.

The homogeneity in the deployed code base leaves the system brittle to attacks. Again, we can understand the risk by considering a case that occurred without malice. In March 2013, a new version of the Bitcoin software was released. The use of a new database inadvertently increased the number of transactions that could be reported at once [4]. The discrepancy resulted in older versions of the software rejecting transactions that the newer version accepted. The resulting operational disruption caused the value of Bitcoins to drop by 33 percent [15].

The frailty is the direct consequence of the interplay between high connectivity and tight logical coupling. The software engineering regime used for developing Bitcoin code does not support graceful operation in the face of discrepancies. For example, the ecosystem currently avoids fixing bugs that would disconnect older clients if the bug does not affect correctness.

Mitigation

Bitcoin is organized as an unstructured peer-to-peer overlay. At the end of the 1990s and in the early 2000s, significant research effort was invested in creating structured peer-to-peer systems. The lessons learned can be applied to use distributed hashtables or alternate approaches to reduce the scaling problems. If node churn is low, such solutions are likely to improve its stability. Reducing the aggregation of trust requires users to adopt wallets that maintain credentials on their devices rather than at proxies. Offline processing of transactions on their behalf can occur through third-party escrow. Bitcoin's scripting capability allows these to be enforced using its multi-signature primitive without trusting intermediaries. Further, as the storage and computational power of mobile devices grows, the need to run light versions of the protocol will decrease. This will reduce the dependence on external verifiers.

The current approach for de-anonymizing users of Bitcoin first clusters the large number of anonymous users based on idioms of use [17] and then connects clusters to real-world identities using associations determined out of band (by transacting with known entities, for example). Changing the patterns in which transactions are carried out, or using intermediaries that mix the inputs and outputs of many users into single transactions can add layers of privacy.

Finally, an increasing number of implementations of the Bitcoin protocol are being deployed. As the diversity in the running codebase grows, so will the resources needed to corrupt all versions of the software. Further, the chance that at least one implementation's other developers will notice grows exponentially.

Related work

Although Bitcoin is a relatively recent construct, by November 2013 the cumulative value of existing Bitcoins had crossed $10 billion. This has motivated numerous analyses, both in academic literature [2] as well as in practice. They can be categorized into prospective studies consisting of theoretical attacks that have been proposed and retrospective ones, which have been undertaken after failures that manifested in practice.

The first group are typically attacks against the steps used for minting, transacting, and validating Bitcoins. The threat model usually assumes that an individual or small group of colluding entities aim to gain more Bitcoins than they are entitled to. Often the attacks attempt to exploit race conditions that result from the fact that Bitcoin is a distributed protocol. For example, a *Finney attack* occurs when an entity validates a transaction but delays reporting it to the network and colludes with the payer who double-spends the amount to another payee. After the payer has received goods or services from the second payee, the original validated transaction is released to the network, causing the second payee's payment to be rejected. In practice, the cryptographic protections, network dynamics, and economic incentives have sufficed to thwart these types of attacks.

In contrast to the first group, the retrospective analyses focus on software flaws discovered in the field. Even though the Bitcoin system itself is decentralized, over time some participants began to aggregate trust by depositing their Bitcoins in a few exchanges. In February 2014, the Mt. Gox Bitcoin exchange filed for bankruptcy after losing $480 million [11] of customer deposits. It blamed the losses on theft made possible

by the fact that multiple Bitcoin transaction identifiers could be derived from the same signature. (This property is known as *transaction malleability*. It resulted from OpenSSL treating signatures with a different number of leading zeros as equivalent [20].) However, such attacks can be guarded against by avoiding central points of failure—that is, participants maintain custody of their own digital cash.

Conclusion

We discussed four security concerns that arise from the high connectivity of the Bitcoin network and the weaknesses that result from each. First, we considered the consequence of relying heavily on communication network connectivity. Second, we covered the effect of connecting nodes with asymmetric (superior) computational power. Third, we reported on the fragility of privacy in the face of high connectivity. Fourth, we highlighted the concerns that come from tight logical connectivity (as occurs with homogeneity in the operational code base). Finally, we discussed possible mitigations.

Acknowledgment

This work was partially funded by the US National Science Foundation (NSF) under Grant IIS-1116414 and Department of Homeland Security (DHS) Science and Technology Directorate. The views and conclusions contained herein are the author's and should not be interpreted as representing the official views of DHS, NSF, or the US government.

References

[1] William Acosta and Chandra, "Improving Search Using a Fault-tolerant Overlay in Unstructured Peer-to-Peer Systems, *IEEE International Conference on Parallel Processing*, 2007.

[2] Joseph Bonneau, Andrew Miller, Jeremy Clark, Arvind Narayanan, Joshua Kroll, and Edward Felten, "Research Per- spectives and Challenges for Bitcoin and Cryptocurrencies," *36th IEEE Symposium on Security and Privacy*, 2015.

[3] "Bitcoin Market Capitalization," *Blockchain Info* – https://blockchain.info/charts/market-cap.

[4] Vitalik Buterin, "Bitcoin Network Shaken by Blockchain Fork," *Bitcoin Magazine*, 12 March, 2013 – http://bitcoinmagazine.com/3668/bitcoin-network-shaken-by-blockchain-fork/.

[5] Michael Casey and Paul Vigna, "BitBeat: Mining Pool Rejects Short-Term Fixes to Avert '51% attack,' " *The Wall Street Journal*, 16 June, 2014 – http://blogs.wsj.com/moneybeat/2014/06/16/bitbeat-a-51-attack-what-is-it-and-could-it-happen/.

[6] David Chaum, "Blind Signatures for Untraceable Payments," *Advances in Cryptology*, Springer, 1983.

[7] David Chaum, "Security without Identification: Transaction Systems to make Big Brother Obsolete," *Communications of the ACM*, Vol. 28(10), 1985.

[8] David Chaum, Amos Fiat, and Moni Naor, "Untraceable Electronic Cash," *Advances in Cryptology, Lecture Notes in Computer Science*, Vol. 403, Springer, 1990.

[9] CVE-2010-5139, National Vulnerability Database, National Institute of Standards and Technology, 6 August, 2012 – https://web.nvd.nist.gov/view/vuln/detail?vulnId=CVE-2010-5139.

[10] Christian Decker and Roger Wattenhofer, "Information Propagation in the Bitcoin Network," 1*3th IEEE International Conference on Peer-to-Peer Computing*, 2013.

[11] Carter Dougherty and Grace Huang, "Mt. Gox Seeks Bankruptcy after $480 Million Bitcoin Loss," *Bloomberg*, 28th February, 2014 – http://www.bloomberg.com/news/2014-02-28/mt-gox-exchange-files-for-bankruptcy.html.

[12] Rip Empson, "Bitcoin: How an Unregulated, Decentralized Virtual Currency Just Became a Billion Dollar Market," *TechCrunch*, 28 March, 2013 – http://techcrunch.com/2013/03/28/bitcoin-how-an-unregulated-decentralized-virtual-currency-just-became-a-billion-dollar-market/.

[13] Ian Grigg, "A Very Fast History of Cryptocurrencies BBTC – Before Bitcoin," *Financial Cryptography*, 8 April, 2014 – https://financialcryptography.com/mt/archives/001491.html.

[14] Olga Kharif, "Bitcoin Economy Widens as Parents Pay Digital Allowance," *Bloomberg*, 24th September, 2014 – http://www.bloomberg.com/news/2014-09-25/bitcoin-economy-widens-as-parents-pay-digital-allowance.html.

[15] Timothy Lee, "An Illustrated History of Bitcoin Crashes," Forbes, 11 April, 2013 – http://www.forbes.com/sites/timothylee/2013/04/11/an-illustrated-history-of-bitcoin-crashes/.

[16] "Regulation of Bitcoin in Selected Jurisdictions," *Law Library of Congress, Global Legal Research Center*, January, 2014 – http://www.loc.gov/law/help/bitcoin-survey/regulation-of-bitcoin.pdf.

[17] Sarah Meiklejohn, Marjori Pomarole, Grant Jordan, Kirill Levchenko, Damon McCoy, Geoffrey M. Voelker, and Stefan Savage, "A Fistful of Bitcoins: Characterizing Payments among Men with No Names," *13th ACM Internet Measurement Conference*, 2013.

[18] Satoshi Nakamoto, "Bitcoin: A Peer-to-Peer Electronic Cash System," *Cryptography Mailing List,* 31 October, 2008 – https://bitcoin.org/bitcoin.pdf.

[19] Arvind Narayanan, Joseph Bonneau, Edward Felten, Andrew Miller, and Steven Goldfeder, "Mechanics of Bitcoin," *Bitcoin and Cryptocurrency Technologies*, 2015.

[20] Anders Nilsson, "The Troublesome History of the Bitcoin Exchange MtGox,"14 February, 2014 – https://anders.io/the-troublesome-history-of-the-bitcoin-exchange-mtgox/.

[21] Carl Shapiro and Hal Varian, *Information Rules*, Harvard Business Press, 1999.

[22] "Silk Road," Wikipedia, retrieved on 1 October, 2014 – https://en.wikipedia.org/wiki/Silk_Road_(marketplace).

[23] Craig Timberg and Ellen Nakashima, "Agreements with Private Companies Protect US Access to Cables' Data for Surveillance," *The Washington Post*, 6 July, 2013 – https://www.washingtonpost.com/business/technology/agreements-with-private-companies-protect-us-access-to-cables-data-for-surveillance/2013/07/06/aa5d017a-df77-11e2-b2d4-ea6d8f477a01_story.html.

[24] Jonathan Zittrain and Benjamin Edelman, "Empirical Analysis of Internet Filtering in China," *OpenNet Initiative*, 20 March, 2003 – http://cyber.law.harvard.edu/filtering/china/appendix-tech.html.

About the Author

Dr. Ashish Gehani of SRI International holds a BS (Honors) in Mathematics from the University of Chicago and PhD in Computer Science from Duke University. His research focuses on data provenance and security. He may be reached at gehani@csl.sri.com.