Bitcoin Concepts, Threats, and Machine-Learning Security Solutions

MOHAMED RAHOUTI, KAIQI XIONG (Senior Member, IEEE),
AND NASIR GHANI(Senior Member, IEEE)

Uuganbayar Gankhuyag

2019-04-02

Abstract

- Bitcoin is a peer-to-peer payment system and digital currency introduced as open source software in 2009 by pseudonymous developer Satoshi Nakamoto where transactions take place among users without any intermediary.
- Bitcoin transactions are performed and verified by network nodes and then registered in a public ledger called blockchain, which is maintained by network entities running Bitcoin software.
- Bitcoin is most successful cryptocurrency and widely traded across the world but Bitcoin's popularity grows, many security concerns are coming to the forefront.
- Bitcoin security depends upon the distributed protocols-based stimulant compatible proof-of-work that is being run by network entities called miners, who are anticipated to primarily maintain the blockchain (ledger).

Abstract

- This paper proposed an intensive study that explores key security concerns.
  - Presenting a global overview of the Bitcoin protocol as well as its major components.
  - Detail the existing threats and weaknesses of the Bitcoin system and its main technologies including the blockchain protocol
  - Discuss current existing security studies and solutions and summarize open research challenges and trends for future research in Bitcoin security.

Introduction

- Bitcoin was originally introduced in 2008. Since then, it has emerged as the most successful cryptographic currency among many competitors, boosting the economy with billions of dollars a few years after being launched.
- Bitcoin differs from traditional on-line banking, there's no third party in this system. Every transaction through bitcoin happens only between users, without any intermediary, e.g., such as an e-bank, a notary, or any other traditional on-line financial service provider.
- Bitcoin is increasingly drawing public attention and moving more and more customers towards using this payment system in a variety of business.
- However, the security, confidentiality and reliability of Bitcoin have also been hot topics. Additionally, to guarantee a reliable and trusted distributed system of monetary transactions, it is critically important for all Bitcoin holders and operators to have a safe environment of monetary operations as well as personal property protection.

Introduction

- Blockchain provides the fundamental framework for all kinds of Bitcoin's operations. In particular, it provides a novel decentralized consensus scheme that stores transactions, money transfers, and any other data records in a secure manner without any involvement from third party authorities.

Introduction

- Bitcoin system and its network infrastructure have been proven vulnerable to a tremendous amount of malicious activities and attacks in the past.
- There have several research studies range from anomaly detection to market return and volatility forecasting in the Bitcoin system.
- Several studies have also focused on utilizing ML techniques for anomaly detection in Bitcoin networks, such as fraud detection and anomalous transactions.
- For example
  - Using unsupervised learning methods to deal with anomalous transactions in Bitcoin system
  - In order for us to fix or even isolate malicious parts of the network until they are debugged, various ML techniques, such as support vector machine (SVM) and clustering, can be deployed to help with identifying those parts that behave abnormally.

Introduction

- In this work, proposed an intensive survey that mainly focuses on the deployment of ML techniques for security threat detection and mitigation in Bitcoin and blockchain systems, along with an analysis of their related concepts.

Bitcoin infrastructure and design

Decentralization

- Bitcoin is the very first distributed crypto-currency system and it is a fully decentralized digital currency system where the monetary power is not controlled by any party.
- This decentralized architecture still has several major limitations:
  - The transactions ledger needs to be publicly preserved by every single node.
  - Ledger transactions need to be checked and legitimized by a distributed entity but not by a centralized authority or party.
  - Unlike centralized economic systems, new bitcoins can be generated by any connected entity.
  - Exchange operations of bitcoins values are completely dynamic and no centralized control is required to handle such operations.

Transactions and scripts

- Transaction format:

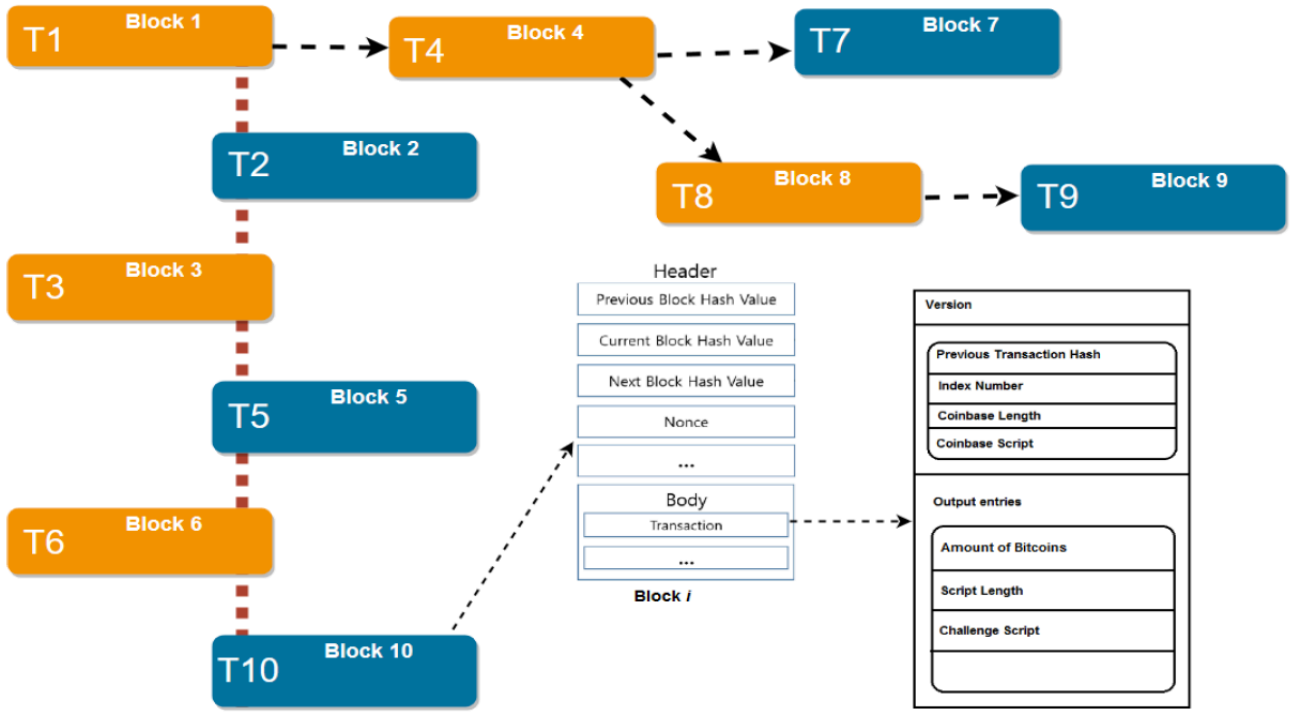

FIGURE 4. A depiction of consensus model of the blockchain with a block and transaction structures.

Transactions and scripts

- Transaction Script:

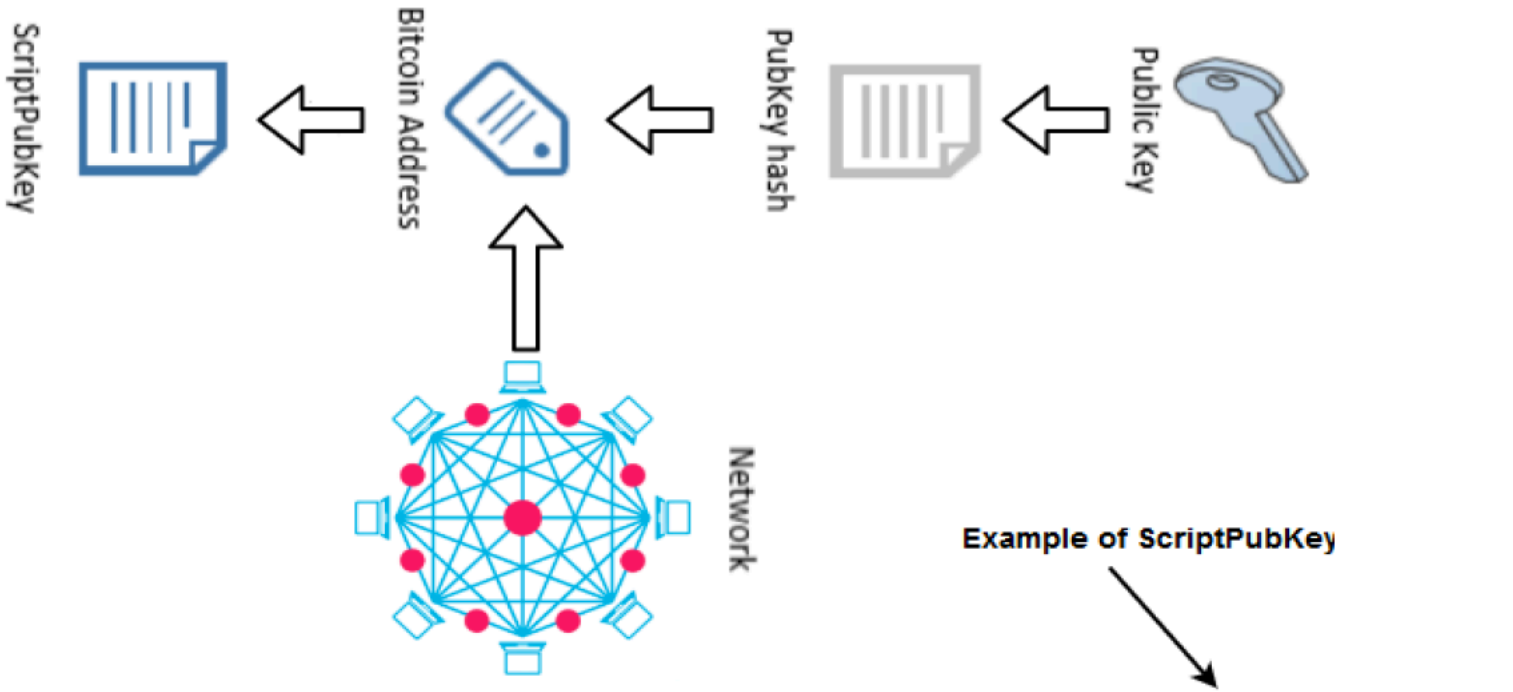

FIGURE 3. ***ScriptPubKey*** **generation from Bitcoin addresses to identify recipients.**

Transactions and scripts

- From transaction to ownership:

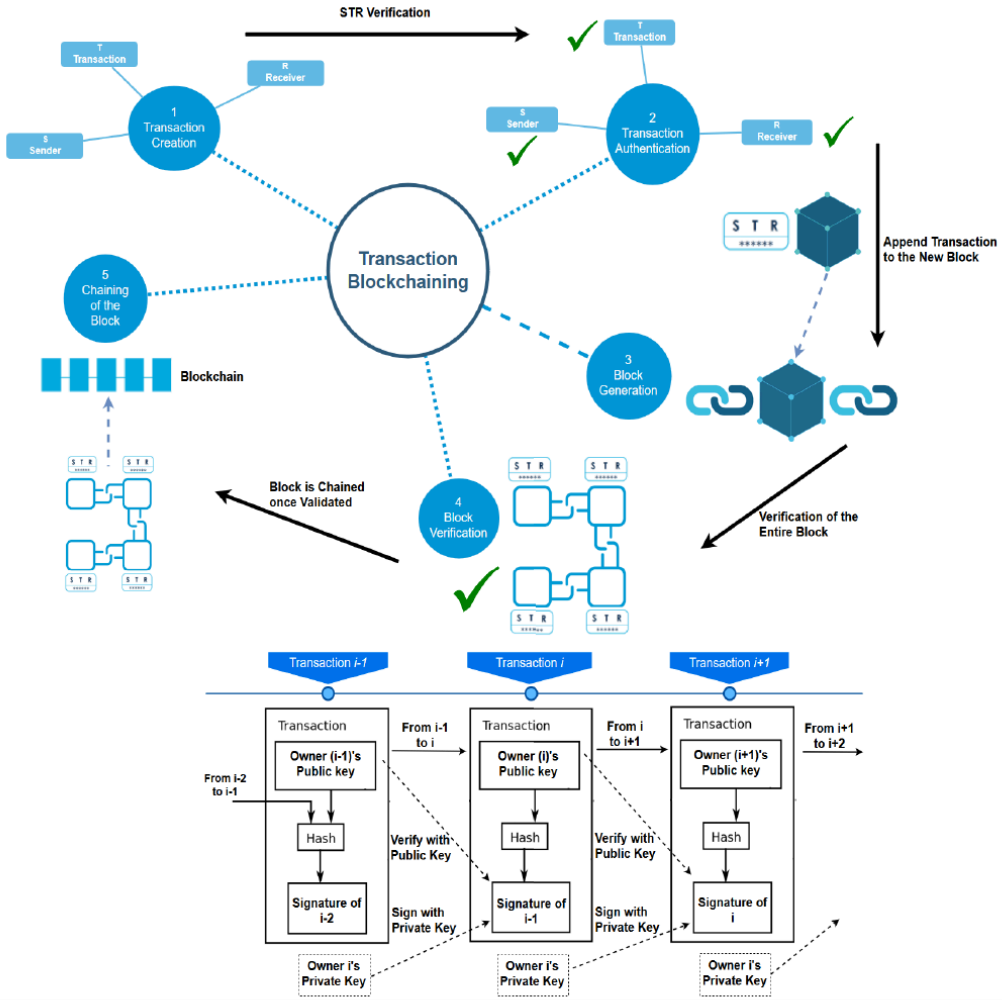

FIGURE 2. Creation and addition of blocks and a transactional process in a blockchain.

Blockchain consensus protocol and mining

- Blockchain is basically a public and append-only-based structure that saves a transaction history in the distributed Bitcoin system in the form of individual blocks by using Merkle tree. if miner nodes validate them through solving a difficult proof-of-work(PoW) puzzle, a new block is successfully appended to the blockchain.
- Operation of appending a new block will be as follows:
  - A miner node will append the new block in its local blockchain and advertise the solution once the valid hash value is determined.
  - if the advertised solution is valid , miner nodes will rapidly verify the received solution of a valid block and update their local blockchain. otherwise, it will be discarded.
- PoW algorithm is used to enable the Bitcoin system to realize a completely distributed consensus. This PoW-based consensus protocol implies a few key rules:
  - Transactions are enabled to spend only unspent valid outputs.
  - Output and/or input entries are rational.
  - Spent inputs are assumed to have correct signatures.
  - Outputs of a coinbase that are a unique type of bitcoin transaction created by a miner node.

Peer to peer communication network

- Bitcoin uses an unstructured P2P-based network along with a reliable TCP-based connection transport.
- An unstructured network suits Bitcoin due to the need for a rapid distribution of data to attain the blockchain consensus.
- The benefit of using an unstructured-based P2P network in Bitcoin is to allow for fast data distribution in the entire Bitcoin network.

Resiliency challenges

Blockchain consensus protocol and mining

- In the PoW based consensus algorithm, the users require no authentication to join the network.
- It makes the Bitcoin consensus model extremely scalable regarding supporting thousands of network nodes.
- However, PoW based consensus is vulnerable to “51%” attacks, in which an attacker has control over 51% of the mining power in the network.
- Under this threat scenario, attackers might create their own transaction block or even fork the local blockchain to converge with the primary blockchain at a later time.

Double spending

- Double spending means spending the same coin twice.
- For example
  - Entity E1 can send a transaction block to entity E2, and then to entity E3, with the same bitcoin.

Mining pool vulnerabilities

- Bitcoin mining pools are an aggregation of resources used by mining entities that share their own processing power in the P2P network.
- Pool hopping attack
  - Attacker attempts to execute a continual and uninterrupted analysis over the amount of shares transmitted by contiguous miner nodes to their pool manager for the purpose of exploring a new transaction block.
- Bribery attack
  - Attacker offers payments to existing miners to deviate from the default protocol and mine on the attacker’s branch.

Vulnerabilities in cryptographic applications

- In Bitcoin, unlike classical public key cryptography, clients are held accountable for their private keys, and therefore each client is responsible for producing their own private keys and maintaining it (rather than a third-party).
- If the private keys are lost, the owning clients will not be able to access their own digital assets in a Bitcoin network.

Goldfinger attacks

- The intention of the majority of mining entities is to explicitly break down Bitcoin network stability.
- For instance, such an attack scenario could occur when a connected entity in the network attempts to harm Bitcoin in order to avert a competition with its own currency.

Feather-forking threat

- Feather forking is a subtle modification of the more well-known punitive forking attack.
- Punitive forking consists in excluding someone from the blockchain through a systematic and unbounded forking operation with respect to those blocks which contain transactions originating from the blacklisted people.
- Although this attack is very dangerous, it is hard to carry out when the attacker does not hold the majority of the hash power of the whole system.

Machine learning-based efforts and countermeasures against threats

- In this paper, mostly concentrate on recent ML solutions and proposals to address the problem of revealing malignants and suspicious activities and actions in Bitcoin and blockchain.
- For example
  - Smith et al. used particular clustering methods to capture malignant activities in a network and were able to classify legitimate system users separately from malicious users, i.e., via k-means-based clustering along with self-organizing maps to design a detection solution.

Machine learning-based efforts and countermeasures against threats

- ML techniques have also been used in several studies to address the aforementioned security threats.
- For example
  - Pham and Lee tried to detect abnormal behaviors in Bitcoin transaction networks by using multiple unsupervised ML techniques, such as k-means clustering and Unsupervised Support Vector Machine (SVM) on two Bitcoin transactions graphs. Specially, one graph presents client transactions as entities and the other clients as entities.

Machine learning-based efforts and countermeasures against threats

- Harlev et al. introduced a novel technique to decrease anonymity in Bitcoin networks using a ML-based supervised approach to predict unidentified network nodes.
- Hirshman et al. also proposed an unsupervised ML technique to detect abnormal conducts in the Bitcoin transaction network.

Machine learning-based efforts and countermeasures against threats

- Monamo et al. investigated the use of the trimmed k-means method to capture deceitful behaviors in the Bitcoin network.
- Also, Monamo et al. described various fraud activities in the Bitcoin network from both local and global perspectives by mean of trimmed k-means and kd-trees.
- Bartoletti et al. utilized data mining and ML-based methods to investigate and capture Bitcoin addresses relevant to Ponzi schemes. In such particular attacks, an adversary shares a falsified transaction block which can threaten investment in Bitcoin.

Machine learning-based efforts and countermeasures against threats

- Moreover, Yin and Vatrapu provided an accurate estimation of the portion of cybercriminal nodes in the Bitcoin network.
- They utilized a large Bitcoin dataset composed of more than 800 observations classified into about 12 categories including observations relevant to cybercrime and uncategorized ones.

Machine learning-based efforts and countermeasures against threats

TABLE 1. Taxonomic classification of proposed security solutions using machine learning.

Reference	Year	Method	Contribution
Pham and Lee [51]	2016	Unsupervised ML techniques	Anomaly detection in the Bitcoin network where clients and transactions are considered suspicious
Yin and Vatrapu [69]	2017	Supervised ML technique	Demonstrate how large is the share of cybercrime-related nodes in the Bitcoin network and nodes/addresses relevant to malicious activities
Monamo, et al. [48]	2016	ML-based multifaceted approach	Bitcoin fraud detection where the fraud is studied from both global and local perspectives using trimmed *k-means* and *kd-trees*
Zambre and Shah [70]	2013	ML-based clustering and classification	Identifying peculiar properties of clients performing anomalous behavior through clustering clients who exhibit suspicious activities
Portnoff, et al. [53]	2017	ML classifiers	Designing a ML-based classier to differentiate between ads posted by the same author vs. different authors and also a linking technique, which uses leakages from the Bitcoin network and sex ad site to link a set of sex ads to Bitcoin transactions and public wallets
Harlev, et al. [31]	2018	Supervised ML technique	Minimizing anonymity in the Bitcoin network through a Supervised ML technique deployment to predict the type of unidentified nodes/entities
Bartoletti, et al. [10]	2018	Data mining technique	Detecting Bitcoin addresses in the network that are related to a Ponzi scheme
Zhdanova, et al. [71]	2014	ML-based micro-structuring technique	Revealing fraud chains through developing a new technique for detection of fraud chains in Mobile Money Transfers (MMT) systems
A. Bogner [11]	2017	ML adoption for graphical threat detection	Providing human operators with an intuitive way to derive insights about the blockcahin system through in-gathering of the system's features into a group of characteristics which are graphically rendered
Remy, et al. [56]	2017	ML-based network analysis techniques	Tracking the clients' activities in the Bitcoin system using community detection on a network of weak signals
Hirshman, et al. [32]	2017	Unsupervised ML technique	Applying ML methods on a Bitcoin transactions dataset to explore anonymity guarantees in the network by clustering the dataset
Monamo, et al. [47]	2016	Unsupervised ML technique	Exploring the use of trimmed *k-means* for simultaneous objects clustering and fraud detection in a multivariate configuration in order to detect fraudulent behaviors in blockchain blocks
Pham and Lee [52]	2016	ML-based superior method	Detecting anomalies in Bitcoin networks through exploring clients where transactions seem to be most dubious where a malignant behavior is regarded as a proxy for dubious activity

TABLE 2. Taxonomic classification of proposed security solutions using machine learning.

Reference	Year	Method	Contribution
Kurtulmus and Daniel [38]	2018	Intelligent problem solving based on ML aspect	Propose DanKu, a byproduct protocol utilizing the distributed nature of smart contracts along with a ML-based intelligent problem solving to solve crowd-sourcing funds for computational research and to efficiently provide a new marketplace without a need for a middlemen
S. Dey [25]	2018	Supervised ML algorithm and algorithmic game theory	Presenting a methodology based on intelligent software agents that supervises the stakeholders' activities in the Bitcoin system in order to detect abnormal behaviors using a supervised ML algorithm along with algorithmic game theory
Liu, et al. [42]	2017	Immune ML-based model	Proposing a ML-based solution to capture double-spending activities in fast Bitcoin payments, where the solution consists of several immune-based blockchain nodes that embrace a detection component
Shaukat and Ribeiro [60]	2018	ML-based Ransomware detection	Proposing a ML solution based on an extensive analysis of Ransomware dataset families in order to provide a layered defense system against cryptographic ransomware in a cryptocurrency system
K. Baqer [8]	2016	Clustering-based method/stress test	Introducing an empirical analysis of a spam campaign/stress test that caused DoS attacks. Namely, a clustering based technique is deployed to detect spam transactions in the Bitcoin network
COINHOARDER [33]	2018	NLP and ML-based phishing ring DNS style detection	Introducing a detection scheme for phishing ring DNS style through ML and NLP techniques where the detection mechanism relies on the observation of the newly registered and/or launched domains
Ermilov, et al. [44]	2017	Address-based clustering methods	Introducing an off-chain information solution along with blockchain information for Bitcoin address separation and classification in order to detect and filtrate errors in users' inputed data and therefore evade insecure Bitcoin usage patterns

Future research direction

- Scalability and blockchain protocol
  - Miner entities can act in a selfish manner by continuing to carry particular blocks of transactions and unleashing them whenever they wish in order to increase their revenue.
  - Such selfish activities will likely create a game theoretic challenge among egocentric miner nodes and the network.
- Cryptography techniques
  - The deployment of clustering techniques based on specific thresholds are designed to address a broad range of threats. However, there are only a few number of methods that use string searching filters for protecting wallets.
- Incentives for miners
  - Malignant miners could conduct an illegitimate activity through the Bitcoin network to acquire extra awarding coins, which will augment the amount of sincere entities in the network. Hence, it is very important to address this challenge by making miner nodes settle to a currency in this cryptocurrency network.
- Preventing backtracks
  - Smart contracts are of a particular interest to financial applications, which incarnate self-enforcing-based contracts entities in financial networks such as Bitcoin. Therefore, this concept could be adopted in Bitcoin infrastructures as the blockchain system drives out the need to depend on authenticated third parties to handle contracts. However, Bitcoin support for such contracts is still very restricted.

Conclusions and My opinions

- Blockchain has demonstrated its potential to transform and mutate classical financial and transactional market models with its key distinctive features, including decentralization, anonymity, and auditability.
- Bitcoin infrastructure is established using the PoW and consensus protocols to protect client transactions and activities, these protocols themselves remain a point of vulnerability and exploitation for cyber threats, starting from the sniffing of network packets to the double spending activities.

Conclusions and My opinions

- In this work
  - First presented an overview of the Bitcoin network and related blockchain technologies and protocols.
  - Then analyzed the common blockchain consensus protocol followed by a discussion of its characteristics, including advantages and limitations.
  - Next presented a taxonomic classification and precise discussion of existing solutions and proposals that use machine learning (ML) techniques to solve common security threats and anomalous behaviors in Bitcoin networks and blockchain.

Thank you