Columbia-IBM Center for Blockchain and Data Transparency
Workshop: IoT and Privacy
November 11th, 2019

Introduction

The Columbia-IBM Center for Blockchain and Data Transparency organized a workshop on November 11th, 2019 on the topic of IoT and privacy. And this whitepaper is a summary of the findings of the workshop. It was organized in 4 sessions that are all related to Internet of Things and privacy: (1) Internet Fraud and Abuse, (2) Supply chain management using blockchains, (3) Blockchains as an authentication mechanism for IoT devices, and (4) IoT edge computing based on blockchains. In the first session, the speakers discussed issues around widespread fraud and abuse on the Internet, specifically focusing on secure hosting and security issues around credential theft. IoT devices will operate at a scale that is an order of magnitude larger than humans on the Internet, and hence attention is needed to secure the infrastructure. In the next session, researchers from IBM presented case studies of the successful use of blockchains in several supply chain management systems deployed out in the field. Both the success stories as well as remaining technical challenges were discussed. The next session moved back to the issue of security and privacy in IoT devices, looking at the privacy issues of Internet-connected televisions and the various kinds of data thefts and leaks that are out there in the wild. Another talk focused on the secure and efficient replication of data across blockchain nodes, which is necessary for a distributed system to scale. Finally, the last session addressed the issue of scalability and trust in smart contract management via blockchains, and techniques to speed up the transaction processing systems in blockchains. Detailed discussion on the sessions and the individual talks follow.

Platforms in Everything: Analyzing Ground - Truth Data on the Anatomy and Economics of Bullet-Proof Hosting

Speakers:
Asaf Cidon, Assistant professor of Electrical Engineering and Computer Science at Columbia University, Session Chair

Damon McCoy, Assoc. Prof. Computer Science and Engineering, NYU Tandon School of Engineering presented the paper for:
Arman Noroozian of Delft University of Technology, Delft Netherlands
Assistance from the Dutch National High-Tech Crime Police and the Dutch Ministry of Economic Affairs

Abstract:
In this talk, McCoy examines multiple potential weak spots that might be used to negatively impact sites that use the newer, so-called, Bullet Proof Hosting (BPH) to host abusive material. He describes the transition from the older model of hosting such sites to the new more agile and diffuse BPH model. Using information from a seized site in the Netherlands, he is able to examine the internal functioning of the business including supply characteristics, upstream providers, demand characteristics, customer and merchant numbers, payment methods, finances, and profits. His conclusion is that, currently, the best hope is to tune detection models based on what we know, thereby increasing the site's operating costs and cutting into the largely thin profit margins forcing BPH sites out of business.

Synopsys:
The traditional or old way that abusive sites (e.g., hosting phishing, bots, child pornography, running anonymous cyberattacks) were hosted was through ISPs where ownership and management, including ownership of hardware and ownership of the autonomous system numbers, was centralized. These hosts were usually de-peered with somewhat poor access to the internet. This model is concentrated in a few bad ISPs, such as McColo Corporation, run by the Russian Business Network, and Troyak. This model only works in places where it is difficult to get court orders to shut them down although hosting sites are relatively easy to find and block.

The new model, or Bullet Proof Hosting (BPH), is more agile and diffuse. In this case, it's not a bunker but a pick-up-and-move environment. If they are disrupted or discovered, they simply leave that host, sign up with another, and move on. This agility makes it harder to track them as does the fact that their primary business is reselling hosting services, not holding actual content. The participating ISPs are typically "low, bargain-basement" ISPs that are not careful about what they host and are, in fact, complicit in the hosting of abusive services. Also, they are hosted in countries that do not invest much in anti-abuse issues. Because they often host a

combination of good and bad services, that is abusive and legitimate services, outright blocking produces a lot of collateral damage.

To date, our view of these BPH sites is from the outside. A number of companies, including Cisco's Open DNS Group and the speaker have done studies that have tracked BPH sites using external signals. What was needed was a look from the inside to understand the various parts of the operation to identify pressure points that could be exploited. That opportunity occurred in 2008 when the Dutch National High-Tech Crime Police Unit and the Thai police, in a joint operation, took down a major BPH operation, arrested two operators, seized back end servers, etc. which gave them a dataset. The BPH was Maxided with secondary business in child abuse material called DepFile.

This dataset, the dimensions of which were described in detail, provided an opportunity to examine the customers, merchants, upstream providers, demand, supply, and the finances of the BPH. In each area, the business model was determined to be robust and not likely subject to pressure. For example, because cryptocurrencies are the main method of payment, the finances were not vulnerable; the upstream supply chains were robust (particularly of child abuse pornography) and they outstripped demand so cutting off some merchants would have little effect – there is a concentration of suppliers but there is also, unfortunately, an oversupply.

The area that did evince some vulnerability was profits. The sale of hosting itself (Maxided) was only marginally profitable – seven years of operation of the BPH netted only $680K, at most, and that is without personnel costs which were not recorded in the files. The secondary business, DepFile, was highly profitable with profits of $4.3M over five years of operation. Maxided was more valuable to its owners as a cheap way to acquire BP servers for the side-business than its own business model. If it had just been a BPH business, it would probably have closed down as not profitable enough. But the side business was quite profitable and it relied on Maxided and made that business useful.

Conclusions:
The two most likely pressure points are ineffective. Disruption via payment channels is unlikely to stop BPH sites because they use crypto-currencies. There is no clear pressure point via upstream providers due to the high number of alternatives. It may be possible to identify and handcuff a few people but this is very hard and very costly.

Broader potential implications are that taking down the providers is costly. A possible alternate route may be to force an increase in operational costs of the BPH provider to render the business even more unprofitable, given the already thin margins of 10 – 20%. More research on mitigation and detection of agile abusive hosting is needed. As we increase our understanding of how these BPH sites operate, we will be able to better identify where research is needed.

Examining Login Challenges as a Defense Against Account Takeover

Speakers:
Periwinkle Doerflier, Maija Marincenko, Juri Ranieri, Yu Jiang, Angelika Moscicki, Damon McCoy, Kurt Thomas
NYU Tandon School of Engineering

Abstract:
Authentication is a difficult problem, as there is no analogy to real-world systems, making it a significant security challenge. To understand how to protect and safely authenticate users, a Google study Dorflier presented a study conducted with Google, which documented the levels and varieties of user attacks (bots, phishing, fancy bear) and examined attack success rates for various modes in four scenarios – log in, account recovery, organic (e.g., new equipment, new location), an experiment group. Prevention included device setting OTP, USB key, authenticator OTP, device prompt, and SMS OTP.

The study determined that a user's high level of urgency to access their account made them more successful in meeting challenges to logins, as did familiarity with challenges (having 2-factor authentication in place). Authentication is an arms race where improvements in authentication are met with parallel improvements by attackers. Attack prevention is also limited by the assumption that phones and users are 1:1 – but people share phones and that needs to be considered in authentication methods.

Synopsys:
Credential theft is one of the most serious cyber problems today. Over two billion credentials were leaked in one event in one week. Somewhere out there someone has your credentials. Based on work done at Google 18-24 months ago, the speaker described an effort to categorize attacks as coming from bots, phishing, or targeted (Fancy Bear or spear-phishing) attacks and examine their success rates using four user scenarios. The scenarios were normal logins with 2-factor authentication; users in the account recovery (forgot password); users shown challenges because they had a new device, were on a new network, etc.; and a random group of users who had the correct login but were challenged simply to see how they managed the challenge. In all, there were some 350,000 attacks examined, the majority were logins with 2-factor authentication, and fewer than 500 targeted attacks.

The scenarios were displayed in a graph with two axes – the urgency of accessing the account and the mental preparedness to be challenged. Results show that if the urgency is high, the success rate of dealing with the challenge is high; if there is little urgency, the rate of success in handling the challenge is low. Likewise, if one is prepared for a challenge (i.e. users in the 2-factor authentication group) then success in dealing with a challenge is higher than those who do not expect it.

The type of challenge was also examined. i) Prevention rates for device-based challenges from bots were 100% when challenged with things such as device setting OTPs (one-time-password), USB keys, authenticator OTPs, the device prompts, SMS OTPs. ii) The prevention of phishing attacks was good but not perfect with all types of device challenges noted above. iii) With targeted attacks, hardware tokens and USB keys are solid ways to prevent attack success. SMS OTP codes are getting out, somehow, and other methods of prevention are less effective.

In terms of scenarios: i) People using 2-factor authentication do well in meeting log in challenges – they are expecting a challenge. ii) Recovery flow users are about 50% successful with the various device-based challenges which may be due in part because they really want access to the account so their motivation is high. iii) Users seeing a challenge organically (they changed locations or equipment, etc.) did not do well. iv) Challenges to the experiment group were not met with great success – that is the users were not expecting a challenge and, in general, they were not successful in dealing with the challenge.

Conclusions:
The more users are comfortable with challenges, the better they do. Part of device-based challenges is that the person may not have the device handy – data reflect that not everyone carries their phone around at all times. However, lockout is unlikely – 97.9% of users shown device-based challenges were able to successfully access their account within seven days.

Users should register a device to their account so that when they are challenged, the identity provider can use stronger, more robust challenges. Security Keys are great for security but terrible for usability. Knowledge challenges are not secure, as phishers are good at getting the information. Security questions are 93% effective but mostly because you have forgotten the answers and so you can't accidentally provide them to a phisher.

Blockchain and Supply Chain Solutions

Speakers:
Ramesh Gopinath, Vice President, Blockchain Solutions, IBM
Laura Loughran, Former Senior Product Manager, IBM Blockchain

In conversation:

Ramesh:
Blockchain will fundamentally transform business processes in all industries. There are two key ideas – first that blockchain allows companies to share information through a shared, decentralized database that is not under the control of any member of the group. Second, the

information that is stored is immutable. Blockchain can be used to address traditional business methods that are inefficient, expensive, and vulnerable. With blockchain, there is consensus, provenance, immutability and with encryption, a secure method to sign documents and share them with appropriate others.

These attributes are important because simple business transactions can get very complicated when there is no end-to-end picture available. Traditionally, no party can see the whole picture of a transaction – supplier, processor, shipper, retailer, etc. Dispute resolution can take weeks. With blockchain, all parties are on a shared database where information can be shared selectively.

While the technology has some challenges, the most difficult part is assembling the participants along the supply chain, some of whom are competitors, and getting them to agree with what can be selectively shared.

All of the participants in a particular blockchain are known so it is a permissioned blockchain. While some may use cryptocurrencies, that is not a requirement for permissioned blockchain use cases since all parties are known. The vast majority of currently operating blockchains are permissioned blockchains; there are very few permisionless blockchains, BitCoin being the most notable where its management and the transactions are anonymous.

IBM's blockchain work has focused on building blockchains for real groups of businesses to share information to solve problems. There is cross-industry participation in over 100 active networks. Among those with high value are those that rely on supply chains, such as food suppliers. Blockchains have three elements – the flow of goods out; the flow of value (money, brand) back; with data in the middle flowing both directions.

IBM has three levels of investment or product lines:

1. The open-source blockchain platform that can be used to build solutions. Originally it was run by IBM in their cloud but now it can be run anywhere.
2. Solutions built for specific communities where IBM assembles the ecosystem of the industry and builds and manages the blockchain system.
3. Services that entail building solutions but not running them.

Here are three applications: **Trust your supplier** which is a blockchain (built in partnership with Chainyard) for building identities for companies so they can easily be onboarded as new suppliers and their data is kept current; **Tradelens** (built in partnershiop with Maersk) which tracks containers on ships facilitating information on container location and route information as well as expediting the massive paperwork that is needed for shipments; **FoodTrust** solves inefficiencies in the food chain and manages quality risk by maintaining end-to-end transparency through data collection at each handoff point. This system was discussed in detail by Laura Loughran in the next talk.

In all of these systems, it is preferred that data come directly from IoT devices, bypassing human error. In some cases, particularly with FoodTrust, the quality of data improved because individuals knew the data was immutable and were more careful that it be correct.

Laura:
Food safety was the primary value driver for the FoodTrust product development, along with supply chain efficiency that would reduce waste and increase the freshness of food. A newer application was its use in building a brand, that is, building a story that proves the quality of the product or its authenticity.

There are three layers to this application:

1. Architecture – the IBM blockchain platform is the base layer where all transactions are recorded and where the ledger is kept
2. FoodTrust platform – the API which strings together the end-to-end view of shipments, document sharing, entitlements, member management, etc.
3. Applications are the top layer – this includes onboarding of companies, traceability, certifications, sharing documents needed to do business. In this layer business partners leverage the system's information to get added value.

In addition,the traceability module can also trace multi-ingredient products. Consider two use cases – the 2006 E coli outbreak where, due to lack of timely information on the source of the contamination, all spinach was removed from shelves nationwide. The industry took six to seven years to recover from this. In another mock trial of FoodTrust, it took store personnel over six days to track the supply chain of a bag of mangos whereas it took FoodTrust, in a similar situation, 2.2 seconds. FoodTrust cannot say where contamination occurred but it allows investigators to home in on the details of the contaminated product rather than search the entire field of options (e.g., all growers, all shippers, all processors, etc.)

Already industries, such as Tunisian olive oil and shrimp export in Ecuador, are using blockchain to build their brands by being able to show customers that their products were "as advertised" – i.e. real Tunisian olive oil and sustainably grown shrimp. The users' API was the basis for entering and retrieving this information which was shared with customers through QR codes, for example. In another example, a coffee company created a private network for end-to-end traceability, including enabling the customer to send a monetary tip to the grower of the beans used in a particular cup of coffee.

Ramesh:
Regarding the governance of FoodTrust – the data is owned by whoever uploaded it, not by IBM. Participants can see their own data but cannot see nor share other data without the permission of the data owner. Private networks, such as the coffee example above, may have full control of the supply chain and so may have a different form of governance.

HyperTrust deals with tracking goods that are, in particular, temperature-sensitive and it grew out of the pilot done with Golden State Foods for moving fresh beef. At its core, it tracts what was received, what was done with it, and where it was sent. The intent is to extend product freshness and shelf life.

There are technical challenges:

1. FoodTrust data comes as encrypted or partitioned which can cause a problem. In the case where the farmer is sending, for example, strawberries, the farmer doesn't know where they will end up so can't set up the supply chain at that point. When the strawberries are finally shipped to a particular store, then the information is linked to the grower to make the supply chain. It takes a sophisticated data entitlement algorithm to make this work.
2. Small farmers are typically not able to or interested in running blockchain nodes. They use trust anchors – companies that run nodes and vouch for the integrity of the data to the Food Trust community - to do this for them. Running a node opens the question of how to share the appropriate amount of data.

Graphene: Efficient Set Recognition for Blockchain Propagation

Speakers:
Brian Levine , Pinar Ozisik, George Bissias, Gavin Andresen, Darren Tapp, Sunny Katkuri
Cryptoeconomics Lab, UMass Amherst

Abstract:
The work done by this group solved an old problem - synchronization between two systems or set reconciliation. Graphene focuses on drastically cutting down on data transmission through the use of Bloom filters and invertible Bloom lookup tables (IBLT).

Synopsys:
Transactions occur and blocks are propagated through the network. Reconciliation between data sets in a blockchain needs to occur as fast as possible to prevent new blocks from being created. Each peer in the network has a mempool of unvalidated transactions. Some blocks have more current information about validated transactions that can be used to clear out transactions from another mempool. It helps to work with a high degree typology database. To clear out the mempool, peers query each other "do you have" and can get data sets in return. Sending whole blocks can be network intensive. Graphene attempts to reduce the network traffic between peers and provides a method whereby there can be intermittent checking with select blocks (nodes???) rather than with each block (node??) each time. To do this it uses a combination of two probabilistic data structures: Bloom filters and invertible Bloom lookup

tables (IBLT) to reduce the traffic between neighboring peers. Using these filters, the overhead shrinks to almost nothing. Bloom filters help represent a set of items and is useful for determining set membership. Calculations are based on the mempool not the block. False positive rate is 1/144M. IBLT is a generalized bloom filter, not an array of bits but of counters.

It is costly to send Bloom Filters or IBLTs when the mempool is large. Graphene combines these two techniques and uses the Bloom filter to reduce the symmetric difference between block and mempool and it uses an IBLT to recover from small errors in the Bloom Filter. One challenge is that it is hard to estimate size of symmetric difference between mempool groups but this is a key element in successfully deploying this technique.

Conclusions:
The Graphene minimizes network costs for set reconciliation. It is applicable for block propagation and mempool synchronization and in general for a variety of systems. Deployment and simulation results show significant improvement. It is deployed on the Bitcoin Cash network via the Bitcoin unlimited client with 700 nodes.

More information can be found at: Cryptoeconomics.cs.umass.edu/graphene.pdf

Watching IoTs That Watch Us: Empirically Studying IoT Security & Privacy at Scale

Speakers:
Danny Y. Huang, Postdoc at Princeton University

Abstract:
How to measure IoT security risk problems at scale, in the wild.

Synopsys:
There is no easy way to get empirical measurements of security risks to home IoT devices - we have no idea what data is being sent, to whom, and from whom. There is a large variety of complex devices and complex user behaviors; for example, some users are confused about how to even interact with their devices. In addition, there are a large number of vendors supplying IoT devices. Lab studies and crowdsourcing have been tried but they are hard to scale to more than 100 devices or 100 users. Internet scanning techniques omit local networks – they cannot tell what's happening on a private home network and it's hard to replicate any results.

This project recognizes the need for user-friendly tools, specifically a one-click tool to provide users with a transparent way for them to see the security and privacy risks of their devices.

Solution: IoT Inspector is a software tool to visualize the activity of IoT devices and identify suspicious behavior. The software can be run by anyone to visualize the activities of their devices and identify potential security problems. IoT Inspector shows who IoTs are talking to by capturing the traffic between the IoT device and the router by spoofing the certificate. Data is uploaded to Princeton where the two visualizations are created for the users. One visualization shows the amount of traffic, even if the device is not being actively used, the other shows who the device is talking to. The tool was released in May there 5,200 users anonymously contributing data from 52,000 devices (TV's, cameras, appliances, cars.)

There are usability and system challenges. Among the usability challenges are recruiting users – currently, that is done by tweeting and talking to the press. Convincing users the project is legitimate and good is a challenge because the system has access to sensitive data and devices. In addition, figuring out how to do data visualization for non-technical users is a challenge. System challenges include having reliable data collection, preventing domestic spying – the software can be used to compromise a phone or a laptop so there are safeguards that the software only works if the device appears to be an IoT device. Performance data on specific IoT devices will be used to inform users and potential buyers about the features and risks of particular equipment.

Conclusions:
This work has opened new research opportunities in security and privacy, device identification, anomaly detection, and opportunities for its use in health monitoring.

The next challenge for the project is in the area of smart TVs that are being tracked by third-party trackers and advertisers. "You watch TV; your TV watches back." Smart TVs are being tracked by 3rd party trackers and advertisers. It is difficult to automate the input and difficult to decrypt the traffic. It is also hard to change/spoof a certificate which is the basis of the current IoT Inspector. Thousands of smart TV channels (i.e. apps) are able to share user data. A paper at CCS '19, "Tracking on smart TVs," looked at who is doing the tracking – the TV or the vendors (third-parties)? What data is being sent? Can users top the flow of data? How is it different from the mobile environment?

There are no known tools to address this area; it's different from the web and mobile environment and difficult to decrypt traffic –we just don't know what's going on since most channels are proprietary. However, we built an open-source automatic TV crawler that we used to study thousands of channels on Roku and Amazon. We saw sensitive information being shared with 3rd parties - some new, some known, such as Google. We found that the "do not track features" do not work and ids are still being sent to 3rd parties. This information has been reported in the Washington Post, the New York Times, and on NPR. The NYC Cyber Command, the New York Attorney General's Office, and the FCC are using the system to investigate child privacy law violations on one particular smart TV platform.

Further information is available at:
https://iot-inspector.princeton.edu/ can down load software for MAC

Trinity: a blockchain based IoT data collection system

Speakers:
Gowri Ramachandran, Senior Research Associate, USC Center for Cyber-Physical Systems and Internet of Things, USC Viterbi School of Engineering

Abstract:
Data-driven IoT apps are on the rise causing an increasing need for real-time data. This work focuses on how to seamlessly collect, share, and manage data through a trusted infrastructure using blockchain.

Synopsys:
This work involves processors and stakeholders. Processors include sensing and actuation, computation such as low-power embedded platforms to edge and cloud computing, and communication using low-power, short-range and LPWAN to 4G networks. Stakeholders include hardware manufacturers, application developers, communication service providers, and system administrators. IoT infrastructure relies on several processes and may involve several stakeholders.

Trinity is a Byzantine Fault-Tolerant distributed publish-subscribe broker-based system with immutable, blockchain-based persistence. It uses a messaging framework used in industry and can be implemented as a centralized or a decentralized marketplace for IoT data. It is a way to share data with multiple clients such as app builders or with multiple organizations. One benefit is that data producers and consumers are isolated from each other. In the Trinity architecture, for each domain, there is a set of publishers and a set of subscribers that are connected through a trusted infrastructure. All data is published to the Trinity infrastructure and undergoes verification on the blockchain before it is sent to subscribers. Brokers are connected to a consensus node and whenever data hits a broker, it will get verified on the blockchain before it gets published to subscribers. All subscribers get the same data and they can go to the ledger to look at the data in the blockchain if there are questions. With this structure, Trinity allows multiple organizations to share information through a trusted infrastructure and have the same data.

In an example of work under development with the Department of Justice in California, data on profiling done by police during stops is being collected in a tamper-proof system using Trinity.

One of the concerns about decentralized marketplaces or IoT 1.0 is that each app is developed independently as a silo and different agencies are using different apps collecting the same data but not sharing it, e.g., data from a camera focusing on a street can be used to monitor traffic, look at parking or trash accumulation, etc. The problem is that the data is owned by a single agency and used for only one application so you end up with three cameras, etc. There is concern about vendor lock where a city opts to go with a proprietary solution from one vendor with no interoperability.

Work is now moving toward the development of I3 – an intelligent IoT integrator as a community marketplace for smart cities data. Other projects include a Streaming Data Payment Protocol (PDPP) and a micropayment platform for use by trusted vehicular services to exchange data. Blockchain would be used for payment and for identity management. Field testing is being done. http://i3.usc.edu/

Conclusions:
Blockchain and the distributed ledger provide tools for applications that involve multiple stakeholders and micropayments. Is the technology ready for adoption? There are issues of performance, public versus private systems (permissioned), interoperability challenges, and finally, the fact that garbage in, garbage out is always a challenge and a physical interface weakness.

Towards Building Trustworthy Blockchain Smart Contracts

Speakers:
Ronghui Gu, Assistant Professor, Computer Science, Columbia University
Co-founder of CertiK, described in this talk
Vishal Misra, Professor of Computer Science and Electrical Engineering, Columbia University

Synopsys:
In traditional systems, exchanging assets requires a trusted 3rd party service provider. With blockchain, this is not required. Instead, users trust a smart contract to correctly encode a transaction's logic. Trusting the contract means trusting the code which may or may not accurately reflect the intention of the transaction. What if that smart contract is wrong?

There are many examples of smart contracts having bugs. The most famous is the DAO (organization) attack. This was a double-spend bug that cost the DAO $50M in 2016. A second type, an integer overflow bug, was launched as a Business Email Compromise (BEC) attack

netting hackers $5B. An integer overflow bug cost the cryptocurrency EduCoin to lose 2B tokens. During the single month of December in 2017, $630M was lost to hackers.

Blockchains and Smart contracts are vulnerable because they are open source and once they are uploaded and there is consensus they are very hard to fix. Hacking smart contracts is not like hacking in the usual sense. Bugs allow hackers to follow the rules of the contract but because of a loophole, they are able to divert assets to their account. Or, the rules have unexpected behavior (e.g., integer overflow) that can lead to an attack.

There are tests available to try and avoid bugs; some are easily deployed but not really effective. "Program testing can be used to show the presence of bugs but never to show their absence" Edsger Dijkstra.

How to improve Reliability?

- Human Review/ Audit (not reliable, not feasible). – most widely used technique in the blockchain world but with 2M active contracts, it is not practical.
- Running tests (not reliable, although highly feasible)
- Runtime monitor (medium reliable, medium feasible)- used in industry where you can shut down a computer or a network and fix the bug, Can't do that with blockchains out in the world.
- Formal Verification (High reliability and high feasibility). This technique uses mathematical methods to verify that the code matches the designer's intention.

Conclusions:

The CirtiK group developed a mathematical technique to prove that the code in a smart contract satisfies the developer's intention (i.e. specification). Of course, you have to trust that the specification is an accurate representation of the intent. The formal validation allows bugs to be discovered and fixed before the contract is uploaded. The process results in a smart contract that includes the specifications, the code, and the proof and provides an end-to-end guarantee of that smart contract.

It is recommended that blockchains may want to add a feature that requires that smart contracts are validated before allowing them to be uploaded, at least for contracts managing large amounts of money or critical processes. If a user cannot provide the proof, the CertiK company can create the proof. Currently, there is no enforced validation of smart contracts by blockchains.

The company is helping most large blockchain players to secure their smart contracts. CirtiK was released in June ’18 and since then has verified 160,000 contracts, auditing more than 88K lines of code.

The company developed a programming language called DeepSEA to help users write code and their specifications or intentions at the same time. Columbia Yale, IBM Hyperledger, ethereum, and the QTUM Foundation are supporting this work.

Scaling Blockchains at the Network Layer

Speakers:

Aleksandar Kuzmanovic, Professor, Northwestern University
Founder & Chief Architect, Bloxroute Labs

Abstract:

The talk describes the traffic loads caused by blockchains and provides a solution to help all blockchains scale. No blockchain should fail due to scalability.

Synopsys:

This talk discusses the problem of scalability in blockchains. Some examples of transaction size include Bitcoin at 3 tps (transactions per second); credit card transactions occur at 5k tps. If cars in the US were to get gas once a week, it would require 450 tps; vending machines used 4x day would result in 1k tps; machine to machine (IoTs) result in 50k tps, and global micropayments yield 70k tps.

Helping all blockchains scale:
As a matter of curiosity, the speaker explained where the standard 3 tps in Bitcoin comes from: Bitcoin transactions 540 bytes per transaction; the system has to send 1MB block every 10 minutes which amounts to 1,900 transactions/ 10 minute which after division comes out to 3tps. This is not scientifically based, it was developed by a someone and has become the standard for Bitcoin

One could solve this by increasing the size of blocks and decreasing the interblock time. However, some people don’t want to touch the blockchain for security reasons others say no

you can increase block size as much as you want. The speaker contends that you cannot arbitrarily increase the block size.

Why can't we make these changes? The pitfall is P2P data distribution. By increasing the block size you increase the amount of data being sent and with randomly located nodes you have likely increased the number of networks over which the data travels. In networking, you only need one place for things to go wrong to disturb the whole flow – the weakest point determines the performance of the network as a whole. When there are middle points, they can be the bottleneck of how fast a block is moving. When propagating to more machines, there will be more places for things to go wrong, thus, the network will be slow.

How does Bloxroute work? Miners build and send blocks to peer nodes, including the open-source gateway. The gateway sends the block to the Bloxroute BDN which propagates the blocks 10-100x faster. With this BDN, transaction indexing goes from 540 Bytes to 4 Bytes, block compression from 30 Mbytes/block to 222 Kbytes – only small identifiers are sent, not the whole transaction allowing for faster distribution. Users can create larger blocks and communication is still increased. Bloxroute has servers around the world using open-source software. The system is neutral and independent of the consensus layer – nothing in anyone's system is changed. One technique used to increase speed is cut through block routing which is an old technique that is useful here. With this technique, once a block is put on the wire it takes time to propagate – in this case as soon as the first bits are received by the server they are sent along without verification. Only when the transmission is complete are the keys sent to unencrypt. Blocks are very small and can move 100x faster; the distance between nodes and servers is small in this network; traffic does not depend on what others are doing – everyone is directly connected to the others so data can flow independently thus limiting performance problems.

The work claims three accomplishments:

1. Scaling bitcoin – simplest protocol – 1k tps on the test network
2. Joined Bitcoin cash network mining test network and were able to move transactions 20 times faster, could sync mempools faster, and fork recover faster proving that large blocks are feasible
3. Joined the Ethereum test network with its shorter shorter blocks. The could push blocks 50% faster; propagation from and to China was faster; and the block size limit was increased by 25% proving that Ethereum can scale

Conclusions:

Bloxroute is the first and only trustless blockchain network. Blockchains are decentralized, no single entity controls them unlike a distribution network where one entity can see all the blocks that go through and that entity has a lot of power about what's going on in the system. This system was designed so if there is a bad actor everyone would know it. It is a centralized network but peers in blockchain are also used to audit so there can be no cheating. This is possible because blocks are encrypted prior to being sent into the Bloxroute network. Only when everything has been propagated is the key sent to unencrypt at the end point and is the content known. Second, if a block does not go through, it will resend it only to a peer and that node will send it on. If you cheat, everybody knows.

Indirect relay ensures the BDN cannot discriminate against a node. Auditing using test blocks help nodes realize if the network is misbehaving

The business model is based on fees. With this system, for users, the transactions go up and the fees go down as much as 100x. The product is currently supporting several blockchains: ONToltooy, JCMorgan Chase QTUM, Conflux, Ethereum.

New challenges remain. Ripple occurs when there are few consensus nodes and supernodes receive a lot of redundant data. There is also a transaction incast problem involving redundant data with a large number of connections.