Using blockchain principles for improving AI research and security

Kresimir Kalafatic,
E-mail:kresimir.kalafatic@gmail.com

Abstract—Distributed ledger (blockchain) is a distributed database whose every written record is signed by private key, every record insertion and change can be traced back to specific public/private key pair and whose atomic operation is irrecoverably committed after the databases mathematically prove their validity and distributed consistency. The mathematical principle of blockchains can be used in improving AI research and security.

Index Terms—blockchain, AI, security, linking research data and code

1 Introduction

Distributed ledger (blockchain) is a distributed database whose every written record is signed by private key, every record insertion and change can be traced back to specific public/private key pair and whose atomic operation is irrecoverably committed after the databases mathematically prove their validity and distributed consistency. The blockchain is based on statistical and cryptographic principles for improving security of data management enabling authorized and audited change and processing of data from its origin to the end of data usage.

Cybersecurity is an important element in every sector of human interaction and lately regulators are introducing new frameworks for increasing baseline security of all institutions under their supervision. Even though formal procedures are used in software and hardware component design, most of the designs have security flaws which have their origin in errors of human personnel or computer models. To reduce the security flaws the design origin has to be proven by mathematics and physics.

Knowing the origin of data enables ranking the data quality of different data sources in machine learning. Sensors can experience degradation during their lifetime and blockchains enable knowing which contaminated training data has to be discarded or repaired. Using multiple parallel sensors performing the same task and using appropriate consensus protocols can reduce the influence of contaminated data sources on the final AI decision.

Having a decentralized database improves data availability and infrastructure resiliency while enabling creation of different multiple AI designs using the same replicated datasets with different splits for training, validation, testing datasets. Decentralized computer models enable using distributed computer resources and splitting the work into smaller parts enabling resource polling architecture.

2 Short overview of some blockchain principles

Distributed ledger (blockchain) implementations are in different stages of development and some of the most popular implementations are bitcoin and ethereum. There are many papers describing the principles of bitcoin and ethereum, so this paper will concentrate on just few important principles.

2.1 Public/Private keys and wallet address

Public and private keys are based on cryptographic principles. Bitcoin and ethereum use the Elliptic Curve Digital Signature Algorithm. Public/Private-key cryptography enables digitally signing data with a private key and anyone who knows the public key of an entity can verify that the signature is valid.

The asset (data or some other) is assigned to the owner based on a wallet address. The wallet address is derived using hash algorithm on public key. The wallet address adds additional layer of security because the public key is hidden until the transaction is initiated from the wallet. Using wallet addresses enables data anonymization which is required by the EU GDPR (General Data Protection Regulation).

2.2 Decentralized database consistency algorithms

Atomic operation in blockchain is irrecoverably committed after the databases mathematically prove their validity and distributed consistency. There exist different types of consensus algorithms which form several groups. ”Proof of work” and ”Proof of stake” consistency algorithms are based on mathematics and operate on economic principles and monetary metrics. Generation of new blocks (tokens) is used for securing database consistency and history while enabling covering operational cost and investment in the infrastructure by selling mined blocks to the network members.

2.2.1 Proof of work

Bitcoin and ethereum are currently operating using ”Proof of work (PoW)” consensus. Bitcoin ”Proof of work” is based on autocorrelation function which uses hash of previously generated block, blocks from latest transactions, address of the miner and other data. The autocorrelation function

uses mentioned data to find nonce which start with defined number of zeros. The first miner who finds the number with required property announces its solution of the autocorrelation function to the other miners. When the new block is accepted by the majority of the miners, the distributed consistency for previous set of transactions is reached. Using autocorrelation functions helps linking the new block with previously generated blocks and transactions. As time passes the solution of the function progresses from the current solution and moves to new solution which depends on newly initiated transactions and history of previous solutions. Current bitcoin mining algorithm uses large amount of electrical power to find a random number with specific property.

2.2.2 Proof of stake

"Proof of stake (PoS)" are based on autocorrelation functions which uses hash of some previously generated block, blocks from latest transactions, address of the validators, the amount of a validators deposit and other data. The basic principle of some PoS consensus algorithms is that the value of the time-locked deposits is much higher than the economic value of possible inconsistent transactions. The participants risk part of their asset in the case of the inconsistency for the financial gain for performing validation service.

Some Proof of Stake algorithms are in one way similar to bank operations. Every bank has capital which is a small portion of the bank asset balance (deposits, credits and other financial instruments and assets received) under banks control. The interest of the banks as legal institutions in preserving financial network viable and secure is legally based on the principle that the penalty of wrong decision will result in the reduction of banks capital, market share and future profitability.

2.2.3 Hashgraph

Hashgraph is a patented algorithm for distributed consistency and is based on graph theory. The algorithm is a simple solution for distributed consistency of atomic operations which doesn't depend on the economic principle or financial monetary value.

2.3 Virtual machines

Beside using blockchain technology for data management, blockchain is also used for code management. The code is also signed in blockchain and the signed code is executed in virtual machines (VMs). For security reasons bitcoin uses VMs which are not Turing complete. The ethereum uses Turing complete VMs. Running the code in VMs requires adequate funds on the account and is a security feature for preventing DOS attack by malicious code and infinite loops on the blockchain network.

2.4 Homomorphic encryption

Homomorphic encryption is an encryption which allows computation on encrypted data, generating an encrypted result, which when decrypted, matches the result of the operations as if they had been performed on unencrypted data. The purpose of homomorphic encryption is to allow computation on encrypted data.

2.5 Assembling the parts

A hash function is function that has a property of mapping data of arbitrary size to data of fixed size. Hash can be used for detecting duplicate data (code, byte array), creating a unique fixed size index of data for quick data lookup or data verification (digitally signed hash). Digitally signing the data and digitally signing the code improves security and enables detection and prevention of unauthorized changes.

Consensus algorithms enable distributed database consistency and secure multi-master resilient infrastructure whose implementation can be vendor independent (no hardware and software vendor lock-in). The validation of initiated transactions is based on mathematics and performed on multiple nodes which can have different HW (CPU,motherboard,..), SW (OS,...) configurations and can have multiple different vendors. Multi vendor configuration reduces the attack surface of the network and the bugs in one vendor HW or SW can be detected during operations. Upgrade and maintenance of the infrastructure can be done gradually using rolling upgrade option. Technical resilience is also extended to business resilience from bad decisions and problems of one vendor company.

If consensus algorithm is based on autocorrelation function the correct current solution of the function depends on some or all of the previous solutions. Origin of data and code can be audited and operation on the data can be reproduced. This is important benefit of blockchain technology because EU GDPR mandates a right to explanation.

3 Using blockchain principles in scientific research

Scientific research involves a systematic objective process of gathering a multitude of data for analytical purpose which will be used to derive a conclusion. The process focuses on testing different ideas through a systematic process which is documented in such a way that other individuals can conduct the same analysis and derive the same conclusion. The scientific research process is a multiple-step process with interlinked steps similar to blockchain. If change is made in one step of the process, the researcher must review all the other steps to ensure that the changes are propagated throughout the process.

The first step would be to establish the separate Certificate Authority for every field of research like mathematics, engineering, physics, chemistry, economics and other which will issue and manage certificates for every researcher or person who requests it.

Blockchain is an important platform and tool in work automatization, process standardization and some blockchain principles can be used in research which has mathematical roots. The format used for example will be Tex which is used in writing most of the research papers, but new format should be designed to simplify automation of the process.

One simple example of using blockchain for documenting research will be described on an example of modeling electrical power of a computer. The description of the example is given to describe the basic idea of using mathematical blockchain or "mathchain".

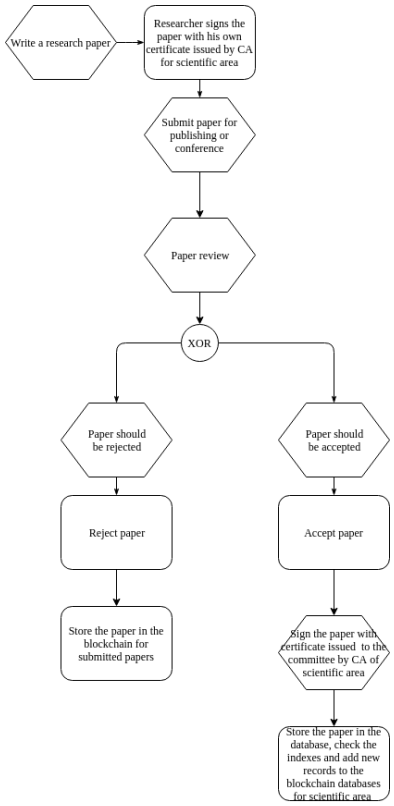

Fig. 1. Example of paper submission and research paper signing by the committee which accepted the paper

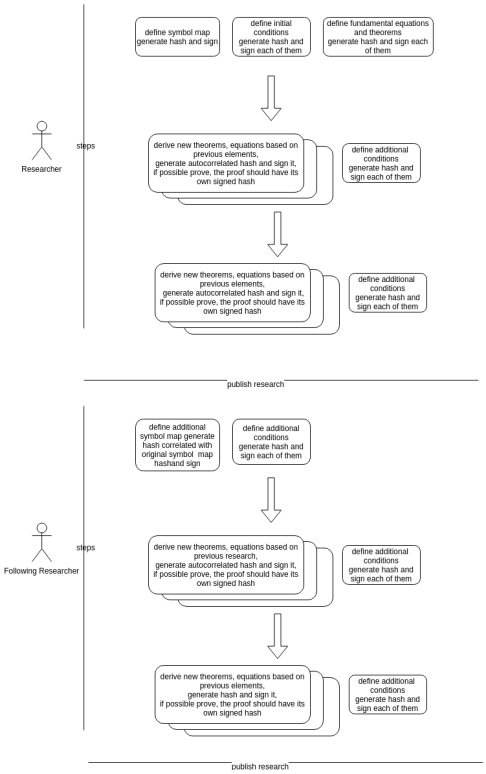

Fig. 2. Example of blockchaining research

3.1 Defining hashes for used mathematical symbols and conditions

The first step in mathematics is defining symbols which will be used to describe multitude of data groups for which researcher believes have some correlation. The symbol definition table should contain symbols, description of the symbols, and a hash of the symbols used. The TABLE 1 is one example of such table. After the table is defined SHA2 sum of the symbol table is calculated. The hash of a symbol table is used as an index for conditions and mathematical equations.

The SHA2 of Table 1 is:

f1195066a0177e3ab54eaf438f81b17fbc51640f37aba065b7f1a5010a35c4ee

After defining symbol definition table, generating index for every symbol and an index for symbol definition table, the conditions for the equation can be defined. The conditions describes the properties of the environment for which the equations are valid. For the given example on electrical power of a computer, the environment would include operational range of computer (in Europe 220V, 50 Hz, temperature from 1.7 C° to 32.2 C°, room ambient). For simplicity in this example condition will be defined as a string ”operational_environment”. The SHA2 is calculated as SHA2 of a string ”operational_environment”:

487703d7b867d3b53e91ec1f11f3b7e588552b0b6c011671b4dc82d669fc39d9

3.2 Defining hashes for mathematical equations

After the conditions hashes are calculated and a hash of defined symbols tables is known, the hash of equation can be calculated. The proposal for generating hash (an index of an equation) is SHA2 sum on following concatenated string:

```
hash{sym};hash{cond1}:hash{cond2}:..;Tex equation
```

Every mathematical blockchain contains the index of symbol definition table ”hash{sym}”, the index of conditions ”hash{condX}” in which the equations are valid and the Tex representation of equation. This blockchain is basically a record linking indexes of symbols, conditions and equations. The hash on this structure would define index for the record which would uniquely and immutably represent the information about the mathematical principle. The equations that are mathematically identical (because of mathematical property) would be represented by one equation - a representative. In the case a researcher wants to check if the discovered equation exists, the researcher would first find the symbol definition in the area of research. For all the symbols in the equation the researcher would generate hash/indexes of all symbols in the equation, and using this indexes find all the records with them. If there isn’t such an equation or conditions don’t match the records found, researcher could generate a new index/record.

The principle is shown on example for electrical power of a computer.

Every change requires energy. Power is the amount of energy transferred per unit of time. Electric power of electrical system is defined:

TABLE 1
Example of symbol definition table

symbol	description	symbol SHA2
I	current	a83dd0ccbffe39d071cc317ddf6e97f5c6b1c87af91919271f9fa140b0508c6c
U	voltage	a25513c7e0f6eaa80a3337ee18081b9e2ed09e00af8531c8f7bb2542764027e7
C	capacity	6b23c0d5f35d1b11f9b683f0b0a617355deb11277d91ae091d399c655b87940d
V	static voltage	de5a6f78116eca62d7fc5ce159d23ae6b889b365a1739ad2cf36f925a140d0cc
α	constant parameter	8ed3f6ad685b959ead7022518e1af76cd816f8e8ec7ccdda1ed4018e8f2223f8
m	constant parameter	62c66a7a5dd70c3146618063c344e531e6d4b59e379808443ce962b3abd63c5a
$E_{short\_circuit}$	short circuit voltage	7b4d928f61ceb002fb3615303093453bd7cf9491d84a2d8a9f81a074c0405d42
f	frequency	252f10c83610ebca1a059c0bae8255eba2f95be4d1d7bcfa89d7248a82d9f111
S	surface	8de0b3c47f112c59745f717a626932264c422a7563954872e237b223af4ad643
L_w	frontal luminace	d8ade83d2dcc64e38b42d877f101e746dc8a07d98c4b1f68c4bbbbe3e3f7499d
I_lum	luminous intensity	148534998ccb5cb8fc1fdb182999aa8ec67f3552d6a7fd181c6debdb578aa7b6
v_w	velocity of pressure wave	4b15a5a6867c654691000e823ae4a17ee31f636de5caf3796b6531dfbe854160
A	area of pressure wave	559aead08264d5795d3909718cdd05abd49572e84fe55590eef31a88a08fdffd
D_p	pressure difference across sound wave	ae94081e451f9069058b99b5031dbafc353fc4796332d2623e304342b6c068e6

$$P_{UI} = UI \tag{1}$$

8d463aa60853f719b7ececcdec1b16fa2b950c5db03ef40c35e269ce7343d4ab

Electric power used by desktop computer is defined:

$$\begin{aligned} P_{total} &= P_{CPU} + P_{mem} + P_{motherboard} + P_{GPU} + P_{speaker} \\ &\quad + P_{mic} + P_{display} + P_{I/O\_devices} + P_{fan} + P_{loss} \end{aligned}$$

cfd631d6d73bdc23de589ff89c42765747cbe0c21c9b0eb3709482bb9507696a

Every component has an individual power consumption equation.

For CPU power consumption is:

$$P_{CPU} = (mV) + (\alpha E_{short\_circuit} f) + (0.5\alpha C V^2 f) + P_{cpu\_loss} \tag{2}$$

1d7c41df24dc5bad67d28167a2e97f6a58d0bfbc65de3d5acdbd1ac9c1b409e0

For GPU power consumption of its components is:

$$P_{GPU} = P_{CPU} + P_{mem} + P_{fan} + P_{gpu\_loss} \tag{3}$$

a0944f4dd78ee880ba109747205b1b8b525ab7afc177f982aff8f79afc76b920

For speaker power consumption is:

$$P_{speaker} = UI \tag{4}$$

50ffad5e7f9fe1945f9b39893ceee96301c543b282a9abcd55d9c97f15c4d73f

, the power produced by sound is:

$$P_{sound} = v_w A D_p \tag{5}$$

a63387181d7e2dbe9e5392e8615341d412de4a468e4039af47bc244aeec466fd

For display power consumption is:

$$P_{display} = P_{panel} + P_{backlight} + P_{display\_loss} \tag{6}$$

ca4ff9f69ebf6094e301b59ab7a7aa8df1f1a074e252468f3b17875ecca605b6

and luminous intensity is a frontal luminance L_w of screen and the active area S of the display:

$$I_{lum} = L_w S \tag{7}$$

80847fa36472ee08b65abe415b078d1877a15dd4d5fe16935feb21a62c0cd7ca

The SHA2 hash of derived equation is an autocorrelation function of the initial symbol table, conditions and Tex representation of equations.

3.3 Explanation

The principle of generating blocks is the same to blockchaining transactions, but in this case we mathematically blockchain together symbol definition tables, conditions, initial and derived equations. The blocks define mathematical equations, conditions and their genesis process using correlation function. Every equation can be traced to the parent equations, so in the case of an error in the procedure, all the equations in which the error propagated can be detected and repaired.

"Proof of Work" principle used in blockchain has similarities with citation of research papers. The more the paper is cited by different researchers, the more the paper is reviewed the more it is trustworthy. Because following researcher references the hash of research paper and hash of the equation in its research for defining new theorems, the linkage from previous knowledge is preserved and secured. If the flaw in one paper or equation is detected, then all theorems and equations which derived from the erroneous element can be easily found and discarded. Every submitted paper would be signed by the researchers, reviewers and committees of conferences or institutions which accepted the paper time stamping it and signing it as validated by the signers. The accepted paper should be stored in the blockchain database established for scientific research. Not accepted papers would be stored in blockchain database for time stamping them and possible future reviews.

Using Tex notation for equations enable generating hash on text string which could be used to provide information for possible merging of theorems or reducing number of identical definitions by human researchers. Using blockchain principle enables the reader of the research paper to easily find the initial equation documentation if he wants to understand the whole process.

Because nowdays most mathematical calculations are done on the computer using procedures and functions in

TABLE 2
Example of some hash entries for using electrical power of a computer

Symbol	Type	Hash
P_{UI}	origin	f1195066a0177e3ab54eaf438f81b17fbc51640f37aba065b7f1a5010a35c4ee ;487703d7b867d3b53e91ec1f11f3b7e588552b0b6c011671b4dc82d669fc39d9 ;8d463aa60853f719b7ececcdec1b16fa2b950c5db03ef40c35e269ce7343d4ab :a25513c7e0f6eaa80a3337ee18081b9e2ed09e00af8531c8f7bb2542764027e7 :a83dd0ccbffe39d071cc317ddf6e97f5c6b1c87af91919271f9fa140b0508c6c
P_{CPU}	origin	f1195066a0177e3ab54eaf438f81b17fbc51640f37aba065b7f1a5010a35c4ee ;487703d7b867d3b53e91ec1f11f3b7e588552b0b6c011671b4dc82d669fc39d9 ;1d7c41df24dc5bad67d28167a2e97f6a58d0bfbc65de3d5acdbd1ac9c1b409e0 :62c66a7a5dd70c3146618063c344e531e6d4b59e379808443ce962b3abd63c5a :de5a6f78116eca62d7fc5ce159d23ae6b889b365a1739ad2cf36f925a140d0cc :8ed3f6ad685b959ead7022518e1af76cd816f8e8ec7ccdda1ed4018e8f2223f8 :7b4d928f61ceb002fb3615303093453bd7cf9491d84a2d8a9f81a074c0405d42 :252f10c83610ebca1a059c0bae8255eba2f95be4d1d7bcfa89d7248a82d9f111 :6b23c0d5f35d1b11f9b683f0b0a617355deb11277d91ae091d399c655b87940d :X
P_{sound}	origin	f1195066a0177e3ab54eaf438f81b17fbc51640f37aba065b7f1a5010a35c4ee ;487703d7b867d3b53e91ec1f11f3b7e588552b0b6c011671b4dc82d669fc39d9 ;a63387181d7e2dbe9e5392e8615341d412de4a468e4039af47bc244aeec466fd :4b15a5a6867c654691000e823ae4a17ee31f636de5caf3796b6531dfbe854160 :559aead08264d5795d3909718cdd05abd49572e84fe55590eef31a88a08fdffd :ae94081e451f9069058b99b5031dbafc353fc4796332d2623e304342b6c068e6
I_{lum}	origin	f1195066a0177e3ab54eaf438f81b17fbc51640f37aba065b7f1a5010a35c4ee ;487703d7b867d3b53e91ec1f11f3b7e588552b0b6c011671b4dc82d669fc39d9 ;80847fa36472ee08b65abe415b078d1877a15dd4d5fe16935feb21a62c0cd7ca :d8ade83d2dcc64e38b42d877f101e746dc8a07d98c4b1f68c4bbbbe3e3f7499d :8de0b3c47f112c59745f717a626932264c422a7563954872e237b223af4ad643
$P_{display}$	derived,partial	f1195066a0177e3ab54eaf438f81b17fbc51640f37aba065b7f1a5010a35c4ee ;487703d7b867d3b53e91ec1f11f3b7e588552b0b6c011671b4dc82d669fc39d9 ;ca4ff9f69ebf6094e301b59ab7a7aa8df1f1a074e252468f3b17875ecca605b6 :X
P_{total}	derived,partial	f1195066a0177e3ab54eaf438f81b17fbc51640f37aba065b7f1a5010a35c4ee ;487703d7b867d3b53e91ec1f11f3b7e588552b0b6c011671b4dc82d669fc39d9 ;cfd631d6d73bdc23de589ff89c42765747cbe0c21c9b0eb3709482bb9507696a :1d7c41df24dc5bad67d28167a2e97f6a58d0bfbc65de3d5acdbd1ac9c1b409e0 :a0944f4dd78ee880ba109747205b1b8b525ab7afc177f982aff8f79afc76b920 :50ffad5e7f9fe1945f9b39893ceee96301c543b282a9abcd55d9c97f15c4d73f :X

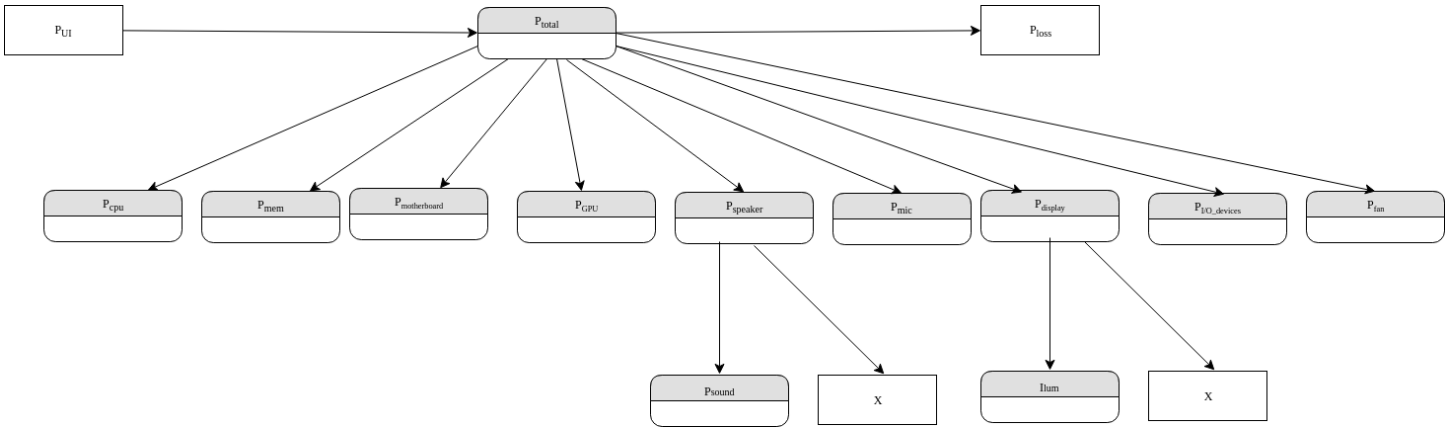

Fig. 3. Example of some relationships between power equations of computer

the program, hash of the code can be linked with the hash of mathematical equation. Linking theorem and code hashes means that finding the error in the equation would automatically detect the code which has to be modified improving computer security infrastructure. Linking conditions with theorem could be used to test if data has the property required by the theorem and the mathematical model.

4 AI INFRASTRUCTURE AND SECURE DESIGN PRINCIPLES OF AI

ARTIFICIAL INTELLIGENCE is a system or device which can handle any task performed by humans or backed by human intelligence. To create AI system designers have to gather a multitude of data for analytical purpose which will be used to derive an AI model. The blockchain technology is a core infrastructure in new industrial revolution ("data revolution") for building AI models which will replace standardized human labor in some areas.

4.1 Data revolution

William Edwards Deming: "In God we Trust, all others bring data."

The data revolution is probably the biggest revolution in the history of human society and it combines the elements of past monetary and industry revolutions. Data revolution using blockchain will gather a multitude of data of superior quality. How will the society use this data has jet to be seen and measured.

4.1.1 Monetary revolution

Monetary revolution was introduction of money in social interaction as a medium of exchange and measuring of

human interaction and spot price of different assets. Introduction of long and short term loans provides a method of measuring the long term and short term value of the asset by the financial institutions for making future investment decisions. Information based on the financial transactions is one data source of economic activity. The second source of trustworthy data for measuring economic activity are taxation records. Taxation and financial data should be trustworthier than statistical questioner data, because in the most cases of asset exchange and ownership taxation data is a percentage of spot price or asset value of the agreed price.

4.1.2 Industry revolution

Industry revolution was introduction of machines as a supplement of human labor. First machines were constructed based on the simple equations and engineering limited by available resources and tools of lower complexity. The new generation of machines were engineered to improve some characteristics of the previous generation. The final decision of using a machine and operational control of the machine was under control of human operators.

4.2 Evolution to AI

4.2.1 Machines

Machines used in the past can be described as a primitive AI for substitution of physical blue collar labor or performing standardized white collar labor based on human defined rules and work flows. They were constructed on the known equations and principles of that time.

The process of such large system infrastructure construction had multiple serial human decision checkpoints from original idea, research, initial analysis, presenting idea to investor, building the prototype, getting regulatory approval, construction of technical and social operational infrastructure, quality control and user.

Computer programs today also have some properties of machine learning. They are developed by human developers and based on developers understanding of received request and resources available. Most computer programs were designed and programed by humans with partial mathematical verification or no mathematical verification. Such development process produces random bugs and computer behavior, one example being the blue screen of death. Blue screen of death happens when data input produces data output outside of the predicted range of OS operation. The computer model based on the developers understanding has stopped, dump of data and reset is done to resolve it.

Some construction designs in the past proved to be inadequate taking into account social, environmental and financial consequences(for example Chernobyl and Fukushima nuclear power plants disasters), so similar problems should be avoided in future designs of critical infrastructure systems.

In mathematically proven construction designs the output and operational range is known and the system reliability depends on system architecture design.

4.2.2 System reliability

There are several system architecture designs which determine the overall system reliability R_a built with n components of individual reliability $R_1..R_n$:

The hash of symbol TABLE 3 of reliability is:

3fc8fe3d54ad2c6dc301a3f8414c175b2774073d3f3965b212a325eddc6dcbe9

For simplicity in this example condition will be defined as a string ”operational_environment”. The SHA2 is calculated as SHA2 of a string ”operational_environment”:

487703d7b867d3b53e91ec1f11f3b7e588552b0b6c011671b4dc82d669fc39d9

Additional condition defined for serial systems is a string ”serial connection of all components”, while for the parallel systems is a string ”parallel connection of all components”.

The additional condition for serial systems has value:

53a41fb36ae50754ba4504884ec3643a9460f60ac4fa102d8a0d43f8d1e73e13

The additional condition for parallel systems has value:

d3fed119d3ed33857630686389a501a8e39ea9d5582deb3d77c3402ce64d814e

1) serial system: in the case of a failure of one component the whole system will fail, the characteristics of serial systems is that overall system reliability R_a will be lower than the least reliable component and is calculated

$$R_{\mathrm{a}} = \prod_{i=1}^{n} R_i \tag{8}$$

8e32b43f74a6d6fce0652fe289e1822da9a737b4686854826fec5b234c03c2eb

2) parallel system: in the case of a failure of one component the whole system won't fail because the other identical component can perform the same task, the overall system reliability of parallel system is calculated

$$R_{\mathrm{a}} = 1 - \prod_{i=1}^{n} (1 - R_i) \tag{9}$$

7863a8eca5cf8c5e24e5c5ab0954ad0353f7af9e7102eccf489eab2d6f624c60

In the case that the individual reliabilities of components are R_i=0.97, to achieve overall reliability R_a=0.999 the parallel system should have two components:

$$n = \frac{ln(1 - R_{\mathrm{a}})}{ln(1 - R_{\mathrm{i}})} = \frac{ln(1 - 0.999)}{ln(1 - 0.97)} = \frac{-6.907755}{-3.506558} \cong 2$$

4.2.3 Machine learning

Machine learning is a field of computer science that uses statistics to build computer models with the ability to ”learn” from data without being explicitly programmed. Machine learning is used for probabilistic-numerical modeling of unknown equations which means that for the predicted solution there can exist a significant deviation from the measured/real result.

TABLE 3
Symbol definition table of reliability

symbol	description	symbol SHA2
R_i	reliability of component	00c00f8247bf728a0ce5a3561853be244d19ee2113a5b04df001dba88679dee8
R_a	reliability of system	33462a02fcce1a84ec45ccb047e91bbc3854a2d6e384ce8f0dc78268af5b1290

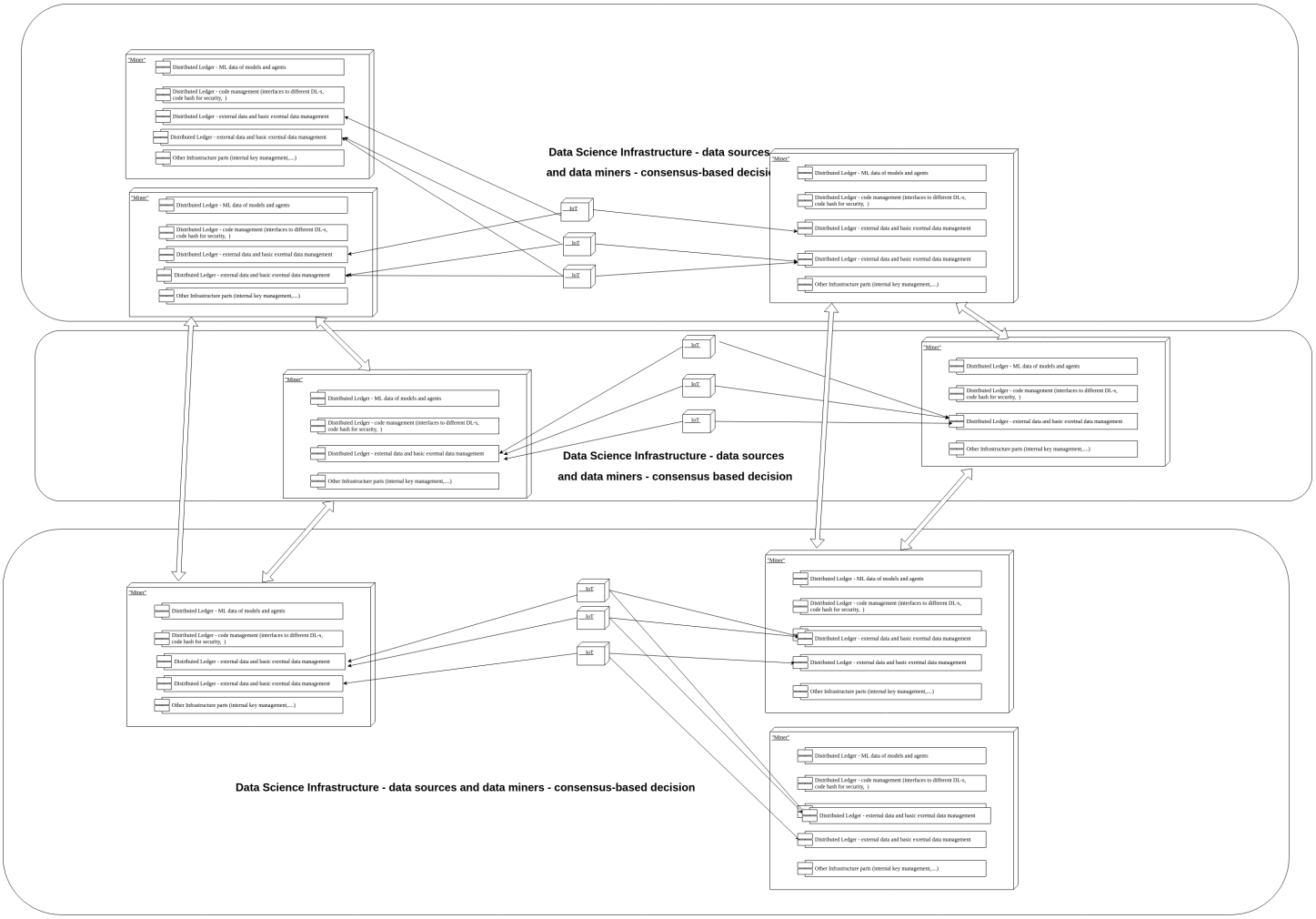

Fig. 4. AI consensus based decision infrastructure

4.2.4 AI security design

The machines in the past where used and operated by humans. The machines which could have significant consequences were operated and controlled by human operators in a serial system design, and supervised by dedicated personnel in parallel system design. The separation of duties principle meant that the human error based on wrong data or decision would be detected with higher probability and disaster contained to the local area. There were multiple such machines with multiple operational control and supervision centers which constituted critical system infrastructure. Having separate infrastructure reduced overall systemic risk.

AI is an computer representation of mathematical model, the security of AI can only be achieved using mathematical and physic principles. AI substitutes human decision process and checkpoints and has an ”evolution” ability making security design and control of AI an important aspect.

Using mathchain presented before introduces following security controls in AI design:

- detail mathematical equation specification of AI design
- detail condition specification of ”operational range” for AI
- detailed historic linkage between mathematical model (equations, conditions) and computer code
- easier detection of error propagation and flaw replacement
- historic audit of mathematical and engineering construction by scientists and regulators
- reduced time of research and education of human AI designers
- compliance with GDPR ”right to explain” regulation

Laws of physics are important security element of AI

because they limit the possible input and output data range and influence of specific AI decisions. Physical characteristics in nature are described with International System of Units which comprise a coherent system of seven basic measurement units. Other units of measurement used in physics can be derived from this 7 basic units using known equations. Using mathchain in physics equations would provide detail information of physical operational range limits of AI.

Most AI systems currently under development are based on machine learning. Machine learning is used in modeling multidimensional problems with unknown and not fully understood dependencies between the variables. This property enables decision-based adversarial attacks. To reduce the surface of such attacks we can revert to the previously mentioned principles. The AI decision which could have significant consequences would be based on consensus decision of multiple AIs built for the same purpose, but trained in different ways. There can exists multiple such AI systems with multiple operational control and supervision centers which constitute critical system infrastructure. Having separate infrastructure reduces overall systemic risk. The design is presented on figure 4. The design enables online evolution and introduction of new AIs with new security features. The wisdom of previously proven AI machines can be preserved and used for consensus control of new generation AI machines.

5 Conclusion

Security is based on understanding and controlling environment and manging risks. To achieve this the environment has to be described using mathematics and physics. Mathematical blockchain enables linkage between code and data in AI, sand boxing and better understanding of AI operational range and understanding possible consequences in the case of AI "blue screen of death" or adversarial attacks on AI.

This paper has presented the idea of using mathematical blockchain principle in research and design. Research is an important element of every area and has its formal description in mathematics. Understanding inter-linkage of mathematical conditions and derived equation is an important element of formal design. Paper described using blockchain principles of serialization and autocorrelation for creating uniquely indexed derived equations and linking them with symbol definition tables, conditions and basic equations from which they originated.

Using mathematical blockchain enables easier understanding and versioning of mathematical model, linking similar and derived research. Additional benefit of this methods is better education and understanding of mathematical origins of AI by new generation of students.

Blockchain technology changed the quote of W. Edwards Deming to: "In God we Trust, all others grant me access to locally replicate your data, code and mathematical blockchain database".

Appendix A
Security analysis of attack vectors possibilities on air-gap computer

For theoretical training exercise lets assume that we want to obtain private key stored on air-gapped computer. The owner is going to reinstall his air-gapped computer because of OS file system corruption. The wallet owner ordered OS installation DVD to his home address. We can replace the original DVD with with virus installed DVD which can detect when wallet software is used to access the private key. Virus will store the key in its own storage. Because the computer is not connected to the network we need to find out what are theoretical ways to send the 256 bit private key from the computer to the outside world without modification of hardware components.

A.1 International System of Units

Physical characteristics in nature are described with International System of Units which comprise a coherent system of units of measurement built on seven base units shown in table. Other units of measurement used in physics derive from this 7 basic units using known equations.

Unit name	Unit symbol	Quantity name
meter	m	length
kilogram	kg	mass
second	s	time
ampere	A	electric current
kelvin	K	thermodynamic temperature
mole	mol	amount of substance
candela	cd	luminous intensity

The SHA2 of SI symbol definition table is:

d8ed77652122ff9bb5c4395ce9ffac7189e4303c7c1f33ea1791b15c1155269d

A.2 Air-gap computer

Air-gap computer is a computer not connected to the Internet. Using only components present at the site we have to extract the data. To achieve this we must return to the basics.

Every change requires energy. Power is the amount of energy transferred per unit of time. Electric power of electrical appliance is defined:

$$P_{UI} = UI \tag{10}$$

8d463aa60853f719b7ececcdec1b16fa2b950c5db03ef40c35e269ce7343d4ab

Air-gap computer has different input and output components connected and some of them are:

- power unit
- keyboard
- monitor
- DVD drive
- camera
- speaker
- microphone

We have to modify the behavior of the mentioned devices to send the private key to the outside world without entering the premises.

A.3 Using different SI units for transmission

A.3.1 Using electric current for transmission - Power unit

Every computer needs electrical power for operation. Desktop power supplies can have a maximal power usage rating from 200 watts to 1800 watts based on type of components installed and power supply purchased. The Energy Information Administration (EIA) estimated in its 2013 Annual Energy Outlook report that approximately 3% of total residential electricity consumption is due to computers and related equipment. The voltage (U) is controlled by electrical power supplier and can be considered constant. The power supply is an AC/DC converter. Current symbol I has index a83dd0ccbffe39d071cc317ddf6e97f5c6b1c87af91919271f9fa140b0508c6c in Table 1 and can be used to search the equations in TABLE 2 which can be used to transmit the data.

$$P_{UI} = UI \tag{11}$$

8d463aa60853f719b7ececcdec1b16fa2b950c5db03ef40c35e269ce7343d4ab

Automatic meter reading system of electrical power is today a standard. If we can obtain the electrical power consumption data, we can use power meter as a reading sensor for data transmission sent by virus. To be able to do this we have to control the power consumption P_{total}.

$$\begin{aligned} P_{total} &= P_{CPU} + P_{mem} + P_{motherboard} + P_{GPU} + P_{speaker} \\ &\quad + P_{mic} + P_{display} + P_{I/O\_devices} + P_{fan} + P_{loss} \end{aligned}$$

cfd631d6d73bdc23de589ff89c42765747cbe0c21c9b0eb3709482bb9507696a

Every component has an individual power consumption equation.

For CPU power consumption is:

$$P_{CPU} = (mV) + (\alpha E_{short\_circuit} f) + (0.5\alpha C V^2 f) + P_{cpu\_loss} \tag{12}$$

1d7c41df24dc5bad67d28167a2e97f6a58d0bfbc65de3d5acdbd1ac9c1b409e0

Increasing CPU load automatically increases CPU voltage and frequency of the CPU. Increasing the CPU voltage and frequency increases the electrical power consumption by the CPU, but also the electrical power consumption is increased by the CPU fan and power supply fan because of increased temperature. Active sensors like camera and microphone also increases power consumption.

From historic data of electrical power automatic meter reading system it is possible find the time with at least consumption and consumption variation (probably at night or during vacation). That time can be used for data transmission. The virus would schedule the programs in the way to control the CPU power consumption so that high power consumption would represent bit 1 of private key and low power consumption in a case of 0.

The virus behavior could be: After the owner installs the OS wait until the user inputs the private key (from the paper) to the wallet. Store the private key on separate space of disk. Override OS poweroff operation with standby and schedule the power-on at the scheduled transmission time. For every bit of private key, if the bit is 1, increase power consumptions as possible, if 0 ,than keep CPU idle. Using remote reading of electrical power meter private key could be retrieved.

A.3.2 Using length, mass and time for transmission - Speaker and microphone

Sound are waves of pressure fluctuations.

The SHA2 symbol TABLE 4 for pressure is:

5b20a0b08ba91e842c03d16818aca1d4ae8f40b6b534afe3dcf40891566c4b6d

The SHA2 condition for pressure is:

487703d7b867d3b53e91ec1f11f3b7e588552b0b6c011671b4dc82d669fc39d9

The equation for the pressure is with measuring unit:

$$p = \frac{F}{A} \tag{13}$$

eeb150d933bf9a0ca9e00fd49ab375e83ae46e60a1be5239a9b9a8970ad36d61

$$1Pa = 1\frac{kg}{ms^2} \tag{14}$$

Looking at the hash/index of a velocity of pressure wave in TABLE 1 4b15a5a6867c654691000e823ae4a17ee31f636de5caf3796b6531dfbe854160 the virus could control speaker for data transmission.

The virus behavior could be: After the owner installs the OS wait until the user inputs the private key (from the paper) to the wallet. Store the private key on separate space of disk. Activate microphone. Use a burner phone to call the account owner during his operation on the computer. The virus using microphone detects the phone ring sound at the predefined time and send a private key as a sound sequence using speakers after the phone line is open.

A.3.3 Using luminous intensity for transmission - Monitor

Looking at the frontal luminace symbol in TABLE 1 d8ade83d2dcc64e38b42d877f101e746dc8a07d98c4b1f68c4bbbbe3e3f7499d , it is visible that this hash for the symbol is present in the TABLE 2, and luminous intensity is derived from frontal luminace and surface. The virus could control display for data transmission.

The virus behavior could be: After the owner installs the OS wait until the user inputs the private key (from the paper) to the wallet. Store the private key on separate space of disk. Override OS poweroff operation with standby and schedule the power-on at the scheduled transmission time. For every bit of private key select appropriate color. Use optical luminance meter for measuring luminance in the room through the window.

A.3.4 Using thermodynamic temperature for transmission - CPU

Increasing CPU load automatically increases CPU voltage and frequency of the CPU. Increasing the CPU voltage and frequency increases the electrical power consumption by the CPU and thermal output.

The virus behavior could be: After the owner installs the OS wait until the user inputs the private key (from the paper) to the wallet. Store the private key on separate space of disk. Override OS power-off operation with standby and schedule the power-on at the scheduled transmission time. For every bit of private key, if the bit is 1, increase power consumptions of components with high thermal output, if 0 ,than keep CPU idle or shutdown the computer for predetermined interval. Using thermal camera for remote reading of temperature retrieve the private key.

TABLE 4
Symbol definition table of pressure

symbol	description	symbol SHA2
p	pressure	148de9c5a7a44d19e56cd9ae1a554bf67847afb0c58f6e12fa29ac7ddfca9940
F	force	f67ab10ad4e4c53121b6a5fe4da9c10ddee905b978d3788d2723d7bfacbe28a9
A	surface area	559aead08264d5795d3909718cdd05abd49572e84fe55590eef31a88a08fdffd

A.4 Conclusion

In theory the possibility of transmitting the private key from air-gaped computer using 6 of 7 SI units of measurement was described. Theoretically it was demonstrated that using different sensors or other devices in the area it could be possible to retrieve a private key without the knowledge of the user. Now, lets look if there exist any research papers which describe such forms of attacks.

The papers describing such proof of concept of attacks are available at "Cyber-Security Research Center Ben-Gurion University of the Negev, Israel" on "Air-Gap Research Page". The paper "Neighborhood Watch: Security and Privacy Analysis of Automatic Meter Reading Systems" shows the proof of concept for hacking the Automatic Meter Reading Systems.

ACKNOWLEDGMENTS

This paper is dedicated to all the professionals trying to secure and keep operational critical infrastructure of the society and next generations who will design the future critical infrastructure. I am grateful to Sanjiv Das for discussion and comments.

REFERENCES

[1] ECB, *CYBER RESILIENCE OVERSIGHT EXPECTATIONS (CROE) FOR FINANCIAL MARKET INFRASTRUCTURES*, Frankfurt, EU: ECB, 2018., SHA2:15500f840303c4b664bd887b3ecb0758e812b761ae6ec77c1957816be692f100

[2] ECB, *TIBER-EU FRAMEWORK How to implement the European framework for Threat Intelligence-based Ethical Red Teaming* , Frankfurt, EU: ECB, 2018.,SHA2:f95857e29cfba418c3d81c02bd9bd13c1791f58fc009e273b66f85f308166e51

[3] Future of Humanity Institute, University of Oxford, Centre for the Study of Existential Risk,University of Cambridge, Center for a New American Security, Electronic Frontier Foundation, OpenAI, *The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation* , 2018., SHA2:f8876c43ae1ac11ea32b91fb1c4cf1b297f6b3fc20422b32cb228a8abfaf1b08

[4] Ben-Gurion University of the Negev, *Air-Gap Research Page*, https://cyber.bgu.ac.il/advanced-cyber/airgap ,Israel, 2018.

[5] Matthew Travers, Newcastle University, *CPU Power Consumption Experiments and Results Analysis of Intel i7-4820K*, Newcastle, 2015, SHA2:b780e18ebf78ec2f3823317200d7a7f4994b78cd196acf81ea2c9256747e1809

[6] Ishtiaq Rouf, Hossen Mustafa, Miao Xu, Wenyuan Xu, Rob Miller, Marco Gruteser, *Neighborhood Watch: Security and Privacy Analysis of Automatic Meter Reading Systems*, 2012., SHA2:1017cba810efb837b047ed6bc6a27e2cbe37445d0e3999f7441833c68e1dcd7e

[7] Satoshi Nakamoto, *Bitcoin: A Peer-to-Peer Electronic Cash System*, 2009., SHA2:b1674191a88ec5cdd733e4240a81803105dc412d6c6708d53ab94fc248f4f553

[8] Vitalik Buterin, *A NEXT GENERATION SMART CONTRACT & DECENTRALIZED APPLICATION PLATFORM*, 2015., SHA2:712053058a2af2f79e332f4222bdfd5d6376b94e0d85c53a4255a608043f078a

[9] Andra M. Maechler,Swiss National Bank, *The financial markets in changing times Changes today and tomorrow: the digital future*, 2018., SHA2:b66d64292a2bf3be543fad1b9556b64bb42292c84ad3a06174a1300052400c41

[10] Financial Conduct Authority, *FCA fines bond trader 60,000 for market abuse*, 22/11/2017, SHA2:d8d156721c3f0da7e6592e979018893d46f3d4c904f1f95cee124e226e484eae

[11] Del Rajan, Matt Visser, *Quantum Blockchain using entanglement in time*, 17/04/2018, SHA2:062592f44df9e4bcf3237526d3c6c9a6c68b12448fd4a8940073da8096929590

[12] Bank of Japan, *Responses to the Great East Japan Earthquake by Payment and Settlement Systems and Financial Institutions in Japan*, 01/10/2011, SHA2:8ee379408095787a086a3ce57b858ed0ae5580dd370a0846268b451a69cd3c49

[13] ENISA, *Security vs Performance Discussion with the Return of Spectre Vulnerability*, https://www.enisa.europa.eu/publications/info-notes/security-vs-performance-discussion-with-the-return-of-201cspectrum201d-vulnerability ,25/05/2018,

[14] ARSTECHNICA, *Remediating Fukushima 'When everything goes to hell, you go back to basics'*, https://arstechnica.com/science/2018/05/remediating-fukushima-when-everything-goes-to-hell-you-go-back-to-basics/ ,11/05/2018,