SEMANTIC ANALYSIS AUDIT IN TRIPLE-ENTRY ACCOUNTING SYSTEMS BASED ON BLOCKCHAIN

Sînică ALBOAIE[1,2], Alexandrina RATA[1], Emil HOROMNEA[1], Mircea VAIDA[2]
Alexandru Ioan Cuza University of Iasi, Iasi[1], Technical University of Cluj-Napoca[2],
General Berthelot 16[1], 26-28 Baritiu str. Cluj-Napoca, Romania [2], abss@axiologic.net, alexandrina.rata@gmail.com, emilhoromnea@yahoo.com, mircea.vaida@com.utcluj.ro

Abstract: Using the blockchain technologies and keeping at least three parallel accounting records, separately located, has as result the concept of triple-entry accounting. The triple-entry accounting is not an alternative for the double-entry accounting, but rather part of a solid system when the two types of accounting are combined. Decreasing the cost of managing trusty accounting by using blockchain is a significant advantage, addressing future requirements of accounting. In this paper, we are proposing an audit method based on semantic analysis of blockchain accounting data to gain resilience of accounting against improper (fraudulent) use both on the public and private sector.

Keywords: *blockchain, triple-entry accounting, semantic analysis, audit*

I. INTRODUCTION

The accuracy of data in accounting records and, respectively, avoiding errors are of critical importance. Eliminating and preventing faulty entries were the cornerstones of accounting development, from single-entry records to double-entry accounting records. The single-entry accounting is a very simple method, but prone to multiple, even accidental, errors. In this case, error traceability is difficult, hard to correct and possibly fraudulent. The double-entry accounting records granted the accounting system considerable improvement, enabling the identification and correction of errors. A side effect of this improvement was the prevention of accidental errors and partial determent of fraud. Unlike single-entry, the double-entry accounting records may not be added by simple mentioning an entry, each entry must be validated both actively and passively. This counterparty partially ensures the accuracy of records. As the need for even higher trust in the quality of accounting records arises, the concept of triple-entry accounting is introduced.

The triple-entry accounting [2][3][4][5] also implies an external party that is cryptographically sealing each entry and, eventually, saves the data in a Blockchain distributed ledger [6][7]. As long as the accounting records are distributable stored and cryptographically sealed, it is practically impossible to modify or delete them.

II. TYPES OF ACCOUNTING DATA

Bearing in mind that the Blockchain ledgers are public, the issue of data privacy arises, while daily transaction is comprising a great amount of important information. The confidentiality of accounting records implies restricted access to use or share the data without the authorization of the accounting services client, except when otherwise provided by enforceable low. One may identify from accounting records and assigned transactions the personnel list and their wages, data which may be used for employee theft, corruption or duplicating the wages list. The payment bills offer information on the credit score, the company suppliers' list, the buying prices (from the invoice found in the Blockchain), the falling due payments. These details may be used for suppliers' theft or corruption. The collection bills comprise details on customers' identity, the selling prices and customers' due dates – all these data may be used by the competition in order to steal customers (by employing the same or lower prices, and larger due margins) or to inflict reputation damage upon the initial supplier by releasing confidential customers' data. A synthesis of data privacy risks is under laid in table 1.

Table 1: Records categories that may be stored in blockchain

Records category	Private information	Potential risks
Wages	Employees list; Salary	Employees theft; Employees' corruption for acquiring commercial secrets; Wages list theft.

Manuscript received March 9 , 2018; revised April 27, 2018

Purchased goods	Suppliers list; Purchased goods/ services lists; Purchasing prices; Due payments.	Suppliers' identification; Suppliers' corruption for purchasing obstruction (holding of goods, raising prices); Strategic suppliers' theft.
Sales	Clients' lists; Provided goods and services list; Sale prices; Due payments.	Customers' identification; Customers' theft by offering better terms (lower prices, larger due payments margins).

III. ACCOUNTING FROM THE PERSPECTIVE OF 3 TECHNICAL DIMENSIONS OF BLOCKCHAINS

In this chapter we will describe three dimensions of blockchain systems, as well as the way they may impact on the accuracy and the application of accounting records.

Table 2: Dimensions of the blockchain

Dimension	Risk prevention measures
The "topological structure" dimension	Establishing a data replication system controlled by various autonomous agents with different interests. If a distributed government system is unlikely, one should strive to a maximum degree of decentralization.
The "data access" dimension	Using hybrid blockchains that are limiting data access and, eventually, are enabling advanced storage capabilities such as "data self-sovereignty" [11].
The "method of truth" dimension	Employing a semantic audit method as it is detailed in this paper.

The "topological structure" dimension

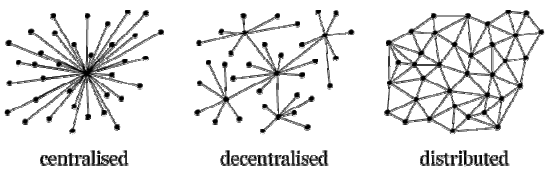

Figure 1: The centralized, decentralized, distributed [1] models

The topological structure represents the way the connection between various system components is carried out. The possible values of the topological structure are: centralized, decentralized, and distributed. Without getting into advanced technical details, Figure 1 depicts 3 architectural paradigms, at the same time enabling insight on the structure, installation and operation of software systems, as well as on the way the user`s experience is capitalized on and the control and social influence are carried out. For application developers, the concept of blockchain based app represents a shift in paradigm for the way software engineers will write the apps in the future. It is also a catalyst in creating decentralized and distributed applications. The decentralized and distributed applications are probably the next step in the development of software architecture. However, this is not a computational phenomenon only. The decentralized applications are meant to trigger a decentralizing trend on social, legal, governmental and business levels, as a result of change on cultural and social environments aiming at general decentralization and shifting the influence power toward networks margins.

The "data access" dimension

The users are considering the blockchain systems as a new type of distributed database, replicated on large scale, which stores their transactions and the according financial effects and not only. Compared to a regular database where system administrators are holding absolute authority when modifying data, the blockchain systems are adding a cryptographic fingerprint on past events and they are distributing these fingerprints to a large number of participants, making practically impossible to change past events in the database. Depending on the replication level of this database and on the level of public access, there are three types of blockchains:

- Public blockchains, well known examples are the Bitcoin [12] and the Ethereum [13] systems;
- Private blockchains, banks and companies have begun various experiments aiming at bolstering the security of their own systems employing blockchain (e.g.: Hyperledger);
- Hybrid blockchains, our research conducted during the PrivateSky [10] project led us to a mixt architecture based on executable choreographies [8], [9], combining public and private blockchain. The URI (Uniform Resource Identifier) concept is the cornerstone of the PrivateSky system, which may be envisioned as an identification key of certain resources (a resource may be simply the name assigned to a value [11], as well as the concept of "validation space" which may be assimilated to a cryptographically sealed key collection that are added in blockchains.

The "method of truth" dimension

The possible values of this dimension may be placed on an

axis where at one end there are the system where data accuracy is validated completely mathematic, and at the other the systems where the accuracy is validated by human operators abiding to processes and rules. One must understand that the blockchain is not panacea by any means. In any human operated system, the data accuracy depends on human management. On the other side, the technical implementation method may reduce or, by contrast, increase the chance of human error or system abuse. We meant to underline in this dimension the authors' belief that the blockchain or any other technical approach are not able to completely solve trust issues and should not constitute an alternative to striving toward better social systems. The Bitcoin crypto-currency was a way to undergo "trustless" financial transactions where there is no need for an advocate to validate the transaction. Although, mathematically speaking, the statement stands, completely "trustless" transactions are not always desirable and may become an obstacle to broader acceptance instead of a strong point.

For example, it is worth mentioning that cryptocurrency theft is practically irreversible, if the authorities cannot achieve physical access to thieves' private key. This irreversibility is a common denominator of physical and crypto currencies, but it is not always desirable. From the accounting perspective, for example, the opposition of businesses against systems that may render them vulnerable to administrative abuse, information leakage or, even, accidental data and value loss must be considered. From the authors' perspective, the favorable feedback of blockchain based systems enabling "trustless" transactions may easily become negative, as these systems require unacceptable levels of trust in the goodwill of governmental agencies.

As any human technology, the blockchain may provide a never before met level of transparency and accuracy regarding economic transactions, which may ultimately lead to large scale prosperity, yet it may always transform in an instrument of coercion and control. Paradoxically, even if these systems might be abusively and discretionary employed by those in power, they may also record – by means of historical data correlation – information against the same.

Table 3: Potential advantages of blockchain

Client	Advantage description
Companies	Reducing accounting related expenses by automated accounting and audit; Eliminating certain fraud and misinterpretation categories; Transparency when dealing with the state, suppliers and clients.
Governments	Decreasing corruption; Prevention of public servants' abuse; Transparency when dealing with citizens and national and transnational companies; Black market elusion.
Individuals	Transparency when dealing with the state, depository institutions, suppliers and employers.

The advantages of automated accounting ledgers, run by blockchain technology, are multiple, both for individuals, companies or even governments. The transparency and real-time monitoring of transactions (with delays of at most several minutes) are the main advantages, simplifying decision making, as well as transactions' control and validation. Several strong points and risks depending on each type of client are listed in Table 3.

As depicted in table 4, the risk prevention may be achieved by approaching the three blockchain systems' dimensions: the "topological structure" dimension, the "data access" dimension, and the "method of truth" dimension.

IV. SEMANTIC AUDIT OF ACCOUNTING DATA

The audit of blockchain stored accounting data will require the development of scripts and programs able to code real world knowledge and invariants and to verify if their properties are observed by the blockchain records. By semantic audit we do not observe abiding to tax legislation and the account operations processes. These processes may be addressed by smart contract and are validated by code assessment or automated testing according to software development methodologies. By semantic audit we approach problems related to the interpretation of blockchain stored data (data semantics).

Employing technical languages of semantic web (such as RDF, OWL) for verifying data semantic properties is not plausible. That is why we believe it is necessary to develop a DSL (Domain Specific Language), enabling even auditors and accountants to learn and use it. This language should be scalable and as close as possible to natural language.

We envision scripts approaching the following cases of semantic audit:

- The audit of transactions traceability;
- The prevention of faking accounting ledgers (fake or partial accounting records);
- The prevention of selective interpretation by authorities.

We propose in this paper a method of semantic audit we named BLOCKAUDIT. We are presenting below an informal description of several real cases that led us to develop our audit method.

In manufacturing industry, the validation key should be the raw stock. A steel bar manufacturer may not produce a much larger quantity of deliverable than the equivalent of purchased raw material. If annual steel bar sales (of all diameters) would sum 375 tons of steel bars, and the sum of purchased raw material is 280 tons while the initial stock is 0kg, even if the technical specifications are taken into account, the difference in quantity – to be highlighted by the blockchain – suggest black market purchase (often without quality certificates) or lack in steel bars quality (by adding unregistered alloys). When identifying such situations, an auditor may verify and validate the supply chain in the blockchain ledgers in order to spot irregularities. The raw materials inflow may be verified based on transport orders

and incoming (the number of runs and laden weight of each transport) in the considered interval. Taking into account this variation, even if the producer holds papers summing the 375 tons, but it purchased transportation services for 280 tons and there are no invoices or justifying documentation for transport included raw stock, it is relatively simple to prove (using blockchain technologies) an error or fraud.

Informal model

The types of above mentioned verifications require two types of information: information on events and structural information. Events information are the type of transaction with the exterior. We have identified three types of transactions: purchase, sale and transport (purchasing raw materials, sale of products, and transport of raw stock or finished goods). Structural information is information on products and services provided by the economic activity and the relationship with raw materials quantities, number of parts, work hours and quantity of sold finished goods or provided services.

The execution of the semantic analysis is based on generating a knowledge structure (graph data) about the real world using structural information. When verifying, to the knowledge graph nodes will be assigned quantitative values resulted from information on events. In other words, the dynamic inputs in the verifying system are events. The time moments specific to verification are chronological data of events occurrence. Each event triggers a quantitative modification in one of the knowledge graph's nodes. Formally, we may consider that the knowledge graph's nodes will be labelled with a quantitative value.

Syntax

We will define formally the knowledge definition language as an internal DSL in a programming language (for exemplification and demonstrations we employed the syntax of the Java Script language).

defineProduct(name)
Defines a finished product node in the knowledge graph;
defineRawMaterial(name)
Defines a raw material node in the knowledge graph;
defineComponent(name)
Defines a part node in the knowledge graph;
composition (P, C, min, max)
Defines a structural relationship between a finished product P and a part or raw material C;
event(t, T, X, amount)
Defines the occurrence of a T type event, at a t type moment, the given event affecting a finished product, raw material or part X. The amount parameter defines a quantity of raw material, finished product, or part (X) that participated to a T type.

Table 5: Proposed primitives (syntactic elements)

The language we propose is quite user friendly and it will consist of Java Script functions call. An auditor (assisted by a programmer) will be able to use a series of primitives (functions) in order to define knowledge. Generating events must be made by programmers by writing blockchain specific code. For exemplification, tutorials and system testing, a primitive able to generate test events will be defined.

The semantics of the analysis method

We layout below a simple example meant to explain the way the primitives described above are to be employed:

```
/*Defining knowledge for running the verification process
based on the 3 events should generate an alert on
insufficient amount of copper to produce the 10 sold
statues */
defineProduct ("Brass statue")
defineRawMaterial ("Copper")
defineRawMaterial ("Zinc")
composition ("Brass statue", "Copper", 101, 105)
composition ("Brass statue", "Zinc", 5, 10)
/* Defining events */
event (buy, "copper", 1000)
event (buy, "zinc", 100)
event (sell, "Brass statue", 10)
```

Table 6: A simple example to clarify the concept from the semantic perspective

Running the semantic analysis implies two types of labeling of knowledge graph's nodes. The first way of labeling is to label with quantities directly resulted from the quantities emerging from the actual events. Intuitively, this type of labeling is assimilable with stocks inventory. It's easy to comprehend that in a well-defined system where all the events are properly generated, at each moment in time, one may know the inventory status (raw materials quantities, amount of parts or finished goods). However, internal management systems cannot store all micro-events required for different production stages. For example, if given a 1kg aluminum bar, 10 grams are subtracted in order to produce a certain part, it is sensible that the internal systems will not be able to generate events for each produced part. That is why only the incoming and outgoing events are relevant (those that leave legal traces as invoices, bills, etc.). The inexistence of events, as well as various events related to rejections, led to the necessity to model the system through the composition relation. The composition relation enables the semantic analysis system to dynamically compute an acceptable minimum and maximum associated to nodes in the semantic graph. Thus, the semantic analysis may be perceived as a stock simulation able to estimate with sufficient accuracy the admissible stock limits at any given moment of the past.

Any breach of acceptable limits will lead to identify errors suggesting accounting management faults or possible fraud. However, defining the knowledge graph is susceptible to human error as it implies know-how on the business model and production processes. In other words, the knowledge graph is formally and exactly describing all the assumptions the auditor made about the on-site reality, while the system searches and identifies the differences between the block chain accounting ledgers and the information generated by auditors.

By formalizing this knowledge, the audited company may employ external experts to identify eventual interpretation errors made by auditors. Thus, the audit is more accurate and through.

V. CONCLUSIONS

The triple-entry accounting is made possible by using the existent block chain infrastructure (with minor adjustments). It is also highly anticipated by the public, due to promises of complete transparency and streamlining. Using it would lend increased substance to the financial score of its participants. Furthermore, the triple-entry accounting ledgers will bolster development by reducing partners' assessment related costs (suppliers, clients, financial institutions). However, the triple-entry accounting does not apply to any accounting record and does not completely eliminates risks, but rather adopting it would greatly benefit the safety and predictability of economic transactions. The audit of future blockchain based informatics systems will involve both auditors and programmers and will enable generating very accurate audit scripts able to integrate complex information on real world. This article aims at providing a possible technique in order to facilitate the communication between these specialists, by enabling the auditor to explain very practically (and well formally defined) what is the focus of an audit of the blockchain stored data.

Furthermore, the suggested model explicitly underlines the assumptions the auditors make on business models (the definition in the knowledge graph), while these models may in turn be validated or invalidated by external experts.

REFERENCES

[1] Paul, B., *On Distributed Communications Networks*, Santa Monica, CA: RAND Corporation, 1962
[2] Ijiri, Y., *Triple-entry Bookkeeping and Income Momentum,* ISBN 086539041X, 9780865390416, 1982
[3] *Triple-Entry Accounting Methods, Warren Henkem*, 1994, http://www.warrenhenke.com/writing/essays/triple-entry-accounting
[4] *Triple Entry Accounting ,Ian Grigg*, 2005, http://iang.org/papers/triple_entry.html
[5] *Triple-Entry Accounting And Blockchain: A Common Misconception*, Forbes, 2017 https://www.forbes.com/sites/forbesfinancecouncil/2017/11/28/triple-entry-accounting-and-blockchain-a-common-misconception
[6] *Blockchain Technology A game-changer in accounting?,* Deloitte, 2017, https://www2.deloitte.com/content/dam/Deloitte/de/Documents/Innovation/Blockchain_A%20game-changer%20in%20accounting.pdf
[7] *Triple Entry Bookkeeping with Bitcoin, bitcoinmagazine*, 2014 https://bitcoinmagazine.com/articles/triple-entry-bookkeeping-bitcoin-1392069656/
[8] Lenuta, A., Sînică, A., Tudor B., *Extending swarm communication to unify choreography and long-lived processes*, At 23rd International Conference on Information Systems Development (ISD2014 Croatia), 2014.
[9] Sînică, A., Ioana, B., Lenuta ,A., Mircea-Florin, V., *Operations on encrypted data in an ORM made for encrypted choreographies*, ICCP (IEEE 13th International Conference on Intelligent Computer Communication and Processing), 2017
[10] PrivateSky, a research europen project: https://profs.info.uaic.ro/~ads/PrivateSky/project/
[11] Sînică, A., Doina, C., *Private Data System enabling Self-Sovereign Storage managed by Executable Choregraphies*, DAIS - 17th IFIP International Conference on Distributed Applications and Interoperable Systems, Neuchâtel, Switzerland,June, 2017
[12] Satoshi N., https://bitcoin.org/bitcoin.pdf
[13] Etherium web site, https://www.ethereum.org/