ACQUISITION INNOVATION RESEARCH CENTER

CLEARED
For Open Publication

Oct 27, 2022

Department of Defense
OFFICE OF PREPUBLICATION AND SECURITY REVIEW

October 17, 2022

FINAL TECHNICAL REPORT AIRC-2022-TR-009

WRT-1049.12c3

BLOCKCHAIN DATA STRUCTURES FOR TRACKING DOD BUDGET SPENDING

PRINCIPAL INVESTIGATOR: Dr. Yeganeh M. Hayeri, Stevens Institute of Technology
CO-PRINCIPAL INVESTIGATOR: Dr. Ian Taylor, Stevens Institute of Technology

Sponsors: Mr. Mark E. Krzysko, SES, Principal Deputy Director for Enterprise Information, Acquisition Data and Analytics, Office of the Under Secretary of Defense for Acquisition and Sustainment
Mr. James Pitcher, Department of the Air Force

Cleared for Public Release.

The views, findings, conclusions, and recommendations expressed in this material are solely those of the authors and do not necessarily reflect the views or positions of the United States Government (including the Department of Defense (DoD) and any government personnel), [or] the Stevens Institute of Technology.

Copyright © 2022 Stevens Institute of Technology. All rights reserved.

The Acquisition Innovation Research Center is a multi-university partnership led by the Stevens Institute of Technology and sponsored by the U.S. Department of Defense (DoD) through the Systems Engineering Research Center (SERC)—a DoD University-Affiliated Research Center (UARC).

This material is based upon work supported, in whole or in part, by the U.S. Department of Defense through the Office of the Under Secretary of Defense for Acquisition and Sustainment (OUSD(A&S)) and the Office of the Under Secretary of Defense for Research and Engineering (OUSD(R&E)) under Contract HQ0034-19-D-0003, TO#0309.

No Warranty. This Material is furnished on an “as-is” basis. The Stevens Institute of Technology makes no warranties of any kind—either expressed or implied—as to any matter, including (but not limited to) warranty of fitness for purpose or merchantability, exclusivity, or results obtained from use of the material. The Stevens Institute of Technology does not make any warranty of any kind with respect to freedom from patent, trademark, or copyright infringement.

Research Team

Name	Org.	Labor Category
Yeganeh Hayeri	Stevens	Principal Investigator
Ian Taylor	Stevens	Co-Principal Investigator
Hojat Behrooz	Stevens	Data Scientist
Mohammad Saghiri	Stevens	Post-Doctoral Researcher

Table of Contents

List of Figures

Executive Summary

This study aims to develop a blockchain data structure and a prototype to track budgeting as funds flow between the DoD elements. The objective of the project is to improve insights into transactions to enable acquisition funding flexibility and to better spend tracking information for DoD planning, management, and macro balancing. The team developed a blockchain data structure and prototype tool to track budget usage as funding flows DoD elements. The report summarizes a range of solutions and use cases that blockchain technology can enhance. Budget tracking has not been tackled by blockchain technology. The existing solutions in other domains only utilize the blockchain as a service. They do not consider how techniques such as graph analyses and machine learning may be used to improve a blockchain-based solution. An inappropriate design of blockchain may lead to a drastic increase in data redundancy of the system, which may result in the high cost of system maintenance in large-scale systems such as government infrastructures. In addition, because of the high volume of data in a blockchain-based system, we need tools, such as graph analyses combined with blockchain, to extract new information and patterns from the data. Therefore, a hybrid design based on blockchain and graph analyses has been utilized for budget tracking in DoD.

In this project, a prototype for organizing a blockchain data structure is proposed with the aid of graph analysis concepts. Since obtaining real-world data for DoD budget management proved difficult, we utilized a high-level data structure based on samples provided during this project as a surrogate for a more realistic dataset. Upon developing a preliminary roadmap, and a graph network demonstrating DoD fund transfers, seven novel use cases have been proposed to show the applicability of this solution. For each use case, we identified which problem has been targeted, how the proposed design can be applied, and how the suggested solution design can be beneficial. Use cases cover several issues related to budget management, data manipulation, and security, optimization, as well as issues related to cost and cost-effectiveness.

The results of this project highlight the followings:

- The immutability of financial data, a vital issue, can be supported by blockchain.
- Graph transaction to demonstrate DoD funds and extract valuable insights is a viable option. Since the existing infrastructure for analyzing DoD data is complex and becomes more challenging upon utilizing blockchain technology, the proposed ideas for analyzing transactions and budgets enhance and secure facilitating data analyses.
- The transaction graph can be used to improve oversight and accountability. In addition, the flows and relationships can be extracted through graphical analysis.
- The combination of blockchain and graph analysis brings transparency and accountability to budget tracking issues in DoD.

This project was conducted based on the following steps:

- Obtaining data and data structure from DoD
- Designing a graph model to highlight the patterns and trends and to obtain various inquiries.

- Designing a blockchain-based solution
- Designing and demonstrating the prototype
- Developing and illustrating several use cases

Since this project focuses on a novel concept and there is no prior design for the blockchain-based system for budget management, we focused on design issues, proposing a prototype to support the idea as a proof of concept. We adapted our design based on provided data samples. The suggested use cases have a high potential to be used as extensions of this project. They include:

- Protecting contract data from manipulation
- Defense industrial base health
- Geographical contract data query improvement
- Improving contingency contracting progress and contract data
- Intellectual property management
- Spare parts contracts
- Predicting missing and inconsistent acquisition data fields

Organization of this Report

This report is organized into multiple sections. The sections are as follows:

1) Introduction
2) Background Concepts
3) Related Projects
4) Motivation
5) Method
6) Blockchain Use Cases
7) Findings
8) Conclusion

Introduction

Recently, Blockchain technology has received much attention because of its rising applications in cryptocurrencies such as Bitcoin and Ethereum. These applications introduced beneficial capabilities such as traceability, immutability, and transparency [1]. In addition, this technology can build a trustable system among a large population of entities that may not be trustable. This feature is required for many applications in financial systems, healthcare, and military services [2].

In a blockchain-based system, a ledger is an immutable shared memory among a set of entities. The shared memory consists of a group of blocks created based on cryptography techniques. Adding every block to the system should be done by a consensus algorithm; therefore, most entities should accept the new block for adding to the ledger [3]. This approach can protect information on the ledger against unauthorized changes. In addition, blockchain technology provides infrastructure for a contract called "Smart Contracts." As contractual or business obligations are met, business rules can be automatically applied [4].

Blockchain technology has a high potential to change governments and industries. According to [5], the Defense Advanced Research Projects Agency (DARPA), the Department of Defense (DoD), and other governmental agencies have organized several research lines to utilize the benefits of blockchain on a government scale. In addition, many industries (i.e., supply chain, logistics, and healthcare) could be considered as potential application areas for blockchain technology. Unfortunately, the number of blockchain-based solutions in governmental systems, specifically in DoD, is low.

Budget management, a well-known domain where blockchain-based solutions can provide game-changing solutions in financial fields, has not yet been considered seriously in DoD. In this project, the blockchain and its applications to DoD are studied. Moreover, some potential use cases in budget management that could be considered in DoD are explored.

Background Concepts

This section introduces some basic concepts about blockchain and smart contracts.

2.1 Blockchain

The Blockchain concept started as a decentralized, immutable ledger framework for transactional information ordering [3]. This technology can revolutionize a wide range of domains, from finance to administration, by promoting security, reliability, and transparency, established via a decentralized and equitable computational model, with a high level of privacy guaranteed [1]. Due to its capabilities, blockchain technology has recently received much critical attention from the academic environment and industry. This technology was initially created to handle monetary transactions within the context of digital currency using a peer-to-peer network. Blockchain technology includes some applications in different fields [6].

- **Public Blockchains:** In this type, anyone can join and participate in the network. The network typically has a mechanism to encourage more participants to join the network. Bitcoin is one of the well-known examples of a public blockchain.
- **Consortium Blockchains:** In this type, a consortium blockchain is a distributed ledger where a preselected set of peers controls the consensus process.
- **Private Blockchains:** In this type, the write permissions in the system are kept centralized to one organization but read permissions may be public or restricted.

Hyperledger is an example of a private blockchain. In these systems, key participants are related to the organization which manages the blockchain.

A system may be created using a combination of the above blockchains. Figure 1 shows some blockchain applications in governments [7].

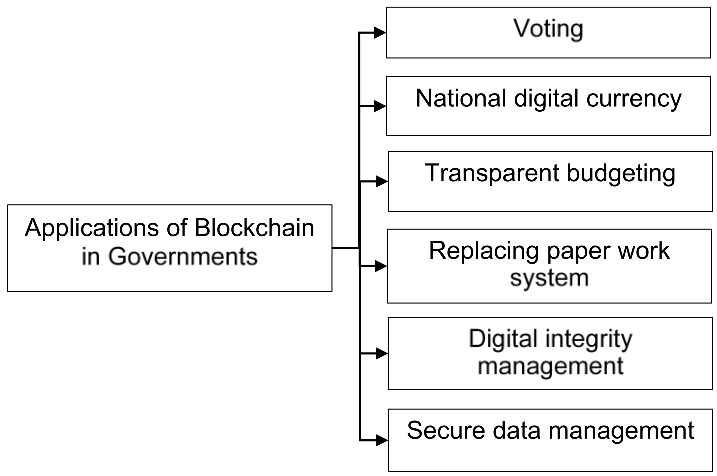

Figure 1 Applications of blockchain in governments

2.2 Smart Contracts

The term "smart contract" was first coined by Szabo (1996) as "*a set of promises, specified in digital form, including protocols within which the parties perform on these promises*." Smart contracts could be defined as the computer protocols that digitally facilitate, verify and exert the contracts created among parties on a Blockchain [8]. Smart contracts are ordinarily deployed and secured by a Blockchain. Firstly, the program code of a smart contract is recorded and confirmed on the blockchain. Secondly, a smart contract is executed among anonymous, trustless person nodes without centralized control or coordination of third-party specialists. Thirdly, a smart contract, like an intelligent agent, might have its cryptocurrencies or other digital assets and can transfer them when predefined conditions are triggered. According to [8], Figure 2 summarizes the benefits of smart contracts.

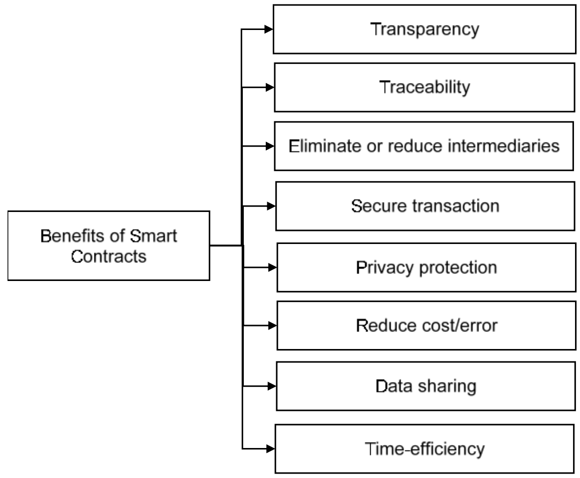

Figure 2 Benefits of smart contracts

There are various smart contract categories, and three popular ones will be presented hereafter. According to [9], the smart contract can be designated based on the following types:

- **On-chain contract:** This smart contract is executed within a chain code. The source code of the contract is defined based on business logic.
- **Off-chain contract:** in this type, business logic is used to design an explicit smart contract.
- **Hybrid contract:** A hybrid smart contract connects Blockchain (on-chain) code with off-chain data and computation.

Related Projects

In the 2018 National Defense Authorization Act (NDAA), the nation's top leaders recognized that blockchain technology may impact cybersecurity and ordered the Department of Defense to study it comprehensively [10]. According to [11], blockchain technology might be developed for the U.S. Air Force to enhance national defense. An application to the supply chain was described in [11] as a solution that could provide provenance information from "cradle to the cockpit." In this work, the application to the supply chain was shown to offer a solution to establish the provenance of every circuit board, processor, and software component. Based on [11], blockchain technology can improve security and efficiency by providing provenance in the DOD supply chain. The author's recommendations influenced the U.S. Air Force to develop organic government blockchain expertise through its Air Force Institute of Technology (AFIT) [12]. The AFIT's mission is to provide the defense community with information to make informed decisions. For supply chain management researchers at AFIT to help the defense community solve its most pressing supply chain tracking problems and solve supply chain security problems, they needed a simple way to study and demonstrate how blockchains work.

Since the original grant in 2017, DARPA has examined multiple blockchain-based systems, including anti-hacking blockchain strategies, blockchain within communications, and hardware network interactions [13]. DARPA has proceeded with caution in its implementation of blockchain technology due to significant issues that have yet to be resolved to ensure proper and efficient implementation [14].

In collaboration with SecureMarking, a supply chain security company, and the University of South Dakota Beacom School of Business, AFIT created a blockchain application based on a multi-echelon supply chain scenario [15]. This scenario involves an Air Force program manager issuing digital tokens. These tokens are then assigned to components. While the physical products move through the supply chain, tokens are transferred between companies based on the blockchain. Some companies in this scenario can add additional information to a token. For example, a component can be recorded on the blockchain when assembled into a product. And companies at any tier can see the end-to-end supply chain for all the parts they have handled. But only the Air Force program manager has complete visibility of all the parts from all the companies at any time.

According to the National Defense Authorization Act (NDAA) [16] of 2018, the DoD has six months to "provide to the appropriate committees of Congress a briefing on the cyber applications of blockchain technology. "This report included a description of potential blockchain-based offensive and defensive cyber applications, an evaluation of efforts to employ such technologies by foreign powers, extremist groups, and criminal networks, and an evaluation of the use or planned use of such technologies by the Federal Government for critical infrastructure networks, and an evaluation of the vulnerabilities of these networks to cyberattacks. In addition, in a recent DoD blockchain research study [17], the author examined a supply chain use case from the pharmaceutical industry to assess the viability of the strategy and its implications for cybersecurity when using blockchain for identifying fake products (drugs) and tracking. This research elaborates that integrating current blockchain technology into the DoD supply chain is not yet economically possible. The main reasons for this are challenges with data privacy, energy efficiency, and the lack of standardized blockchain interfaces among all participants.

According to [18], a governance-centered blockchain network implementation is suggested. The network enables the DoD to act as a regulator and protect all transactions' security and traceability. These recommendations included further advancements in blockchain development using proof of stake (PoS) within Ethereum or Hyperledger to address energy issues. Because DoD will decide who can participate and standardize all participant interfaces, this will alleviate data privacy concerns.

In [18], a white paper highlighted use cases of blockchain or distributed ledger technology (DLT) that benefit DoD. The paper's author argued that DoD would become even more dependent on digital technology and critically reliant on secure, timely, accurate, and trusted data due to the next generation of blockchain technologies, Artificial Intelligence, smart drones, robots, and additive manufacturing. However, as the value of data has increased, cyber warfare has emerged to pose a threat to the United States. Inappropriate actors and decisions may weaken American capabilities and disrupt crucial U.S. defense assets, ranging from communication systems to supply chains.

In [5], the authors identify a specific DoD use case, extrapolate requirements, and perform a thorough assessment of the different layers of the blockchain stack to identify the existing state of the art and undertake a gap analysis of the technology for this context. They describe a platform that meets many of these challenges and show how we architected, designed, and implemented a solution for this use case for deployment at Naval Air (NAVAIR) systems. This solution connects transactions from two separate blockchain systems, Consensus Quorum and Hyperledger Fabric, using a graph-based approach that preserves privacy while enabling full transparency across the military and supplier networks.

In [19], SIMBA Chain reported a $200,000 contract with the U.S. to utilize its blockchain system for securing sensitive data for Research and Development (R&D) sections. The DOD aims to improve this critical data's integration, security, auditability, and controlled access in this project. This project proposes an Authenticity Ledger for Auditable Military Enclaved Data Access (ALAMEDA) which will enable government agencies to securely share documents and scientific datasets across many relevant entities with the ability to control and track who gets access.

According to [12], the U.S. Air Force (USAF) has reported it will begin using SIMBA Chain in its supply chain management. A cloud-based Blockchain-as-a-service (BaaS) platform was introduced through a Defense Advanced Research Projects Agency (DARPA) grant. The platform, developed by Indiana Technology and Manufacturing Companies (ITAMCO) and the University of Notre Dame, has been awarded a government contract to explore how blockchain can enhance supply chain capabilities. The platform was selected following the call by the Department of Defense through the Small Business Innovation Research (SBIR) program for technologies to streamline supply chains. USAF's complex supply chain moves millions of parts across the globe, which requires ongoing, documentable precision. SIMBA Chain will be used to scale the additive manufacturing supply chain.

The authors of [18] tried to provide a wide range of case studies, including blockchain in DoD, some of which are explained as follows:

- Multi-Factor and Multi-Party Authentication
- Securing Spacecraft Tasking and Control
- Improving the Efficiency of Defense Logistics and Supply Chain Operations

Most use cases reported in [18] have not been implemented yet. On the other hand, the number of use cases in DoD has increased recently, although few implemented and practical solutions have been reported.

Motivation

Blockchains provide a secure way to transition assets without needing a central party/coordinator [20]. These systems are known by the following features [6].

- A decentralized consensus algorithm replacing the centralized control
- A system where a consensus is needed to agree to add (or remove) data; this results in a non-changeable medium.

Since no single point of failure exists and consensus algorithms can protect data against unauthorized changes, blockchain-based solutions have recently received much attention. Several use cases have been reported in the literature [7], two of which are explained below;

- For cryptocurrency, blockchain technology enables the movement and tracking of currency [2].
- For the supply chain, blockchain technology can track items (parts, produce, etc.) across multiple disparate organizations and systems, with each entity contributing but not controlling the information [21].

For tracking DoD budget/spending, we can use blockchain to pull data from disparate systems, create a non-reputable auditable record, and even extend this concept to active tokens providing movement of DoD funds. Compared with other projects conducted in DoD, such as those reported in [5], this project mainly focuses on budget management issues.

Method

The main objective of this project is tokenizing DOD spending to understand all the different layers and how funds flow between the various DOD systems. Therefore, the following steps have been taken to achieve this objective.

- Step 1- Understanding the existing contract vehicles and how funds currently move around.
- Step 2- Building the relationship between all the DOD elements and the data through a Graph model.
- Step 3- Create a smart-contract-based system that maps the relationships we have created in the Graph model so we can record them on a blockchain.

The rest of this section is organized as follows. Since data will play a vital role in this project, the first subsection is dedicated to data issues. The second subsection focuses on graph analyses and how we extract valuable information from data using graph analyses. The last subsection explains how the smart contract and blockchain should be organized.

5.1 Data Issues

Data is an essential part of the proposed approach. Because of security and privacy issues, utilizing recent information about organizations, programs, and DoD contracts was impossible. In addition, a fragment of data provided to the project's technical team was not well organized, and it had a lot of missing information. Since this project is a kind of proof of concept, the mentioned problems were not critical in the proposed approach. Nevertheless, an essential part of this project and its extensions relies on data and related issues, which should be addressed in practical solutions. Figure 3 shows a sample dataset used in this project, and Figure 4 shows the statistics of the dataset.

FPDS.award>re	FPDS.award>re	FPDS.award>	FPDS.award>	FPDS.award>	FPDS.award>	FPDS.award>
3/24/14 0:00	3/24/14 0:00	7884.14	7884.14	TRIDENT II M	UNITED STA	NOT A MANI
3/24/15 0:00	3/25/15 0:00	8718.48	8718.48	TRIDENT II M	UNITED STA	NOT A MANI
8/30/22 0:00	8/30/22 0:00	0	0	UH-60M BLA	UNITED STA	MFG IN U.S.
9/30/20 0:00	9/30/24 0:00	14390	14390	TRIDENT II M	UNITED STA	NOT A MANI
9/30/10 0:00	9/30/10 0:00	47440	47440	TRIDENT II M	UNITED STA	MFG IN U.S.
3/4/14 0:00	3/4/14 0:00	3361.25	3361.25	TRIDENT II M	UNITED STA	NOT A MANI
3/31/14 0:00	3/31/14 0:00	3742.25	3742.25	TRIDENT II M	UNITED STA	NOT A MANI
8/18/15 0:00	8/18/15 0:00	3823.52	3823.52	TRIDENT II M	UNITED STA	NOT A MANI
10/14/11 0:00	10/14/11 0:00	11155.5	11155.5	TRIDENT II M	UNITED STA	NOT A MANI
7/15/12 0:00	7/15/12 0:00	109365.76	109365.76	TRIDENT II M	UNITED STA	MFG IN U.S.
10/31/14 0:00	10/31/14 0:00	29447.54	29447.54	TRIDENT II M	UNITED STA	NOT A MANI
10/31/15 0:00	10/31/18 0:00	30066.39	130768.44	TRIDENT II M	UNITED STA	NOT A MANI
11/22/10 0:00	11/22/10 0:00	91647.96	91647.96	TRIDENT II M	UNITED STA	MFG IN U.S.
5/24/10 0:00	5/24/10 0:00	87476.17	87476.17	TRIDENT II M	UNITED STA	MFG IN U.S.
1/10/05 0:00		0	0	DDG 51		
10/31/09 0:00	10/31/09 0:00	9522	9522	TRIDENT II M	UNITED STA	MFG IN U.S.
4/6/10 0:00	4/6/10 0:00	90751	90751	TRIDENT II M	UNITED STA	MFG IN U.S.
12/20/10 0:00	12/20/10 0:00	29724.16	29724.16	TRIDENT II M	UNITED STA	MFG IN U.S.
9/30/11 0:00	9/30/11 0:00	9635.94	9635.94	TRIDENT II M	UNITED STA	MFG IN U.S.
1/12/13 0:00	9/30/13 0:00	8670.23	8670.23	TRIDENT II M	UNITED STA	NOT A MANI
12/30/09 0:00	12/30/09 0:00	14929.78	14929.78	TRIDENT II M	UNITED STA	MFG IN U.S.
12/30/10 0:00	12/30/10 0:00	32722.8	32722.8	TRIDENT II M	UNITED STA	MFG IN U.S.
3/4/11 0:00	3/4/11 0:00	18529.36	18529.36	TRIDENT II M	UNITED STA	NOT A MANI
5/29/20 0:00	5/29/20 0:00	4249478.94	4249478	DHMSM	UNITED STA	NOT A MANI
4/23/21 0:00	4/23/21 0:00	124250	124250	TRIDENT II M	UNITED STA	NOT A MANI

Figure 3 Sample dataset

Dataset statistics	Number of variables	100
	Number of observations	57,698
	Missing cells	1,677,333 (28.8%)
	Duplicate records	0 (0.0%)
Variable types	Numeric	13
	Categorical	81
	Boolean	4
	Unsupported	3

Figure 4 Dataset statistics

5.2 Graph Analyses

In this part, contracts and corresponding information can be represented as a graph. This approach opens new approaches for data analyses, as explained below:

- By analyzing the graph and its properties, new knowledge about the budget flow can be extracted.
- The graph may be visualized from different perspectives to extract more insights from the data and flows.

In what follows, a brief explanation of the activities done in this project phase is given. A sample of data from [22] has been used for network analysis. The data is included all the U.S. government spending categorized into the three levels of governmental organization and contractors. The datasets contained all allocated budgets for the specific governmental organization and their contracting details. The DoD has held 284 columns describing one specific budget for a particular program/project and its contractor. The main spending flow is presented in Figure 5.

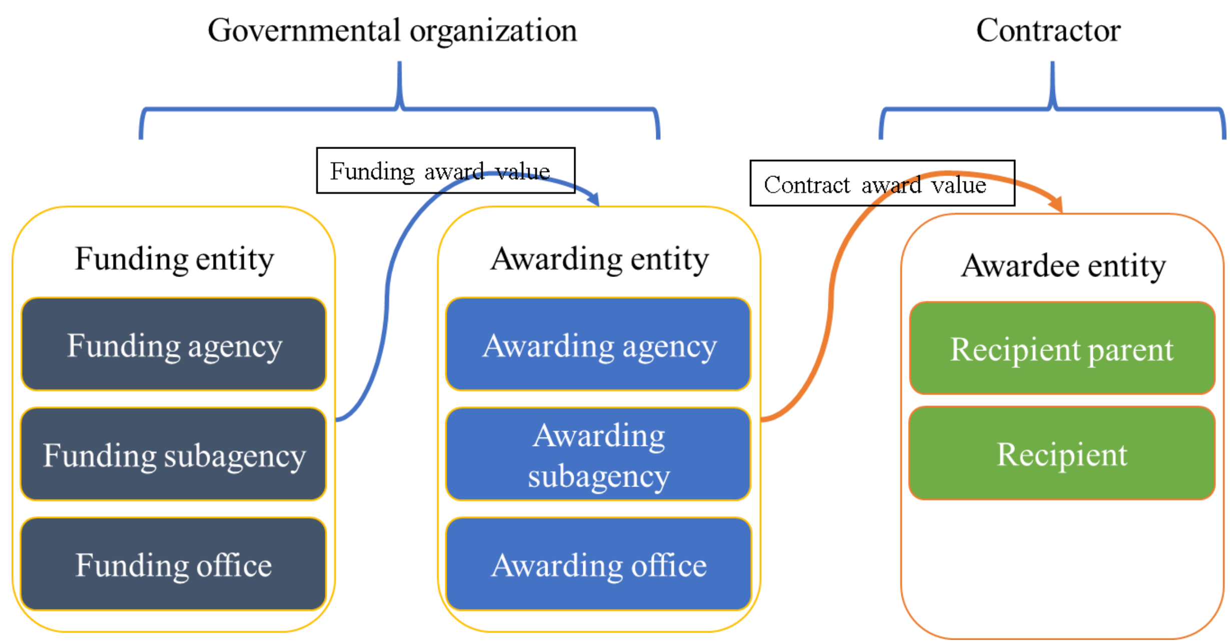

Figure 5 USA spending data structure format

The data are in CSV format and partitioned in 2-gigabyte slices for each government organization separately. They have included details about the origin of the budget and spending format and the details field about the geographical implementation of the contracts. It has a simple 3 level structure for the contracted and a two-level structure for a contractor. A graph has been created by applying the defined format and creating a graph presentation for each dataset record, as presented in Figure 6.

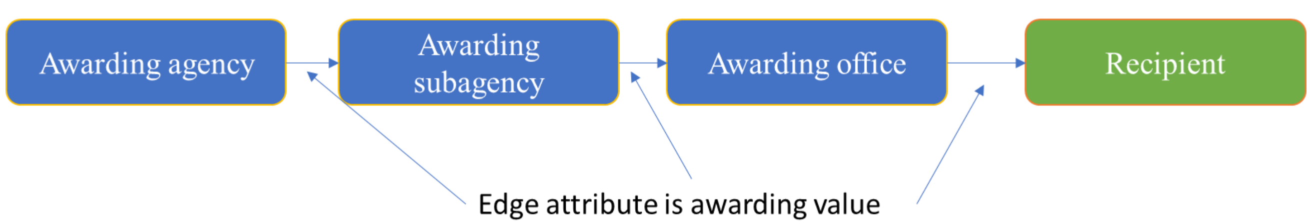

Figure 6 General graph format

A partial DOD dataset from 2020 for about 30,000 records has been retrieved and preprocessed to create the graph. The 30,000 records have been aggregated for each pair of the connected graph. The aggregation would be the total amount of the individual award values. 4,023 unique nodes and 5,294 edges have been created from the aggregation. The Gephi has been used for analyzing the graph. The first analysis has been done for clustering the graph based on the value of the contracting. The clustering has been done to implement the most weighted interconnected communities. The results are presented in Figure 7. the clustering separates the local communities that are most interconnected. It shows that some contractors make a contract with more than one community. Also, some individual contractors only work for different governmental organizations.

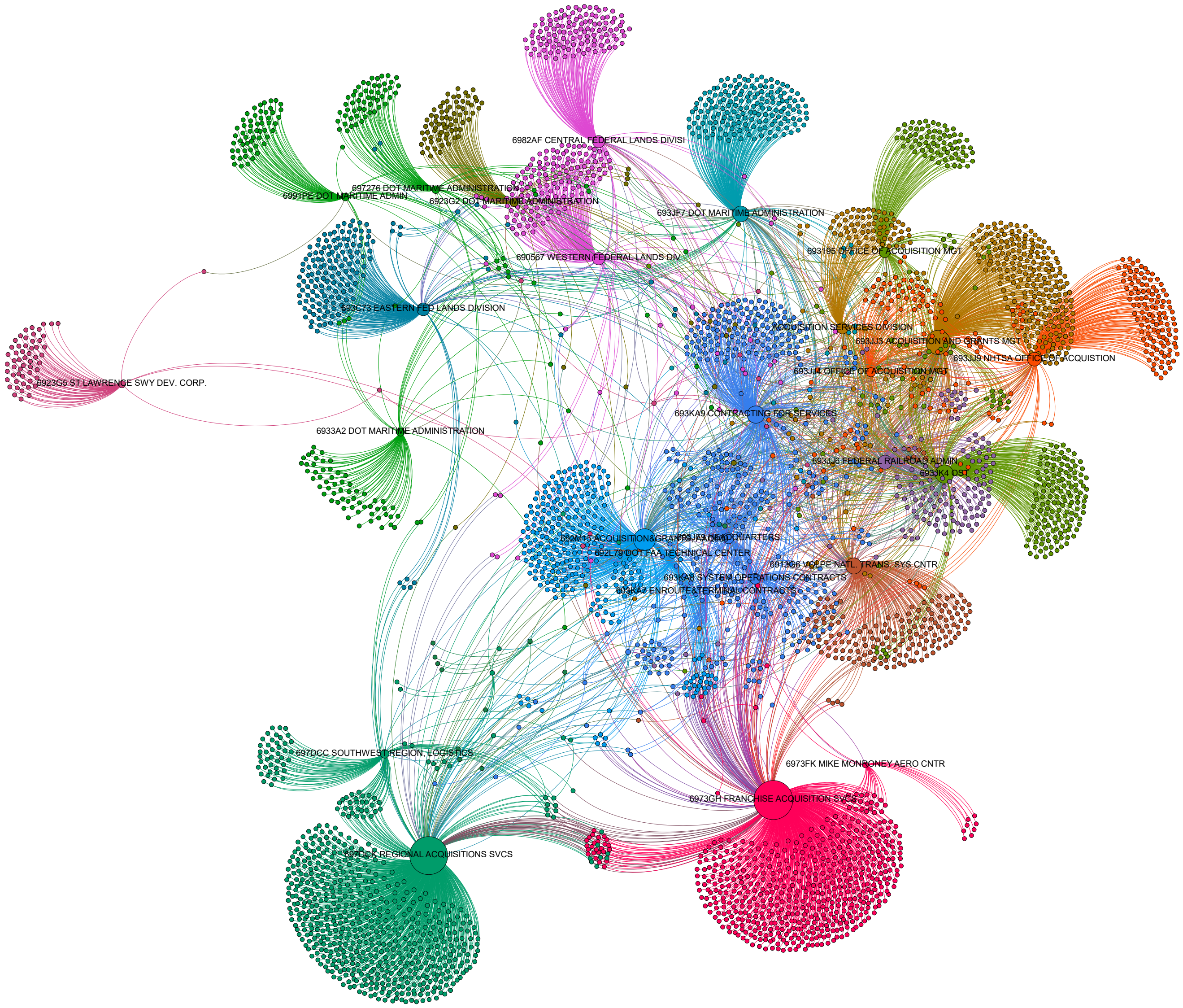

Figure 7 Clustering the contracts based on their weighted contract values

The spending budget within the governmental organizational structure is another part of the presentation shown in a Sankey exhibition in Figure 8. it shows that the organizational structure smoothly impacts the budget split. The budget flows from the top organization to the bottom without intercommunication or transfer between the sub-organizational entities.

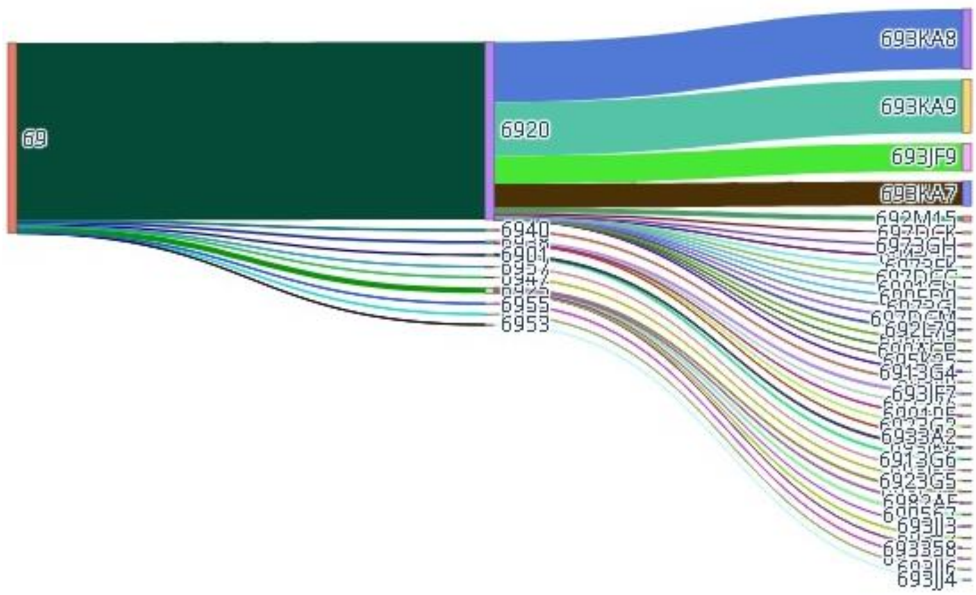

Figure 8 Budget flow in governmental organization

The last part of the analysis presents the clustering of the awarding graph by the graph degree. It clusters the graph based on the weighted frequency of the weighted relation of an entity. It clusters the contractor based on the amount and number of contracts.

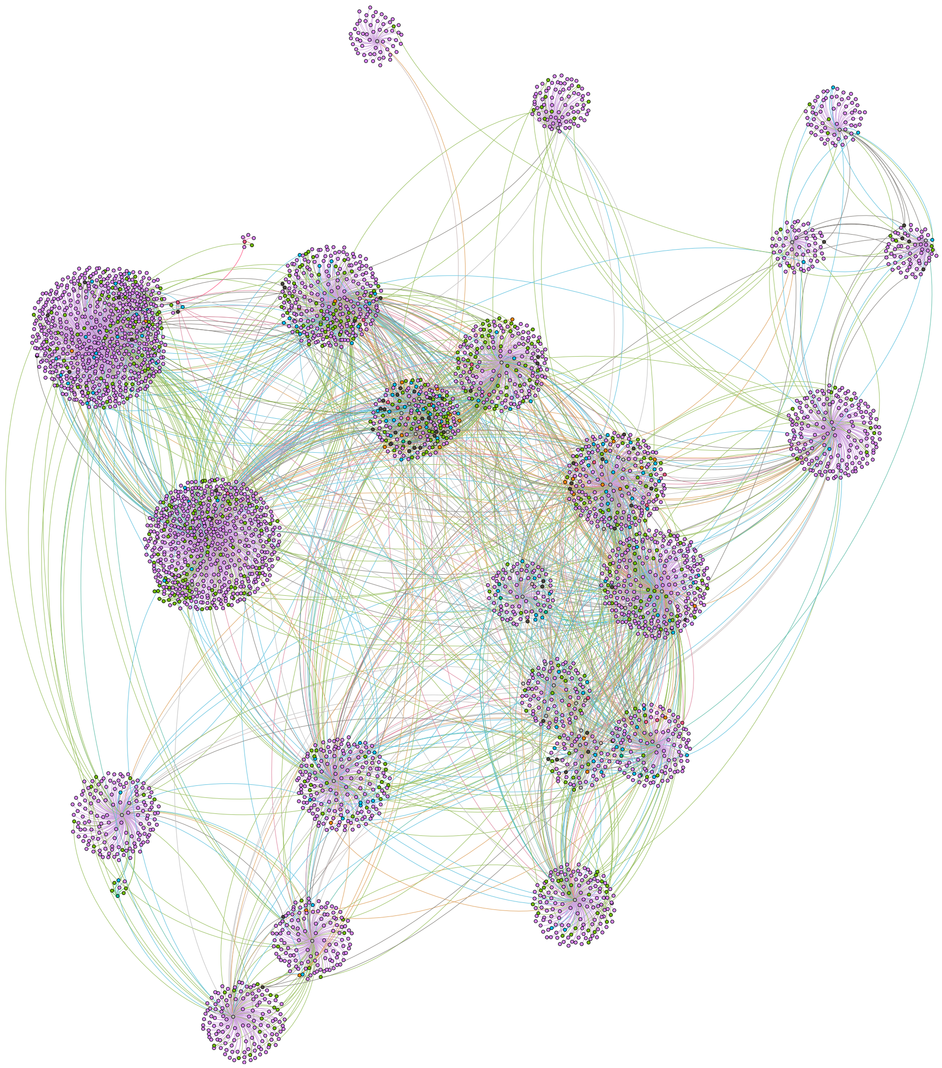

Figure 9 Graph clustering based on the weighted in-degree

Opportunity for analyzing the dataset graph: The dataset contains various categorical, quantitative, and qualitative attributes of each contract. For example, it specifies the geographical location of the contract implementation and the contractor. It could be used for examining the geographical distribution of the contract implementation or contractor.

5.3 Blockchain and Smart Contract Deployment

Blockchains provide a secure way to transition assets without needing a central party/coordinator. The first issue in these systems is how the blockchain and smart contracts should be organized. In this project, Figure 10 illustrates how we used the graph relationships to create smart contracts for representing and recording DoD program funding on a blockchain. In this figure, the nodes represent blockchain digital DoD program assets (programs, departments, and companies) and the edges represent transactions between entities. To implement the blockchain and smart contract, we used SIMBA chain [23], which provides the required technologies as several services. Therefore, many issues related to implementing a blockchain were moved to the Simba ecosystem.

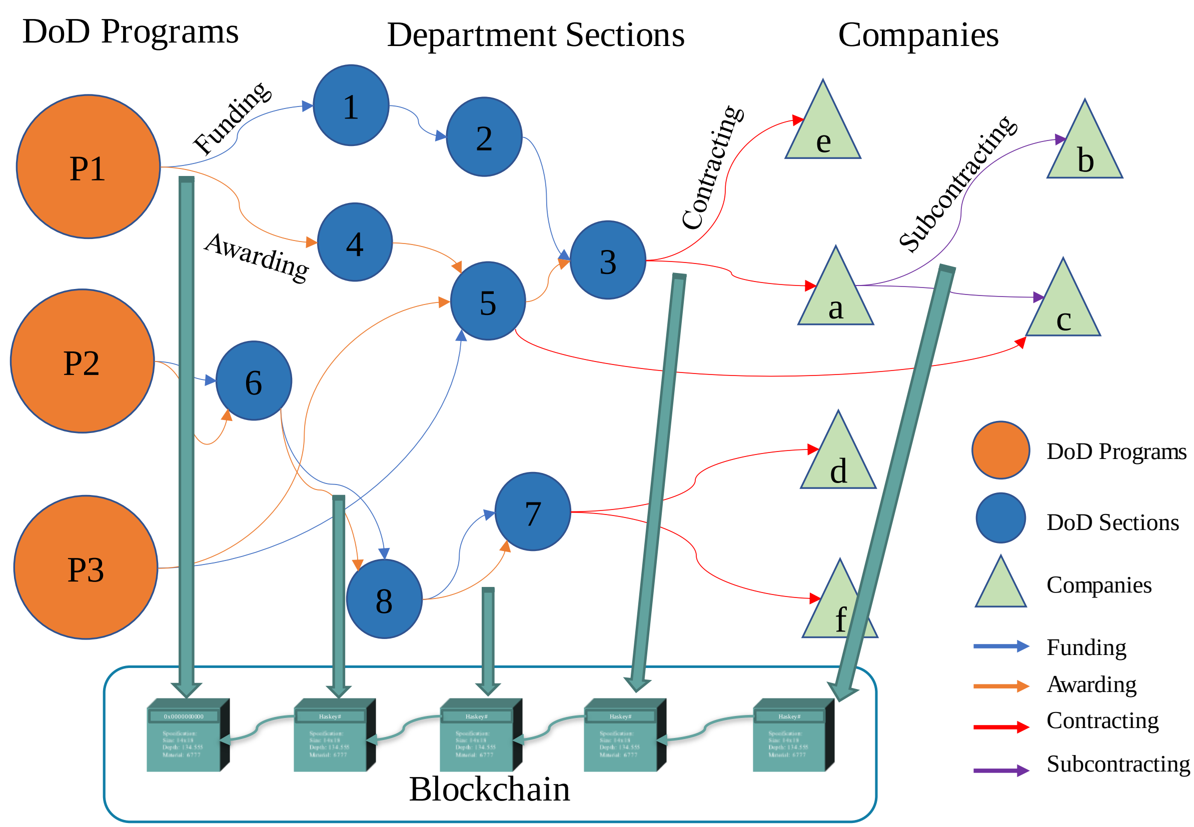

Figure 10 Graph relationships and blockchain

Figure 11 shows the Simba Chain interface for designing smart contracts. This interface has the following benefits.

- The SIMBA Smart Contract Designer allows a user to create Smart Contracts by graphically specifying the assets and the transition of those assets.
- This resulting graph is automatically converted to Solidity (Smart Contract language for Ethereum) for deployment onto a blockchain.

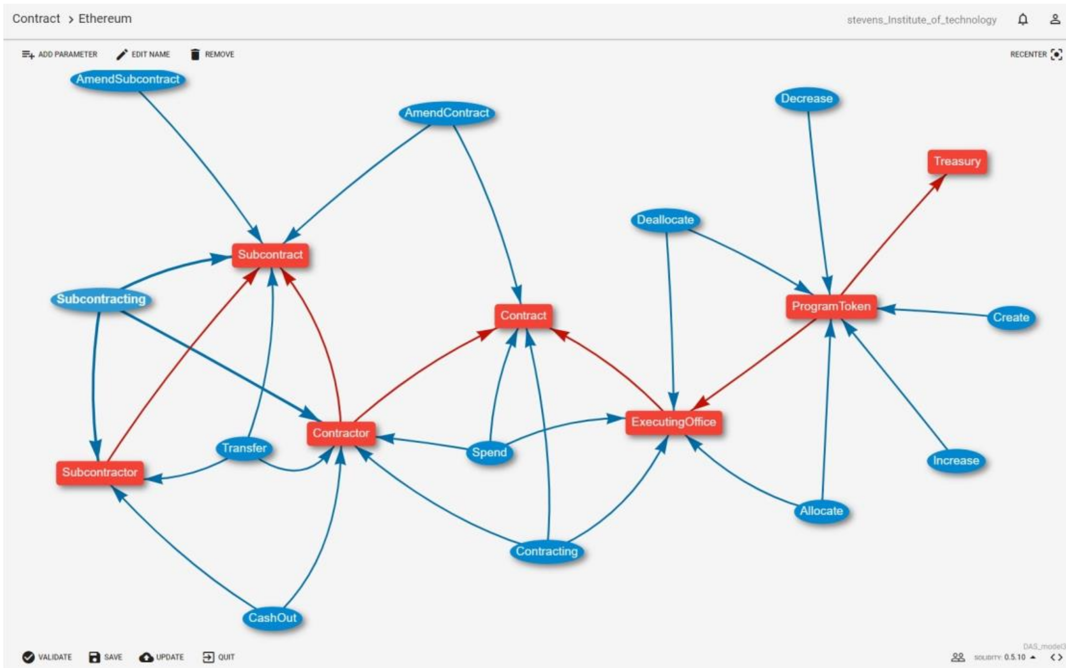

Figure 11 SIMBA Chain interface for designing a smart contract

Blockchain Use Cases

In this section, seven use cases are briefly explained. Evaluation, Implementation, testing, and the overall analysis of these use cases can be further explored in the next phase of this project.

6.1 Use Case One: Protect contract data from missing and manipulation

DoD relies on acquisition data for decision-making and managing the weapon-system acquisition portfolio. Therefore, improving acquisition performance is a high priority for the Secretary of Defense (OSD) office. DoD acquisition data are structured (e.g., relational databases) and unstructured (e.g., text). The data are collected for various purposes (e.g., statutory, regulatory, management, and decision-making) from centralized and decentralized locations. They have been collected relatively long, some dating back more than 50 years. Technology advancements have improved data collection performance, quality, accessibility, archiving, analysis, and reporting capabilities. Such improvements, however, have increased the risks of coming with externalities such as those associated with data security, redundancy, disconnection, and data integrity. Acquisition data are typically gathered from several sources resulting in suboptimality of quality and accuracy of input data, which could result from inconsistent updating policies among various elements of DoD. These challenges become crucial when paralleled with data sensitivity and security concerns [24].

The contracts and subcontracts datasets are part of the DoD acquisition system and contain financial records, including but not limited to the purpose of contracts, initial appropriation funds, the contracts period, contracts awardee office, and the spending amount. The contract financial records present the contract activities during its timeline. The related subcontracts are also part of the primary contract information record. However, some fields and financial transactions were missing over time.

Any improvement in the quality of acquisition data and contract financial transaction archiving could improve the DoD's policymakers' and managers' decision-making process and impact the efficiency of programs' budgeting by reducing planning costs and process time. The planning process is expedited by quickly accessing qualitative, accurate, and properly formatted data without requiring intermediate reporting and data improvement efforts.

Data missing and a blockchain system can address manipulation problems. A blockchain can manage data flow among contracts in a transparent and traceable fashion without any intermediary. Inserting the data and contract into a blockchain provides an accurate, traceable infrastructure for contracts to protect data integrity and immutability. Blockchain efficiently resolves data missing and manipulation in a shorter period and lower cost. We may need to design some application-specific consensus algorithms to protect the integrity of data inserted into the blockchain. The consensuses process can prevent data manipulations in a self-organized manner.

6.2 Use Case Two: Defense Industrial Base Health

Major unpredictable events disrupt the world's socioeconomic crucially. They affect differently, including the shortage of munitions and materials because of the high demand (the Ukraine war) or slowdown in raw material supply in production and delivery (COVID-19 pandemic) [25]. However, their impact is likewise reflected in cost surges and schedule delays. Those disruptive events also impact the DoD's entire acquisition accordingly. For example, the Ukraine war caused a massive drawdown of U.S. munitions and material stocks. The war shifted the supply of Russian-based material to US-based material while there was no replenishment plan. The demand growth impacts the costs and schedules directly. A similar scenario happened to DoD's acquisition of COVID-19's long-time impairment to a slowdown in raw material supply in production and delivery. Those disruptions reduce U.S. defense supremacy and increase defense supply chain vulnerability and risks.

Moreover, the U.S. defense industrial base health is targeted by 14017 presidential orders (America supply chains) to comprehensively review supply chains in critical sectors, including the defense industrial base (DIB) [26]. It intends to strengthen the national industrial base during times of disruption.

DoD acquisition system barely traces the supply chain products and information to their origins. The contractors only present their subcontractors' components and materials specifications and features. However, having complete supply chain information on all products helps risk assessments and mitigation contingency plans for supply shortages. Having all supply chain information for acquired products allows DoD to plan for an alternative supplier or assess the risk of regional disruption. The contractors should provide supply chain information. However, the information is untrustworthy as intermediate companies could manipulate and alter that information. That information also is sensitive data, and its security is crucial for the companies.

A blockchain-based solution for the supply chain could provide required information to the DoD acquisition system on a real-time basis to plan for alternatives and risk assessment. The supply chain information in the blockchain model is secure and trustable as they are produced and encrypted by the original supplier. The information on the components and raw materials is passed along the blockchain without interference from intermediate companies. The originality of every product and its movement can be traced to a high level of security with a blockchain-based token management system. Without the proposed solution, paperwork and redundant processes may appear in different parts of the supply chain, which imposes a higher cost.

6.3 Use Case Three: Geographical Contract Data Query Improvement

Recently federal acquisition regulatory council (FAR) established a new regulation [27] to increase content requirements for federal government procurements governed by the buy America act. According to presidential order 14005, the regulation profoundly tries to maximize the use of products, materials, and services supplied in the U.S. The change is intended to improve supply chain resiliency in crucial sectors such as defense, public health, agriculture, and transportation.

DoD fundamentally supplies its needs and requirements by contracting and subcontracting through private companies. The acquisition system contains various information about those contracts and subcontracts. The information described the contract terms, such as scope, schedule, and financial terms required by the department. Additionally, the contractors and purchased material, and the implementation location are also defined in the system. Geographical information describes the origin of the contractors, the source of production of supplied material and goods, and the project implementation place. That information explains where a purchased product is produced, where to install or implement it, where the contractor's origin is, and where the funds are. Although these data should be well established and collected in the acquisition system, there is missing, incomplete, and invalid location information. Those problems generally happen because several disconnected distributed systems are responsible for collecting information for each contract and subcontract. Typically, that location information is not validated. As a result, geographically tracking the supplied products and services is not accurate enough.

A blockchain data structure could help feed the location base information of the DoD contracts in an immutable, distributed manner. The location-based information would be provided to the system by participant individuals such as contractors, subcontractors, and the department agency as a contractee. A consensus algorithm will validate that information to protect the data's integrity, accuracy, and connectivity. Validated policymakers and managers could swiftly use accurate location-based data to increase the U.S. base supplement aligned with the regulations and laws without needing intermediate data preprocessing and reporting activities. As a result, it reduces the time and cost of those activities for DoD.

Additionally, it is possible to tokenize and manage the products by blockchain. The information about products' tokens can be imported into a shareable file within the blockchain. Then the blockchain provides a fast and secure verification process to check the product's originality. As a result, we will be able to organize complex queries for a prolonged product supply chain. Without such a system, verification for a complex or non-complex product supply chain needs longer, more paperwork, and a higher cost.

6.4 Use Case Four: Improve Contingency Contracting Progress and Contract Data

DoD relies on contractors worldwide to support contingency operations such as armed conflict and humanitarian crises. DoD's information system—synchronized pre-deployment and operational tracker-enterprise suite (SPOT-ES) [4]—traces and reports information about contracts and contractor personnel supporting applicable contingency operations. However, SPOT-ES could not present data by the type of relevant contingency operations [28]. This obstacle happens because of a lack of association between operations and practical contingency operations in SPOT-ES. As a result, information on many quarterly deployment records on contractor personnel supporting applicable contingency operations misses in SPOT-ES. Without clarifying the responsibility for resolving missing or inaccurate data in SPOT-ES within DOD guidance, the reliability of data in SPOT-ES is at risk. The root cause analysis shows that two main issues cause the problem:

- Operational types and names change over time, even within the same area of operations
- Data is considered sensitive due to operational security concerns.

As a result, DOD's ability will be hindered when there is a need to find the contractor staff's location during an emergency or when contractors exit at a contingency location.

Therefore, improving the efficiency and performance of the DOD's capability to locate the contractor workforce's location during an emergency or when contractors exit at a contingency location declines overall operational costs and delays.

A solution to the problem could be building an integrated, decentralized, secure, and unmanipulable SPOT-ES to trace and report information about contracts and contractor personnel supporting applicable contingency operations. Apply blockchain and smart contracts to create an immutable database to ensure that data on operations, exercises, and other activities in SPOT-ES are linked with "applicable contingency operations" securely and decentralized. A blockchain-based solution, consisting of several self-executable contracts, can check the status of data and make the appropriate decisions (suggestions) in a fully self-organized manner without any intermediator. The solution also can reduce outside attack risk. In addition, Blockchain technology prevents the data from being missing and disconnected or unintegrated. Additionally, blockchain encryption protects a secure repository for individual system users. Finally, the blockchain data structure provides an accessible data flow for tracking the information for planning for future wartime or crises based on the contractor's location and capabilities.

6.5 Use Case Five: Intellectual Property Management

Existing DoD acquisition processes and regulations regarding Intellectual Property (IP) are not appropriate for utilizing the potential of private industry partners such as non-traditional companies [29]. Some of the issues are explained as follows. The size of information regarding IP is increasing, and there is no scalable system that considers vital abilities such as security and privacy in data management. In addition, different pricing for licensing agreements, disconnection between related sectors, and budgeting, increase the complexity of IP

Management There are repositories such as SAM.GOV and a project reported in [29] to explore a considerable amount of data related to IP However, most problems, such as security, traceability, consistency, and transparency, have not been appropriately addressed.

Blockchain can capture IP outputs, register them across the DoD without a centralized system, and associate them with awards. This solution eliminates useless intermediaries. In addition, the information about IPs will be protected against unauthorized changes by consensus algorithms.

In addition, using blockchain, we can control how the IP can be used, under what conditions the IP can be used, and the contractual obligations for using that IP The smart contract can be used to manage the flow of data regarding IP Applying a smart contract will result in more efficient exploration and re-using of IP across the DoD, remove duplicate efforts, and fully automate licenses and other IP requirements. A blockchain-based solution can decrease the management cost and reduce the duration required to execute contracts in the system.

6.6 Use Case Six: Spare Parts Contracts

DoD spends billions of dollars yearly on spare parts for maintaining planes, ships, and equipment. The defense logistics agency (DLA) is the primary supplier of those spare parts. The spare parts are generally acquired through contractors. The Federal Acquisition Regulation (FAR) obliges DLA contracting officers to control if contractors proposed prices are fair and reasonable before awarding contracts. However, for non-competitive contracts such as sole-source contracts, contracting officers must obtain certified cost or pricing data (information contractors certify as accurate, current, and complete) to determine whether the offered prices are fair and reasonable. Two significant issues happen in these specific situations.

First, the contractors delay giving certified cost or pricing data. According to a study by GAO, this delay could cause a long delay in contracting and supplying required spare parts [30]. For instance, the Airforce reports 1,154 days of delays in supplying Aircraft auto-flight parts between 2015-2019.

Second, overcharged prices could happen for specific solo contractors' spare parts suppliers. For example, TransDigm, one of the Pentagon's largest spare-parts suppliers, collected excess profits on 105 parts from early 2017 to mid-2019, according to a two-year inspector general's inquiry into the company's business model [31]. TransDigm is a key sole-source provider of spare parts for Pentagon airplanes and helicopters, including F-16 jets and AH-64 Apache, and CH-47 helicopters. DoD executed 4,697 contracts with TransDigm and its subsidiaries, valued at $634.7 million between 2014 and 2019.

Recently House proposed legislation to force government vendors to disclose their cost data to help the government negotiate for spare parts, which Pentagon officials support. The legislators believe access to cost data would enable contracting officers to negotiate fair and reasonable prices.

Although the idea of passing legislation to force governmental vendors to disclose their cost data is an essential step toward sole-contracting price transparency, the tractability of those data is unclear. The disclosed costs could be manipulated and altered by the vendors, and verifying their validity is far more complicated.

A blockchain-based solution for the supply chain information and the fund could provide required information to the contracting officers on a real-time basis to estimate the fair and reasonable spare part price. The supply chain funds and information in the blockchain model are secure and trustable as they are produced and encrypted by the original supplier. The information and cost of the components and raw materials are passed along the blockchain without interference from intermediate companies. The final cost of every product based on its components and raw material costs can be estimated with a high level of security with a blockchain-based token management system. Blockchain protects the data against manipulation and alteration by applying a consensus algorithm.

The House targets the spare parts supply chain information and fund data for the Governmental suppliers. Those data help the DoD to reduce the delay in estimating the fair price and also give a better negotiation tool to the contractors' officers to reduce the cost of spare parts by cutting unfair contractors' profits.

6.7 Use Case Seven: Predicting Missed and Inconsistent Acquisition Data Fields

DoD ultimately relays on acquisition data for decision-making and managing the weapon-system acquisition portfolio. Therefore, improving acquisition performance is a high priority for the secretary of defense (OSD) office. DoD's acquisition data are in structured (e.g., relational databases) and unstructured data formats (e.g., text). Such data are collected for various purposes (e.g., statutory, regulatory, management, and decision-making) from centralized and decentralized locations. The data has been collected for a relatively long (some dating back more than 50 years). Technological improvement has helped DoD improve data collection performance, data quality, data accessibility, data archiving, and analyzing and reporting capability. However, technological advancements have setbacks in data integrity, redundancy, disconnection, and security. Acquisition data are sourced through several information systems supporting various functional business areas. The problems arise from two main issues, including 1) quality and accuracy of input data and 2) different updating process policies. Those challenges are more crucial when paralleled with data sensitivity and security concerns [24].

Any improvement in the quality of acquisition data could improve the DoD's policymakers' and managers' decision-making process and impact the efficiency of programs' budgeting by reducing planning costs and process time. The planning process is expedited by swiftly accessing qualitative, accurate, and properly formatted data without requiring intermediate reporting and data improvement efforts. The improvement should be done by defining a functional data governance framework to monitor, control, and manage the acquisition systems (tools) and the acquisition data. However, those processes only could improve the data quality and accuracy for the future and cannot enhance oversights in currently collected data.

Machine learning (ML) approaches are data-driven methods that can learn from historical data and predict missing and inconsistent data. The methods fundamentally look at the historical data, find the hidden internal relation between them, and record those relations as knowledge (motifs). Then, any missing data fields or the inconsistent transaction would be predicted by comparing them to the learned knowledge and trying to find the best fit. This approach proves its capabilities in various application areas, such as next-word prediction in text editors.

The graph format of the relation between the contracts, contractor, contracted, subcontract, and subcontractor adds another layer of information to the acquisition system data. Those graph relations also could be used for graph analysis of the data for predicting the missing links such as missing contracts or specific contract missed fields such as subject. Mixing graph theories with ML strategies give another opportunity to improve the acquisition of data quality and accuracy by applying methods such as graph neural networks.

Findings

This study aimed to develop a blockchain data structure based on the graph analysis concept to solve some financial problems in DoD. As a result of this study, we may point out the following findings:

- Graph analyses will be vital to organize and also analyze blockchain data structures. With appropriate analyses, we may reduce the data redundancy of blockchain systems. In addition, we can extract valuable knowledge about the changes in the structure and data of blockchain-based systems.

- Since the changes in blockchain-based structure occur very fast, graph analyses and machine learning systems will be required to manage financial systems based on blockchain. If continued with the status quo, the reaction time of a human agent for decision-making and handling events such as hacking and failure will not be acceptable.

- The complexity and cost of financial systems are increasing because of the management mechanisms and intermediaries. The hybrid systems based on blockchain and graph analysis techniques will reduce the cost and response time of systems because of removing intermediaries and also many useless human agents and processes.

- Blockchain-based systems support a high level of traceability for transactions. In addition to tracking savings and alternative consumptions, we can move from reactive audit and data insights to a proactive tokenization mechanism that can distribute funds while enabling transparent spending and accountability.

Conclusion

This project reviewed applications of blockchain technology and graph analyses tackling multiple challenges of DoD. Then, a method for organizing a blockchain system based on graph analyses for budget management problems has been suggested. We also discussed the potential of hybrid solutions based on blockchain and graph analyses which will have many applications in DoD. It should be noted that all of the reported blockchain based solutions that are suggested in the literature for DoD only focus on using pure blockchain-based systems. These solutions lead to high redundancy and also complexity in the system. Therefore, the solution proposed in this project will be useful to mitigate the mentioned problems with the aid of graph analyses. Seven use cases have been suggested that focus on different scenarios related to budget management problems. Suggested use cases improve DoD's budget management from various aspects, such as transparency, traceability, and immutability. In addition, because of reducing intermediaries and online verification capabilities, blockchain-based solutions could significantly reduce the time and cost of services.

Project Timeline & Transition Plan

What is the long-term transition goal for the research if continued?

The proposed study can be extended to design a governmental infrastructure for managing the budgets of DoD during the digital transformation era. Existing mechanisms are not efficient, and this issue becomes more challenging after rising of other blockchain-based ecosystems for digital money and digital assets. Furthermore, a novel machine learning-based method can reinforce the proposed design to provide better budget management.

List the potential tools, guides, educational units, or other artifacts that resulted from this research that external sponsors might use if the long-term transition goals are met.

Students were trained as part of an applied blockchain course to learn how scenarios are developed using the tool to address various problems. The combined usage of Graph network and blockchain was tested and proved efficient in the visualization and analysis of the technology. The team will move forward to develop an open-source platform for the implementation of blockchain applications.

Which of these might be or are planned to be incrementally delivered in a future research task?

In addition to the open-source software mentioned above, use-case scenarios can potentially be developed further

Did you identify any transition partners? Are there other advocates or potential adopters of this research?

The research team worked with SIMBA Chain team because we used blockchain as a service. Collaborative thoughts and ideas were exchanged to find mutual paths between this project and other DoD projects SIMBA Chain has taken on.

Was the research piloted with a potential transition partner? Are there others who would conduct pilot research use if fully funded?

We collaborated with the SIMBA Chain team, which provided us with blockchain as a service. The Stevens team can collaborate with blockchain service providers such as SIMBA Chain to implement the scenarios for testing purposes if fully funded.

Appendix A: Cited and Related References

[1] J. Bambara, P. Allen, K. Iyer, S. Lederer, R. Madsen, and M. Wuehler, *Blockchain: A practical guide to developing business, law, and technology solutions*. McGraw Hill Education, New York, 2018.

[2] F. Tschorsch and B. Scheuermann, "Bitcoin and Beyond: A Technical Survey on Decentralized Digital Currencies," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 3, pp. 2084–2123, 2016.

[3] D. Mingxiao, M. Xiaofeng, Z. Zhe, W. Xiangwei, and C. Qijun, "A review on consensus algorithm of blockchain," in *2017 IEEE international conference on systems, man, and cybernetics (SMC)*, 2017, pp. 2567–2572.

[4] F. Dietrich, D. Palm, and L. Louw, "Smart contract based framework to increase transparency of manufacturing networks," *Procedia CIRP*, vol. 91, pp. 278–283, 2020.

[5] F. Asuncion *et al.*, "Connecting supplier and DoD blockchains for transparent part tracking," *Blockchain: Research and Applications*, vol. 2, no. 3, p. 100017, 2021.

[6] A. M. Saghiri, "Blockchain architecture," in *Advanced applications of blockchain technology*, Springer, 2020, pp. 161–176.

[7] D. D. F. Maesa and P. Mori, "Blockchain 3.0 applications survey," *Journal of Parallel and Distributed Computing*, vol. 138, pp. 99–114, 2020.

[8] S. N. Khan, F. Loukil, C. Ghedira-Guegan, E. Benkhelifa, and A. Bani-Hani, "Blockchain smart contracts: Applications, challenges, and future trends," *Peer-to-peer Networking and Applications*, vol. 14, no. 5, pp. 2901–2925, 2021.

[9] C. Li, B. Palanisamy, and R. Xu, "Scalable and privacy-preserving design of on/off-chain smart contracts," in *2019 IEEE 35th International Conference on Data Engineering Workshops (ICDEW)*, 2019, pp. 7–12.

[10] M. Thornberry, "National Defense Authorization Act for Fiscal Year 2018, H.R.2810, 115th Congress Public Law 91," Dec. 12, 2017. http://www.congress.gov/ (accessed Jul. 02, 2022).

[11] N. B. Barnas, "Blockchains in National Defense: Trustworthy Systems in a Trustless World," Air University (U.S.). Air War College, Jun. 2016. Accessed: Jun. 30, 2022. [Online]. Available: https://www.hsdl.org/?abstract&did=825348

[12] "USAF to use blockchain-based SIMBA Chain for its supply chain operations," May 17, 2020. https://supplychaindigital.com/technology/usaf-use-blockchain-based-simba-chain-its-supply-chain-operations (accessed Jul. 01, 2022).

[13] "DARPA Blockchain Programs." https://coincentral.com/darpa-blockchain-programs/ (accessed Jul. 02, 2022).

[14]"DARPA wants to take a look at blockchain's security rules - FedScoop." https://www.fedscoop.com/darpa-wants-take-look-blockchains-security-rules/ (accessed Jul. 02, 2022).

[15]"Air Force Institute of Technology releases free 'Blockchain for Supply Chain' Tools." https://www.mmh.com/article/air_force_institute_of_technology_releases_free_blockchain_for_supply_chain (accessed Jul. 02, 2022).

[16]"The 2018 National Defense Authorization Act .," Jun. 30, 2022. https://docs.house.gov/billsthisweek/20171113/HRPT-115-HR2810.pdf/

[17]D. Woodfield, "The Emerging Impacts of Blockchain Technology on DOD Asset Cyber Security," AIR COMMAND AND STAFF COLL MAXWELL AFB AL MAXWELL AFB United States, 2019.

[18]V. Adams. M, Alonso, W. Henry, D. Hyland-Wood, W. C. Jansen, V. Kodumudi, A. Nannra, J. Neidig, N. S. Matt Nelson, J. Stevens, and W. H. S. Tennenbaum, I. Taylor, J., "POTENTIAL USES OF BLOCKCHAIN BY THE U.S. DEPARTMENT OF DEFENSE." [Online]. Available: https://www.crowell.com/files/Potential-Uses-of-Blockchain-Technology-In-DoD.pdf

[19]"U.S. Department of Defense to use blockchain to secure sensitive data," *Ledger Insights - blockchain for enterprise*, May 13, 2020. https://www.ledgerinsights.com/us-department-of-defense-blockchain-secure-sensitive-data/ (accessed Jun. 30, 2022).

[20]C. Aristidou and E. Marcou, "Blockchain standards and government applications," *Journal of ICT Standardization*, pp. 287–312, 2019.

[21]S. Wang, D. Li, Y. Zhang, and J. Chen, "Smart contract-based product traceability system in the supply chain scenario," *IEEE Access*, vol. 7, pp. 115122–115133, 2019.

[22]"Government Spending Open Data | USAspending." https://www.usaspending.gov/ (accessed May 24, 2022).

[23]"SIMBA Chain," *SIMBA Chain*. https://simbachain.com/ (accessed May 24, 2022).

[24]M. McKernan, N. Y. Moore, K. Connor, and M. E. Chenoweth, "Issues with access to acquisition data and information in the Department of Defense: Doing data right in weapon system acquisition," RAND Corporation Santa Monica United States, 2017.

[25]"Defense Industrial Base Faces Short and Long Term Challenges," *Federal News Network*, May 03, 2022. https://federalnewsnetwork.com/contracting/2022/05/defense-industrial-base-faces-short-and-long-term-challenges/

[26]"Executive Order 14017 on America's Supply Chains," *FDA*, May 2022, Accessed: Oct. 06, 2022. [Online]. Available: https://www.fda.gov/about-fda/reports/executive-order-14017-americas-supply-chains

[27]"Federal Acquisition Regulation: Amendments to the FAR Buy American Act Requirements," *Federal Register*, Mar. 07, 2022. https://www.federalregister.gov/documents/2022/03/07/2022-04173/federal-acquisition-regulation-amendments-to-the-far-buy-american-act-requirements (accessed Jul. 05, 2022).

[28]"GAO: DOD Should Better Document Contingency Contracting Progress and Improve Contract Data - H.S.," 5 Oct. https://www.hstoday.us/federal-pages/government-reports-and-summaries/gao-dod-should-better-document-contingency-contracting-progress-and-improve-contract-data

[29]"I.P. Data Repository," *The Acquisition Innovation Research Center*. https://acqirc.org/publications/executive-summary-ip-data-repository/ (accessed Jun. 29, 2022).

[30]"U. S. Government Accountability. Spare Parts Contracts: Collecting Additional Information Could Help DOD Address Delays in Obtaining Cost or Pricing Data," 30 June. https://www.gao.gov/products/gao-21-388 (accessed Oct. 06, 2022).

[31]"Pentagon Contractor Accused of Overcharges Is Pursued by Lawmakers," *Bloomberg.com*, 18 Apr. Accessed: Oct. 06, 2022. [Online]. Available: https://www.bloomberg.com/javelin/com/article/basic/published/2022-04-18/transdigm-s-alleged-pentagon-overcharges-pursued-by-lawmakers.