Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

BAT: Block Analytics Tool Integrated with Blockchain Based IoT Platform

CHATHURANGI EDUSSURIYA[1], KASUN VITHANAGE[1], NAMILA BANDARA[1], JANAKA ALAWATUGODA[1], MANJULA SANDIRIGAMA[1], UPUL JAYASINGHE[1] , AND GYU MYOUNG LEE.[2]

[1] Department. of Computer Engineering, University of Peradeniya, Peradeniya, LK.
{edussuriya.c,kas.vith,namilad,alawatugoda,manjula.sandirigama,upuljm}@eng.pdn.ac.lk
[2]Department of Computer Science, Liverpool John Moores University, Liverpool, UK. g.m.lee@ljmu.ac.uk

Corresponding author: Gyu Myoung Lee (g.m.lee@ljmu.ac.uk).

This work was supported by the Institute for Information Communications Technology Promotion (IITP) grant funded by the Korea government (MSIT). [2018-0-00261, GDPR Compliant Personally Identifiable Information Management Technology for IoT Environment].

ABSTRACT The Internet of Things (IoT) industry is revolutionizing the physical world into a new digital era. However, the exponential growth of the number of IoT devices has raised number of issues to be addressed. Even though, the huge amount of data generated by these IoT systems can be used to derive useful insights and predictions, the authenticity and integrity of the data cannot be specifically verified. In this study, we provide solutions to scalability, security and privacy issues of IoT with the use of blockchain technology with its decentralized, immutable, secured architecture by introducing a blockchain based IoT platform. Moreover, novel approach called Block Analytics Tool (BAT) integrated with the platform is introduced. It provides an environment for the data produced by the IoT devices and systems to be analyzed and learned to make predictions about the future while ensuring the authenticity and integrity of data with the use of the tamper proof architecture of blockchain technology. BAT enables the Machine Learning and Artificial Intelligence applications to be built using the data stored in the platform in an optimized manner with increased efficiency. In this paper, we design and implement the proposed system. In addition to that, a pharmaceutical supply chain is used as the use case scenario to show the functionality of the proposed system. A model to forecast the demand of pharmaceutical drugs is built using a real-world data set to demonstrate the use of BAT. Furthermore, the performance of the BAT with the platform is evaluated using the case study.

INDEX TERMS IoT, blockchain, data analytics, smart contracts, Access management

I. INTRODUCTION

INTERNET of Things (IoT) plays a significant role in the convenience of human daily life at present through various innovative applications and services. Further, it empowers the concept of autonomous systems creating new social paradigm. The enormous amount of data generated by these services and systems usually are stored in on premises servers and cloud servers depending on the context. However, these types of systems are vulnerable to several issues, including single point of failure due to its centralized architecture. Gartner [1] has predicted that there will be 21 billion IoT devices at the end of 2020. The traditional client-server architecture of the current storage system will not be able to withstand the growing large number of IoT devices. These systems communicate and store critical, secure and privacy sensitive data such as medical records, financial records [2]. Moreover, this recorded data could be used to identify patterns and anomalies as well as make decisions about the future. Unauthorized access to these data could lead to many security and privacy concerned issues. There are many incidents which have recorded violation of the privacy of the users of IoT devices [3]. The users have little or no control over the access management to the data. Furthermore, the data could be modified, deleted from the systems without the concern of the users which has raised issues such as counterfeit of medicine in pharmaceutical supply chains. Owing to the data protection issues, GDPR (General Data Protection Regulations) [4] came into effect in the EU region trying to

protect the data of the citizens with rights and regulations. Nevertheless, more and more data breach incidents specially related to pharmaceutical supply chains are reported every day.

This paper presents a platform which makes use of the blockchain technology with other state of the art technologies in solving the above identified problems. The decentralized, peer to peer architecture of blockchain can address problems related to centralized architectures. Access control mechanisms are used to protect data and prevent unauthorized access. The tamper proof mechanisms of blockchain provides data integrity making the system resistance to unapproved modifications. Smart contracts are used to invoke the communication between IoT devices verifying the authenticity of data sources.

This study introduces a novel approach to build machine learning, Artificial Intelligence (AI) applications on the data stored in the blockchain with the Block Analytics Tool (BAT). In BAT, the storage system of blockchain which is designed for transaction processing, is combined with data analysis tools in an optimized manner without decreasing the performance. An index system which is specifically designed for transactions, called Block index is introduced to query the blockchain while reducing the cost associated with data retrieval. BAT enables decision making and predicting the future on trusted, secured data which are stored in a decentralized manner without an involvement of third parties. The platform with the BAT, facilitates learning on huge amount of data produced in the systems integrated with IoT devices.

We implement a real-life case study to show the functionality and we evaluate the performance of the proposed system of platform with BAT. Drug counterfeiting is a critical issue in pharmaceutical supply chains as it impacts human life. Using the IoT devices such as RFID tags and temperature sensors connected to the platform, the necessary transactions and functions of the supply chain are modeled. Using this architecture, counterfeiting of drugs can be stopped while making sure favorable conditions for drugs such as temperature is maintained throughout the supply chain. Furthermore, a predictive model to forecast the demand of pharmaceutical drugs using the data stored in the platform is implemented using the BAT.

The rest of the paper is organized as follows. Section II gives a brief description about blockchain technology and a literature review of the existing related work. The design principles of the platform are explained in Section III whereas the design of the BAT is explained in the Section IV. The implementation of the case study is presented in Section V. The results of the performance analysis of the proposed system of platform with BAT is elaborated in the Section VI. Section VII concludes the paper while presenting the future directions of the study.

II. RELATED WORK

The blockchain technology is a computing paradigm which provides distributed architecture for different parties to build trust in a trustless environment without third party involvement [5]. Bitcoin was introduced by Satoshi Nakamoto as a pure decentralized peer-to-peer electronic cash in 2008 which marked the initial implementation of blockchain [5]. Blockchain is a distributed database or a ledger which contains timestamped records. These records are known as blocks. Blocks are protected by cryptographically and linked to the previous block [6]. A transaction in blockchain is verified by peers in the network. Without knowing each identity a peer can verify a transaction and add it to the blockchain using the cryptographic hash of the block. Transaction history is visible via public keys but participants are anonymous. Peers need to verify a block before adding a transaction to the blockchain and distributed peers should agree on the order of the transactions before the block is added into the blockchain to maintain the integrity. This is known as the consensus mechanism. This ensures blocks are valid within the network. There are different types of consensus mechanisms used by blockchain technologies such as Proof-of-Work [7], Proof-of-Stake [7], voting-based consensus [8].

Developing of IoT platforms using blockchain technology has been attracted by many researchers and developers due to many reasons.

The existing IoT platform architecture is highly centralized and with the rapidly developing IoT industry, the centralized architecture will not be efficient or scalable to embrace the growing number of IoT devices. The decentralized architecture of blockchain, will be efficient and it will be able to manage the increasing number of IoT devices and resist a single point of failure.

Access control is the ability to control who has access to the data. There were huge data breaches when using apple fitbit [3] where the data of the users were accessed by 3rd parties without the consent of the users. With the blockchain technology the user can be sure that their data is not used without proper authorization. The data becomes tamper proof which secure the integrity of data. If the data is changed or altered, the blockchain technology can figure out where the data has been changed and the timestamp of the block change which makes the system tamper proof.

Most of the blockchain based IoT platforms are focused on the access management control. FairAccess [9] is one of the platforms which improve access control by using a novel type of blockchain. Seyoung H. et al [10] also has proposed a platform using blockchain for the access management of IoT devices. Ethereum [11] is used as the type of blockchain and access management is controlled using smart contracts. The device act as a node in the blockchain network in these studies. However, when the device acts as a node, the performance of the blockchain reduces due to the power and performance constraints in the devices.

Lei H. et al has created a blockchain platform [12] with a novel method of access control. The platform introduced is created in a layered architecture where modifications can be done to each layer without affecting the other layers. The performance of the platform is comparatively high with the opti-

mized execution procedure of the system. However, the data storage system and the querying mechanism of data is not optimized in this study. Oscar N. [13] proposed a platform which mainly focuses on the access control of distributed networked sensors which are connected geographically. The scalability of the system is increased using a specific type of node called management hub nodes as it would enable connection of networks simultaneously. Due to the constraints related to the processing time, the proposed platform could face limitations such as the denial of the addition of devices to the network, unauthorized user accessing the information to be processed by the management hub.

Researchers have shown interests in utilizing the blockchain technology specifically for the industrial IoT (IIoT). BPIIoT [6] is a blockchain platform for IIoT which can be used to create distributed applications (DApps) for manufacturing. Using the Dapps provided by the BPIIoT platform a machine can perform transactions with another machine or a consumer without a third-party control. However, less focus is towards the data management of the platform.

Lightweigh Scalable Blockchain(LSB) [13] is a novel blockchain, specially designed for constrained IoT devices which addresses many security and privacy issues. It is a time-based algorithm, used as consensus instead of using PoW and PoS which make the blockchain more lightweight. The introduced blockchain will reduce the overhead, which is very useful in real time applications. The study is more concentrated on the consensus and security mechanisms where the data storage management is not optimized in the approach.

There are numerous studies conducted to explore the data storage management of blockchain technology. BeeKeeper [14] is an IoT platform which is used for homomorphic computation and secure storage. The architecture of the beekeeper is based on a beehive. The devices are considered as bees and the blockchain network together with the servers, act as the beehive. The system creates more beehives with the addition of more devices to the network. The data collected is processed with homomorphic computations [15] without learning. Sapphire [16] is an IoT platform which is specially proposed for data storage management. This system is designed in a manner where the storage architecture could be specifically used for the data analytics application of IoT. The hashing mechanism used by the system is Location and Type Sensitive (LTS). Hence this architecture is not suitable to be used in instances where geographical location is not particular.

IoTA [17] is the most popular IoT blockchain platform which is using its own blockchain tangle. It is using a novel blockchain architecture and the platform has proved to be working with the best performance. Tangle, the type of blockchain used by the IoTA also act as a cryptocurrency. However, IoTA uses PoW as the consensus mechanism, which reduce the overall performance of the platform.

Apart from these blockchain based IoT platforms, studies have been conducted to improve the data analytics application of blockchain [18]. Massimo B. et al [19] describes a mechanism to integrate blockchain data with other existing databases. Paul T. [3] describes the way blockchain and big data could be used to store medical data in a more secure manner. Also, the paper elaborates about the advantages and disadvantages of using blockchain technology to store medical data.

However, the studies conducted to build data analytics applications on the data stored in the blockchain still have many constraints and drawbacks. Section IV provides an overview of these drawbacks while introducing a new approach to mitigate these constraints and drawbacks.

III. ARCHITECTURE OVERVIEW OF THE PROPOSED PLATFORM

In this Section, we present the proposed architecture of the platform with high level details about the components of the platform. The Section III-A provides detailed explanations of the way each component connected for the intended functionality of the platform. For this study Hyperledger [20] is chosen as the candidate blockchain technology as it is the most suitable type of blockchain for handling business logic currently [21].

Figure 1 shows the architecture overview of the platform. A modular architecture is adopted with a layered structure which makes sure that each layer could be designed separately without altering the basic architecture. The design is basically divided into 4 layers. Each layer is interfaced with the other layer through a communication medium.

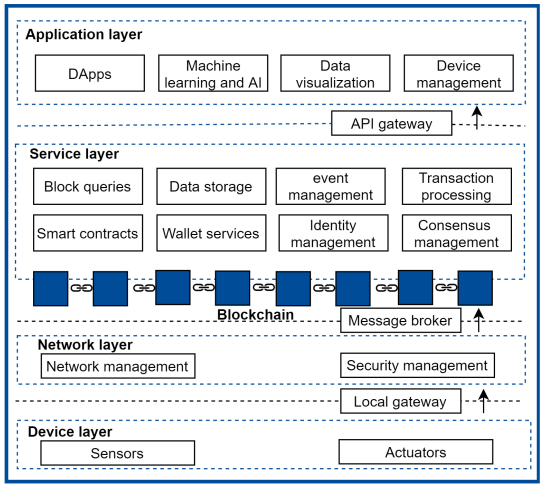

FIGURE 1. The architecture overview of the proposed platform.

The bottom layer of the platform consists of the devices which can be connected to the network. The device layer is connected through a local gateway to the network layer.

The network layer manages the device layer. The security and the network management of the device layer is performed by the network layer. Basic security protocols are implemented for the transmission of data and control

between the device layer and the blockchain such as data encryption to prevent unauthorized listening and understanding of the content. In the proposed system MQTT [22] is used as the protocol for communication of IoT devices as it is lightweight and more suitable for the communication between constrained devices as it has a small header [23]. Furthermore, the network layer uses Transport Layer Security (TLS) [24] for the encryption of data. The network layer is connected with the blockchain network through a message broker. In the implementation, MQTT broker is used as the message broker.

The blockchain is the main actor in the platform. The service layer is found interconnected with the blockchain. Smart contracts are one of the most important services provided by the service layer. A smart contract is a special code/ program where there are few conditions mentioned [25]. When the conditions are met by a specific user or a device, mutual authentication or access could be granted. A transaction occurs when the input variables are satisfied with the logic function given. From one instance to the other the specific logic function changes. The use of a smart contract removes the necessity of use of a central authorization mechanism or a 3rd party access of the system. The platform creates different smart contracts according to the specific situation.

Wallet services help the platform in the process of identity management of the platform. These wallets are produced by each Certification Authority of the blockchain network. A wallet contains digital certificates and security keys which can be used for the identification of a component connected to the network. The validation of the certificates provided by the components are performed by the particular organization which issued the specific wallet to the user. This makes sure that the identity management service of the platform is decentralized between the organizations of the blockchain. Furthermore, each component connected to the platform can be granted with levels of privileges with access control. A user can have privileges of an admin, writer or a reader [21]. Hence, the control a user has over the platform is reduced, i.e., client user of the platform will not be able to alter the configurations of the platform.

The transaction processing is one of the major uses of the platform. A consensus mechanism is used for the ordering and validation of transactions. The platform uses a permissioned voting-based consensus mechanism as explained in the Section II. An endorsement process is performed [26] for the validation of the blocks. The endorsement policy defines which organization needs to approve the transaction. In our proposed platform, all the organizations connected to the network should validate the transaction. Event management is another service found on top of the service layer. In the proposed platform the data is requested from the device layer with a triggering of an event.

Blockchain stores data of all the transactions processed in the platform. Data related to peer to peer communication is stored as blocks without the control of a central authority in the data storage. Hence blockchain acts like a data warehouse

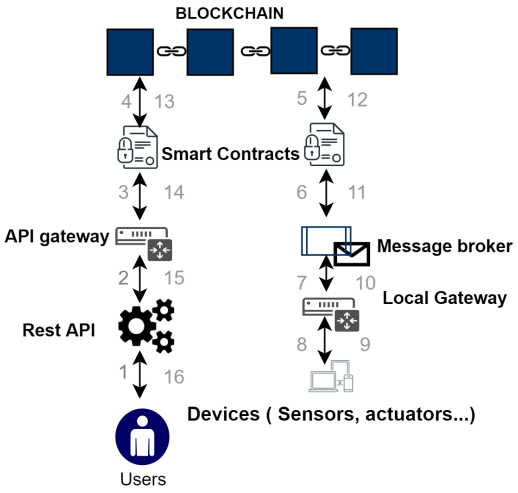

FIGURE 2. Transaction execution procedure .

[27] which stores data from different sources (IoT devices, DApps data, management data...). This data is very useful especially for industries and business. With the proposed system data can be retrieved and data visualized easily specially for business analytics.A novel approach to query the blockchain is introduced in the Section IV.

On top of the services provided by the blockchain, the application layer which exposes these services to the external users can be seen. This layer is interfaced with the service layer using an API gateway. The transaction processing and interacting with front end applications are developed through the DApps.

In the proposed system, a special architecture called BAT facilitates the development of ML and AI applications using the data stored in the blockchain, as explained in the Section IV In the platform, each IoT device is registered before data transmission from the device. Device management in the application layer is used to handle IoT devices. The registration and management is done through a smart contract specifically created for that purpose. Each device is also provided with a wallet to prove its identity. Hence this mechanism will make sure that unauthorized devices cannot communicate in the system. This provides a solution to the problem of integrity of IoT devices.

A. EXECUTION PROCEDURE OF A TRANSACTION

Figure 2 shows how the different services of the layers in the architecture collaborate together to perform a transaction.

First of all, when a user wants to interact with the proposed platform, the user should first provide the identification certificates which are stored in the wallet of the user to the blockchain network. The wallet is issued to a user by a specific organization established in the network.This process is done through the API gateway connected to the rest API. If the user fails to produce the proper credentials, the users will not have access to the platform.

After the authentication, the request for the specific transaction or the process is submitted to the API gateway through the application. The smart contract will be invoked once the request is received by the blockchain network. If the necessary conditions are fulfilled, the endorsement peer will approve the transaction and commit the transaction. The other entities of the network are informed about the transaction through the orderer [20] of the network and all the peers will update their ledgers about the transaction. Then, the response of the transaction is sent back to the application through the gateway.

If the particular smart contract requires data readings of the IoT devices connected to the network, an event is triggered through the chaincode and the network subscribes to the message broker (MQTT broker) of the system. Through the local gateway, the sensors and actuators will publish the encrypted data.

IV. BLOCK ANALYTICS TOOL (BAT)

The storage system on the blockchain is specifically designed to handle blocks of data and store the transactions as objects. The storage system is specially optimized to increase the efficiency of transactions in the blockchain. It is neither optimized to perform complex queries nor as an efficient data retrieval schema. Hence in most of the studies that have been conducted off chain database which contains the same data stored in blockchain is used instead of using the on-chain database. As recommended by the Hyperledger Fabric developers [20], mirror storage facility can be used to run concurrently with the blockchain storage system which replicate all the data in both locations for the data analsysis. However, that is a huge waste as it double of all the resources used to store and analyze data. The mirror storage stores redundant data that would not be useful for any future reference. Moreover, updating the storage system at two locations concurrently, reduces the processing time of a transaction which affects in reducing the performance of the platform.

In some studies the on-chain network is used to store the security key and the data would be stored in the off-chain database [28]. Hossein S. et al [29] have taken the basic cloud architecture and decoupled the data plane and control plane of the architecture and restructured the architecture to be used as an IoT platform. The usage of blockchain technology in the control plane has enabled the control of data and access control by the proposed system itself without a centralized authority. The data is stored in a separate storage and the hash pointer is stored at the blockchain. In this mechanism, everytime a transaction is processed, the data stored in the off-chain network has to be retrieved to identify the state of the block. For instance, in supply chain transactions, the current owner of the specific object has to be identified to perform the transaction between the current owner and the new owner of the object. For this purpose, off-chain database has to be queried. Hence, the performance of transaction processing will be reduced drastically.

Owing to the reasons explained about the constraints in the state of the art approaches in designing ML and AI applications on data stored in the blockchain, we designed a novel approach which would reduce the cost of a mirror storage or an off-chain database and would optimize the querying mechanism of blockchain.

The processes and the functionality of each and every part of the BAT are explained below as shown in the Figure 3.

The user provides the necessary configuration files through an API which would define the user specific requirements. As shown in the 3, the user A and user B can use 2 instances of the BAT and they can process different data analysis tasks parallel without affecting the other task. Hence, the tasks can be separately run in different environments as specified by the user through the configuration files.

The user interactions are always performed through smart contracts as shown in 3. Hence, even if a user tries to change the data of the blockchain though the BAT, it cannot be executed as the queries are performed in only one of the ledgers in the blockchain network. Other ledgers and peers in the network would not be updated about the change of data which would fail the endorsement/consensus of the transaction which results in denying the alternation.

A. BLOCK INDEX

As mentioned earlier rich queries can degrade the performance of the blockchain system. Furthermore, to query a single block, all the blocks in the blockchain have to be searched. Acquiring the data from the blockchain acts as a bottleneck of the proposed system. Hence, we propose a novel approach that can be used to optimize the data acquiring process from the blockchain.

A special indexing system called block index is proposed in this architecture to reduce the search time in the blockchain. This index system would be more suitable when blockchain is used for transactions especially in business logic. Blockchain technology is used to handle transactions between two parties. A unique ID is given for all the transactions recorded in the blockchain. In transactions what happens is the change of the ownership of this specific object while the other details about the object remains the same.

For an example, in vehicle trading between two parties what happens is that the ownership of the vehicle changes while specifications about vehicle such as engine capacity, fuel capacity, dimensions, power remains the same.

Moreover, when we want to search through a blockchain, if we want to get details about a particular object, the search pointer goes through many blocks that contain the same details which increase the search time to a large extent.

To address these drawbacks, we propose the block index system. In the block index system, the first transactions that occur related to a particular object are stored in the first column of the index. When more transactions occur related to the same object, they are stored in the second, third ... columns of the index.

In the architecture of the blockchain, transactions are recorded in a manner where all the data would be duplicated

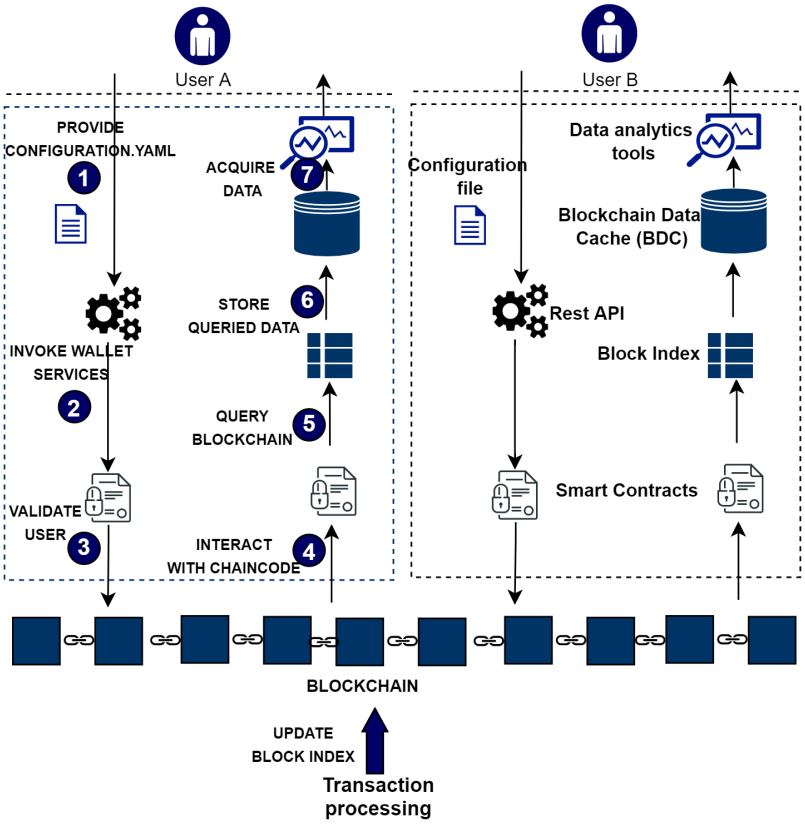

FIGURE 3. Overview of the Block Analytics Tool (BAT).

into the newly created block with the changed parameter. Thus, in our index system, if we want to query details that would not change over transactions, we can easily go through the first column of the index and find the necessary block.

For instance, figure 4 shows part of a blockchain which is used to store data about trading of cars produced by a company. The blockchain is used to keep track of the current and previous state of the cars.

Imagine that we want to find a car with specific details. In the given example, block P represents a car with the given specifications. If it queries through the same blockchain architecture, the pointer would have to go through all the blocks. However, in the proposed block index system, the pointer only has to go through the first column to acquire the information.

This reduces the time complexity (T(n)) of query time from O(n) to O(m) where n is the input size and n>m. In the index system, the width of a row represents the number of transactions related to a specific object. The adjacent columns of the index contain the details about the same object with ownership and few other details changed. This becomes very efficient once the number of transactions for objects are higher. After initially creating the blockchain network, every time a transaction occurs, an event is triggered to update the block index. Instead of updating a database with the same set of details, updating an index would save processing time as well as the storage.

1) Implementation of the block index

Algorithm 1 Pseudo code to add a block to the block index

Input: x[ID] - ID of new block
Output: column of the index, row of the index
Initialization:
nRows0 = number of rows in column 0
nColumnsi = number of columns in row i
IndexColumn = 0
for $i = 0$ to $nRows0$ **do**
  if $(i[ID] = x[ID])$ **then**
    for $IndexRow = 0$ to $nColumnsi$ **do**
      IndexColumn = IndexColumn+1;
    end for
  end if
end for
return $IndexColumn, IndexRow$

In the implementation, CouchDB [30] database is used as the state database. CouchDB provide an index system which can be used to query the database easily. The proposed block index makes use of this system. Algorithm 1 shows the pseudo code of the block index. When a block is created in the blockchain or transfer object, through the chaincode

the block index is updated. If the block contains details about a newly created object, the block index would store the transaction of the object under the column 0. When the object undergoes another transaction, the transaction would be recorded in the second column through the chaincode. When querying the blockchain, the couchDB make use of this block index created to retrieve the information.

B. BLOCKCHAIN DATA CACHE

Through the configuration files the user has the capability of requesting a specific set of data according to the requirements. Then, through the BAT the data is queried from the blockchain using the Block Index and the retrieved data is stored in the Blockchain Data Cache(BDC). The BDC acts as a temporary storage for the retrieved data. In our study we support both on-premise and cloud storage systems to be used as the BDC. In the creation of the BDC, the event manager and the ordering service of the blockchain network makes sure about the synchronization of the blockchain network with the BDC to maintain the ACID (Atomicity, Consistency, Isolation, and Durability) properties [31] of the blockchain database and the BDC. The created BDC can be pulled down with the destruction of the resources phase in the BAT.

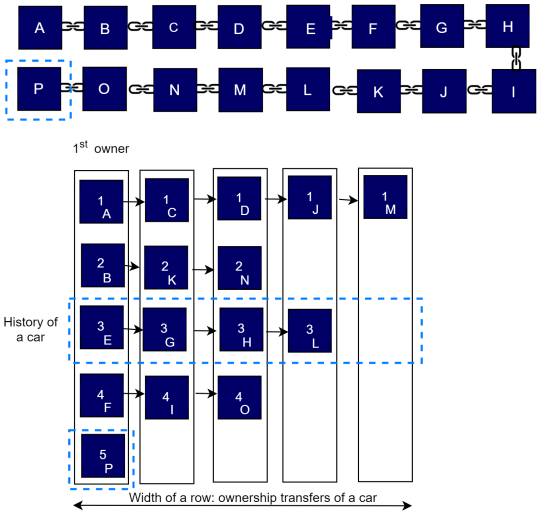

FIGURE 4. Architecture of Block Index.

C. BLOCKCHAIN DATA ANALYSIS

The data analysis, ML tools and frameworks (Tensorflow [32], Pytorch [33] ...) as specified by the user through the configuration files, can be integrated with the BDC. Unlike the blockchain data storage system, the tools can perform rich queries, preprocess data stored in the BDC. Furthermore, as the BDC in the study are isolated from one another, the tools can write data to the BDC as needed. The tools can utilize the new blocks which are added to the blockchain while building the model, as the BDC is updated with the addition of data.

V. CASE STUDY

A. USE CASE CONCEPTUAL SCENARIO

A pharmaceutical supply chain is used as the use case scenario in the study. Drug counterfeit is a severe problem in pharmaceutical supply chains. This occurs due to the lack of transparency in the transportation of drugs through the supply chain. Certain medicines also require to have proper conditions maintained. These details also remain hidden and no trusted method to monitor these conditions. According to the World Health Organization, more than 10% of medicines worldwide are counterfeited [34]. Counterfeiting happens in several ways such as manipulating the expire date, producing with no active chemical ingredients, wrapping in forged packages. [35]. After distributing these ill-treated drugs, users are unable to identify these counterfeit medicines. There is no proper mechanism to verify the integrity of whether the original package is distributed by the 3rd party logistics company [34].

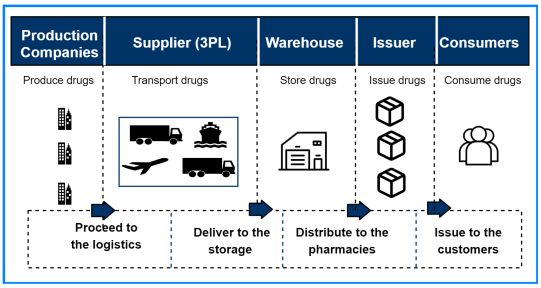

FIGURE 5. Use case scenario.

Most of the above-mentioned problems can be addressed by establishing trust between the parties in the supply chain. A blockchain based IoT platform can be used to process transactions that occur in the supply chain. RFID tags are used to prevent tampering expire date, manipulating dosage information in the medicine package. Drug manufacturers add initial package data into the Blockchain and RFID tags which can be used to verify the details at any given point using the proposed platform. To track whether the drugs were transported using correct conditions sensors could be used. For example, temperature sensors are used to check whether the temperature of drugs was maintained during transportation or during storage while processing the transaction. It is assumed only sealed packages with RFID tags are accepted by the parties in the supply chain.

Figure 5 shows the scenario of the use case chosen. In the use case, the production companies produce medicine, 3PL (3rd party logistics) supply the medicine to the warehouse. Through the warehouse, the medicine is distributed to the pharmacies or the issuers. Most of the time drug counterfeiting occurs through the suppliers and the warehouse.

B. SYSTEM ARCHITECTURE OF THE CASE STUDY

The implementation and analysis of the proposed system was performed in the gcloud virtual machine instance (configuration of Name- c2-standard-4, Zone- us-central1-f, vCPUs-

8, Memory- 30GB) [36]. The main actors of the scenario are the supplier, warehouse and the issuer organizations. The network is initiated by the warehouse organization and the initial configuration of the network is configured by the warehouse entity. The network is controlled according to the rules imposed by the network configuration. The channel of the blockchain is named as the NCK channel. The channel order is maintained by the orderer connected to the channel. The channel is governed by the channel configuration which has the policies related to all 3 entities. Separate certificate authorities are maintained by each entity for the validation of the transactions processed in the blockchain. Each organization consists of 2 peers and 1 peer is used as the anchor peer which is used to communicate with the other entities. In each peer a copy of ledger containing details about all the transactions are installed. The peers contain copies of smart contracts installed in the network. 4 different applications are created as DApps which are used by different organizations for transactions and processing of data. 2 smart contracts are mainly used by the applications.

- Create batch - a manufacturing company uses this application to create a batch to be delivered to one issuer company. Details about the batch of drugs such as name, dosage, quantity, manufactured and expiry date are added as a block to the blockchain. An RFID tag is assigned for this particular batch.
- Transfer batch - This is used when a batch is transferred from one entity to the other.

Through the application, when an RFID tag is read by a specific entity, the value of the temperature sensors is read to the system and it is checked whether the batch has maintained favorable temperature up to that point. If the temperature has not been maintained at the range or if there is any breach in the data related to RFID tag, the smart contract will automatically be invalidated. The smart contracts and the endorsement policies will make sure that the data is read or added by an authorized person who has necessary wallet configurations.

Four different applications are used by production companies, suppliers, warehouse and issuers for the transactions. Through another application, an end user can examine where a specific batch is located on that occasion. Eventually, the end user at the issuer can see how the batch has been transported, stored and issued by the organizations while maintaining the favorable conditions for the batches with complete transparency. Figure 6 shows the user interface of the application at the end of transferring a batch through the supply chain.

C. DEMAND FORECASTING OF PHARMACEUTICAL DRUGS

All the transactions occur in the supply chain are recorded in the blockchain. Blockchain also acts as a data storage system for transactions. These data can be used for data analytics of the supply chain. The difference between the supply and

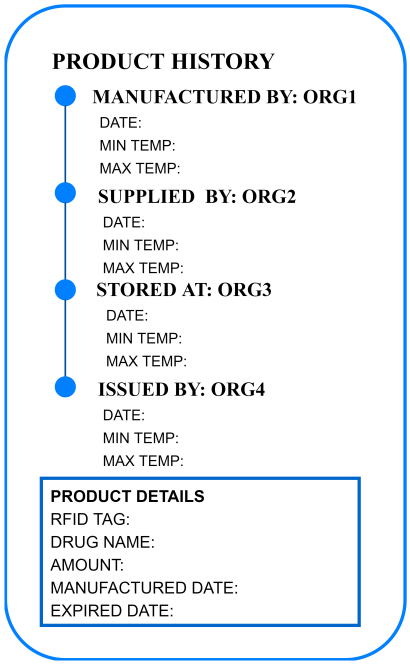

FIGURE 6. User output of product history at the end of transferring a batch through the supply chain.

demand in supply chains is known as bull whip effect [37]. Maintaining this effect in pharmaceutical supply chains is very important. Demand forecasting in the pharmaceutical industry is critical as the availability of drugs at the needed time, impact on patient's life [38]. Furthermore, the demand for drugs by different pharmaceutical companies is a complex combination of the necessity of drugs, shelf life, regulations and the cost associated with drugs.

The consumption method [39] of forecasting of drugs is the usage of historical data of past consumption of drugs.

As the data is recorded in the blockchain in a well-ordered manner, it can be used to predict future requirements of drugs easily.

We used this use case scenario to experiment the block analytics application designed. With the help of hospital prescribing dispensed in the community dataset [40], batches of drugs were created. The data about hospital prescribing data related to the Manchester University NHS Foundation [40] were used for the study. The batches were added in a manner where batches would be transferring through the supply chain at different stages. A block contains details about RFID tag, drug name, dosage, quantity, cost, organization (e.g.-production, supplier, etc), temperature sensor readings, transaction time, manufactured and expired dates. During a transaction the organization, temperature sensor readings change while the other details remain the same. The block index was updated with the addition of batches to the platform.

Figure 7 shows part of the block index which shows the state of transfer of batches through the supply chain. The transaction ID is used to map the block index to the

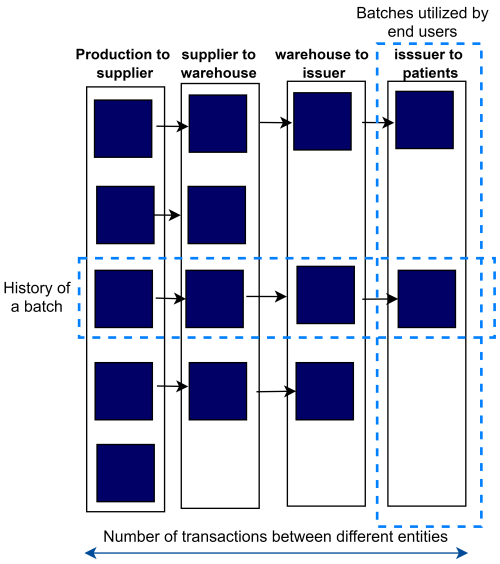

FIGURE 7. Block index for the use case scenario.

blockchain.

For the demand forecasting analysis, the details about drug name, quantity, cost and issued date of batches had to be extracted. According to the block index designed, the details about issued batches to the patients are available in the 4th column. Using the block index, the specific batches could be extracted easily without querying the full blockchain. If the block index was not used for this purpose, the querying would be more complicated with more functions and algorithms.

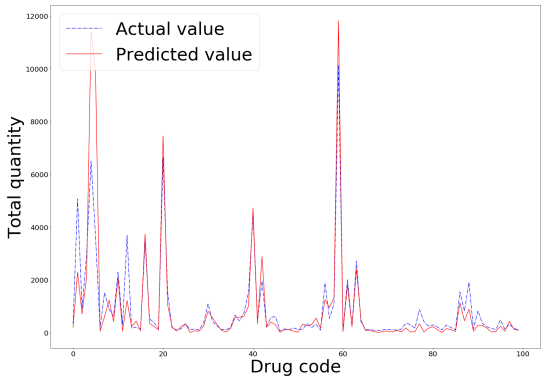

FIGURE 8. Predicted and actual value comparison for the demand forecasting.

The retrieved blocks are stored in the BDC. For the study, non-relational database MongoDB [41] was used as the candidate on-premise database for the BDC. Data were preprocessed using the BAT and ML model was built on Tensorflow. The demand analysis of supply chain data is a time series analysis as the previous storage, seasonal weather patterns, average salary is affected by the time. Pharmaceutical drugs show correlations between each other as many drugs are prescribed as a set for a specific disease. While preserving these correlations, the time series analysis model was built using xgboost regressor [42] with SVR (Support Vector Regression) [43] as the base estimator.

Figure 8 shows the demand prediction and actual values of total quantity of certain types of drugs for the month October of 2019 for the hospitals in Manchester University NHS Foundation [40]. The model was built with the an average accuracy of RMSE (Root Mean Square Error) value of 0.5213. The model utilize previous 12 months of data while analyzing the correlations of the drugs and the demand of drugs for the next month can be predicted. For an instance, the drug quantity needed by "Amlodipine-Tab 5mg" of drug code "0206020A0" can be predicted with an RMSE of 0.3134.

The BDC can be pulled down after the analysis. The BAT saves the state of the model, BDC, and query terms are saved and when the model is required again, the BAT will invoke the state and the BDC will be updated with the newly added blocks of batches to the blockchain.

VI. PERFORMANCE ANALYSIS

In this Section we present the analysis of the performance of the proposed system of platform with BAT. Figure 9 shows

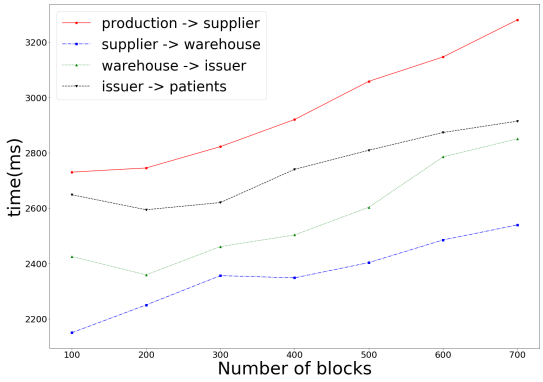

FIGURE 9. Total transaction processing time.

the total transaction processing time taken by different applications for creating and transferring batches with the help of IoT devices connected. The transaction processing time increases when the number of blocks in the blockchain increases and highest transaction time is taken for the creation of the new batch which is the transaction between production and supplier companies. This is due to the additional device registration time related to RFID and temperature sensors.

The traditional methodology of the usage of off-chain database with the introduced novel approach is compared to present the performance between the two methods. Transaction processing in a secured environment is the ultimate goal of using blockchain technology. Hence, the performance of the transaction processing should not decrease with the addition of other services in to the platform.

Figure 10 shows the total event processing time spent to update the block index and the addition of a new block

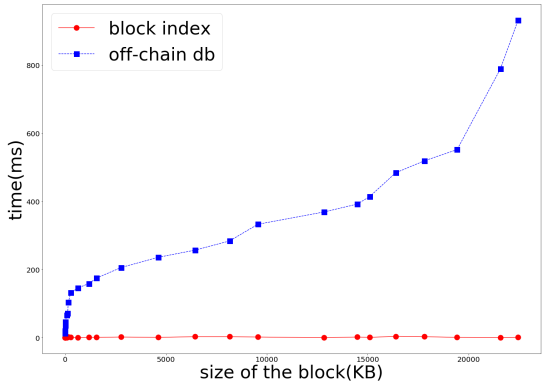

FIGURE 10. Total event processing time.

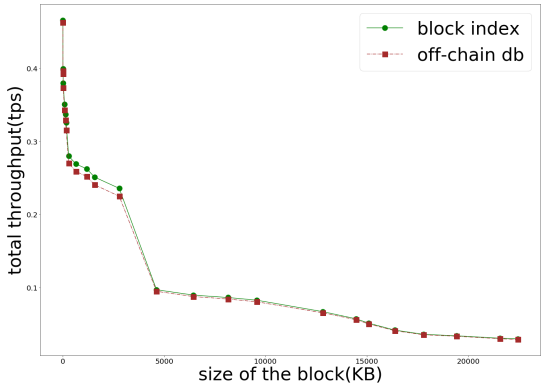

FIGURE 11. Throughput of transactions.

to the couchdb. As explained at the Section IV, more time is exploited to duplicate the blocks into another off-chain database. When the block size increase, the time taken for the event processing increase exponentially. On the other hand, a constant time between 0-5ms is used to update the block index.

Transaction throughput is the number of transactions which can be processed in a second. The throughput decreases drastically with the increase of size of a block. Furthermore, the throughput with the usage of block index is high and it is notable when the block size is less than 5MB.

Furthermore, the off-chain database will duplicate all the data in the blockchain adding huge redundancy of data whereas the size of the block index will remain less than 1MB most of the time.

A new index system is introduced to query the blockchain in this paper. The performance between the 2 approaches of using the block index and the normal querying architecture is compared as shown in the table1 through the scenarios in the case study. In the table, the batches issued to the users means retrieving data of the 4th column of the block index. Selecting quantity for the batch X is getting data about a batch, which is querying through the first column and getting the data without going through all the blocks.It is very convenient to obtain the data related to one organization or transaction

TABLE 1. Performance analysis of the block index

Query	With block index (ms)	Without block index(ms)
Select blocks which are issued to users	1235	2786
Select quantity for batch x	312	561
Get batches of drug type z	376	542
Get transaction history for batch x	527	533

through the block index which increase the efficiency by 100% as shown in the table. The search pointer moves only through the first column and checking the block types where drug type = z, to get the details about the batches of drug type z. The search pointer would not have a special exploitation in using the block index to obtain the transaction history for a batch, as it would have to go through all the columns to get the transaction history.

VII. CONCLUSIONS AND FUTURE WORK

The rapidly growing autonomous systems which integrate IoT devices, network and storage systems have many vulnerabilities. In this paper we design and implement a blockchain based IoT platform which address scalability, data security and access control problems found in these autonomous systems. The devices can be connected to the blockchain and monitored through the end user applications. Furthermore, we introduce a novel approach to integrate machine learning and Artificial Intelligence (AI) technology to perform secure learning and prediction methods using data analysis with the introduction of BAT. The BAT uses a new method to query the blockchain with a block index. This index is ideal to be used in systems where blockchain is used to record and manage business transactions. The queried data is stored in a BDC which act as a temporary data storage to be utilized by data analytics, ML platforms and frameworks. After the data has been analyzed, the created BDC could be pulled down using the BAT. The implementation of the case study depicts how the proposed system of blockchain based IoT platform with BAT meets the defined design principals and requirements. This novel approach of BAT utilizing the BDC saves resources such as storage facilities compared to the traditional approach of creating mirror storage which duplicate resources. Furthermore, the total transaction processing time for the new approach is comparatively low due to the minimal event processing time related to the creation of block index compared with the off-chain database. The usage of block index reduces the cost associated with queries saving processing time which increase the efficiency of the proposed system. In addition to that, the architecture can effectively be utilized as the state of the created forecast models can be reused with the addition of new blocks upon request by the user. Moreover, the usage of smart contracts to invoke the BAT make sure that unauthorized parties will not have the

access to use or modify the data.

In the future we will mainly focus to bring the platform towards edge computing with the optimization of the BAT for real-time processing. The current implementation of the proposed BAT is dependent on the performance and the functionality of the Hyperledger blockchain. Future versions of the BAT will also be compatible with other types of blockchain technologies.

REFERENCES

[1] I. Lee and K. Lee, "The Internet of Things (IoT): Applications, investments, and challenges for enterprises," *Business Horizons*, vol. 58, no. 4, pp. 431–440, 2015.

[2] A. Panarello, N. Tapas, G. Merlino, F. Longo, and A. Puliafito, "Blockchain and iot integration: A systematic survey," Sensors, vol. 18, p. 2575, 08 2018.

[3] M. Z. A. Bhuiyan, A. Zaman, T. Wang, G. Wang, H. Tao, and M. M. Hassan, "Blockchain and big data to transform the healthcare," in *Proc. Int. Conf. Data Processing and Applications*, ser. ICDPA 2018. New York, NY, USA: ACM, 2018, p. 62–68, accessed: Jan. 22, 2020. [Online]. Available: https://doi.org/10.1145/3224207.3224220

[4] P. Voigt and A. Von dem Bussche, "The eu general data protection regulation (gdpr)," A Practical Guide, 1st Ed., Cham: Springer International Publishing, 2017.

[5] S. Nakamoto, "Bitcoin: A Peer-to-Peer Electronic Cash System," *Journal for General Philosophy of Science*, vol. 39, no. 1, pp. 53–67, 2008.

[6] A. Bahga and V. K. Madisetti, "Blockchain Platform for Industrial Internet of Things," *Journal of Software Engineering and Applications*, vol. 09, no. 10, pp. 533–546, 2016.

[7] V. Gramoli, "From blockchain consensus back to Byzantine consensus," *Future Generation Computer Systems*, 2017, accessed: Jan. 22, 2020. [Online]. Available: http://dx.doi.org/10.1016/j.future.2017.09.023

[8] G.-T. Nguyen and K. Kim, "A survey about consensus algorithms used in blockchain." *Journal of Information processing systems*, vol. 14, no. 1, 2018.

[9] A. Ouaddah, A. Abou Elkalam, and A. Ait Ouahman, "FairAccess: a new Blockchain-based access control framework for the Internet of Things," *Security and Communication Networks*, vol. 9, no. 18, pp. 5943–5964, 2016.

[10] S. Huh, S. Cho, and S. Kim, "Managing iot devices using blockchain platform," 01 2017, pp. 464–467.

[11] G. Wood, "Ethereum: a secure decentralised generalised transaction ledger," *Ethereum Project Yellow Paper*, pp. 1–32, 2017, accessed: Jan. 21, 2020. [Online]. Available: https://ethereum.github.io/yellowpaper/paper.pdf

[12] L. Hang and D. H. Kim, "Design and implementation of an integrated iot blockchain platform for sensing data integrity," *Sensors (Switzerland)*, vol. 19, no. 10, 2019.

[13] A. Dorri, S. S. Kanhere, and R. Jurdak, "Towards an optimized blockchain for IoT," in *Proc. 2017 IEEE/ACM 2nd Int. Conf. IoT Design and Implementation*, no. October, Pittsburgh, PA, USA, 2017, pp. 173–178.

[14] L. Zhou, L. Wang, Y. Sun, and P. Lv, "BeeKeeper: A Blockchain-Based IoT System with Secure Storage and Homomorphic Computation," *IEEE Access*, vol. 6, no. 8, pp. 43 472–43 488, 2018.

[15] C. Gentry, "Fully homomorphic encryption using ideal lattices," in *Proc. 41st ACM Symp. on Theory of Computing*, ser. STOC '09. New York, NY, USA: ACM, 2009, p. 169–178.

[16] Q. Xu, K. Mi, M. Aung, Y. Zhu, and K. L. Yong, "New Advances in the Internet of Things," *Studies in Computational Intelligence*, vol. 715, pp. 119–138, 2018.

[17] S. Popov, "The tangle," ABA Journal, no. FEB., pp. 1–25, 2016, accessed: Jan. 23, 2020. [Online]. Available: https://iota.org/IOTA_Whitepaper.pdf

[18] C. G. Akcora, M. Kantarcioglu, and Y. R. Gel, "Blockchain Data Analytics," in *Proc. - IEEE Int. Conf. on Data Mining*, vol. 2018-November, no. December 2018, 2018, p. 6.

[19] M. Bartoletti, S. Lande, L. Pompianu, and A. Bracciali, "A general framework for blockchain analytics," *SERIAL 2017- Colocated with ACM/IFIP/USENIX Middleware 2017 Conf.*, 2017.

[20] Hyperledger, "Hyperledger – Open Source Blockchain Technologies," 2019, accessed: Jan. 23, 2020. [Online]. Available: https://www.hyperledger.org/

[21] E. Androulaki, A. Barger, V. Bortnikov, C. Cachin, K. Christidis, A. De Caro, D. Enyeart, C. Ferris, G. Laventman, Y. Manevich et al., "Hyperledger fabric: a distributed operating system for permissioned blockchains," in *Proceedings of the Thirteenth EuroSys Conference*, 2018, pp. 1–15.

[22] U. Hunkeler, H. L. Truong, and A. Stanford-Clark, "Mqtt-s—a publish/subscribe protocol for wireless sensor networks," in *3rd Int. Conf. on Communication Systems Software and Middleware and Workshops*, 2008, pp. 791–798.

[23] N. Naik, "Choice of effective messaging protocols for iot systems: Mqtt, coap, amqp and http," in *IEEE int. systems engineering symposium*, 2017, pp. 1–7.

[24] T. Dierks and E. Rescorla, "The transport layer security (tls) protocol version 1.2," 2008.

[25] V. Buterin et al., "A next-generation smart contract and decentralized application platform," *white paper*, vol. 3, no. 37, 2014.

[26] P. Thakkar, S. Nathan, and B. Viswanathan, "Performance benchmarking and optimizing hyperledger fabric blockchain platform," in *IEEE 26th Int. Symp. on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems*, 2018, pp. 264–276.

[27] B. Devlin, Barry/Cote, and L. Doran, Data warehouse : from architecture to implementation. Addison-Wesley Longman Publishing Co., Inc., 1997.

[28] L. Bai, M. Hu, M. Liu, and J. Wang, "BPIIoT: A Light-Weighted Blockchain-Based Platform for Industrial IoT," *IEEE Access*, vol. 7, pp. 58 381–58 393, 2019.

[29] H. Shafagh, L. Burkhalter, A. Hithnawi, and S. Duquennoy, "Towards blockchain-based auditable storage and sharing of iot data," in *Proc. 2017 Cloud Comput. Security Workshop*, ser. CCSW '17. New York, NY, USA: ACM, 2017, p. 45–50, accessed: Jan. 23, 2020. [Online]. Available: https://doi.org/10.1145/3140649.3140656

[30] Apache, "Apache CouchDB - About," 2019, accessed: Jan. 21, 2020. [Online]. Available: https://couchdb.apache.org/

[31] J. Han, E. Haihong, G. Le, and J. Du, "Survey on nosql database," in *6th int. conf. on pervasive computing and applications*, 2011, pp. 363–366.

[32] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin et al., "Tensorflow: Large-scale machine learning on heterogeneous distributed systems," arXiv preprint arXiv:1603.04467, 2016.

[33] N. Ketkar, "Deep Learning with Python," *Deep Learning with Python*, pp. 195–208, 2017.

[34] H. Campbell, "Dangers of drug importation: a case of counterfeit cancer drugs," pp. 1–3, 2016, accessed: Jan. 22, 2020. [Online]. Available: https://catalyst.phrma.org/dangers-of-drug-importation-a-case-of-counterfeit-cancer-drugs

[35] Bayer, "Background Information on Counterfeit Drugs," pp. 1–7, 2014, accessed: Jan. 20, 2020. [Online]. Available: https://www.bayer.com/en/background-information-on-counterfeit-drugs.aspx

[36] "Virtual machine instances," accessed: Feb. 24, 2020. [Online]. Available: https://cloud.google.com/compute/docs/instances

[37] R. Metters, "Quantifying the bullwhip effect in supply chains," *Journal of Operations Management*, vol. 15, no. 2, pp. 89–100, 1997.

[38] I. N. Rachmania and M. H. Basri, "Pharmaceutical inventory management issues in hospital supply chains," *Management*, vol. 3, no. 1, pp. 1–5, 2013.

[39] G. Candan, M. Taskin, and H. R. Yazgan, "Demand Forecasting In Pharmaceutical Industry Using Neuro-Fuzzy Approach," *Journal of Military and Information Science*, vol. 2, no. 2, p. 41, 2014.

[40] NHS, "Hospital prescribing dispensed in the community | NHSBSA," 2019, accessed: Jan. 21, 2020. [Online]. Available: https://www.nhsbsa.nhs.uk/prescription-data/prescribing-data/hospital-prescribing-dispensed-community

[41] "MongoDB," accessed: Jan. 21, 2020. [Online]. Available: https://www.mongodb.com/

[42] R. E. Schapire and Y. Singer, "Improved boosting algorithms using confidence-rated predictions," *Machine Learning*, vol. 37, no. 3, pp. 297–336, dec 1999, accessed: Jan. 23, 2020. [Online]. Available: https://doi.org/10.1023/A:1007614523901

[43] X. Huang, A. Maier, J. Hornegger, and J. A. Suykens, "Indefinite kernels in least squares support vector machines and principal component analysis," *Applied and Computational Harmonic Analysis*, vol. 43, no. 1, pp. 162–172, 2017.

CHATHURANGI EDUSSURIYA is currently pursuing the B. Sc. degree in Computer Engineering from the University of Peradeniya Sri Lanka. Her research interest includes Data Analytics, Blockchain, AI, IoT and Cloud Computing.

KASUN VITHANAGE is a final year undergraduate at the Department of Computer Engineering, Faculty of Engineering, University of Peradeniya Sri Lanka. At present, he mainly researches about Blockchain-based Decentralization, Distributed Computing, Cloud Computing and IoT

NAMILA BANDARA is currently pursuing the B. Sc. degree in Computer Engineering from the University of Peradeniya Sri Lanka. His research interests include IoT, Blockchain and embedded systems.

JANAKA ALAWATUGODA received the B.Sc. degree specializing in Computer Science from the University of Peradeniya, Sri Lanka in 2012 and Ph.D. from Queensland University of Technology (QUT), Australia in 2015. Currently, he is a Senior Lecturer in the Department of Computer Engineering, University of Peradeniya, Sri Lanka. Janaka was a software engineer in IFS RD and an intern at Nippon Telegraph and Telephone (NTT) Corporation, Japan. He is a member of Computer Society (Sri Lanka), Association for Computing Machinery (USA) and the International Association for Cryptologic Research. His main research interest is cryptography, focusing leakage-resilient cryptography, public-key cryptography and authenticated key exchange protocols.

MANJULA SANDIRIGAMA is a senior lecturer at the Department of Computer Engineering, Faculty of Engineering, University of Peradeniya Sri Lanka. His main research interest is cryptography and security protocol development. He also specializes in patents and technology transfer. Manjula obtained his BSc in Electrical Engineering from the University of Peradeniya. After several years of experience in Sri Lanka Telecom, he obtained his MSc and Ph.D. in Computer Science from the Ehime University of Japan. Manjula is also an Attorney-at-Law in Sri Lanka. His present interests include Blockchain and other technology related regulation work.

UPUL JAYASINGHE received the B.Sc. degree in electronics and telecommunication engineering (first-class honors) from the University of Moratuwa, Sri Lanka in 2010 and the M.Sc. from the Asian Institute of Technology, Thailand in 2013. He received his Ph.D. in Computing from Liverpool John Moores University (LJMU), UK in 2018. Dr. Jayasinghe is also offered the Best Thesis Award for his Ph.D. thesis by the Faculty of Engineering, LJMU, UK in the same year. Currently, he is attached to the Department of Computer Engineering, University of Peradeniya, Sri Lanka as a Senior Lecturer. He has worked as a researcher in the Centre for Wireless Communication, University of Oulu, Finland, and Computer Communications and Applications Laboratory, EPFL, Switzerland. His research interests include IoT, Blockchain, Data Analytics, AI, Networking and Security, and Wireless Communication.

GYU MYOUNG LEE received his BS degree in electronic and electrical engineering from Hong Ik University, Seoul, Rep. of Korea, in 1999 and MS, and Ph.D. degree from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Rep. of Korea, in 2000 and 2007, respectively. He is currently a Reader in the Department of Computer Science at Liverpool John Moores University, Liverpool (LJMU), UK. He is also with KAIST Institute for IT convergence, Daejeon, Rep. of Korea, as an adjunct professor. His research interests include future networks, Internet of Things, multimedia services, and energy saving networks including Smart Grid. He has actively contributed to standardization in ITU-T as a Rapporteur (currently Q16/13 and Q4/20), oneM2M and IETF. He is also the chair of the ITU-T Focus Group on data processing and management to support IoT and Smart Cities Communities.