Blockchain-Integrated Framework for Data Security: An Application Based on IoT Data and Deep Learning

Doygun Demirol, Resul Daş, Mehmet Özdem, Ceren Nur Cansel, Davut Hanbay

***Abstract*—The rapid development of the IoT (Internet of Things) ecosystem leads to the creation of big data environments that require real-time analysis. In this comprehensive data ecosystem, anomaly detection and data security emerge as critical requirements. This paper presents a comprehensive approach that integrates a deep learning model developed for anomaly detection in IoT network traffic and a blockchain-based data storage structure designed to ensure data integrity. In the research, network traffic data of a sample device from the N-BaIoT dataset is used. The developed deep learning model was able to classify attack and normal traffic patterns with high accuracy. The proposed model achieved an accuracy rate of 99.79% in anomaly detection, demonstrating its effectiveness in classifying IoT network traffic. Data security is ensured with the Fernet encryption algorithm, while data integrity is protected using blockchain technology. Experimental results show that the proposed system achieves significant performance metrics in terms of both anomaly detection accuracy and data security verification. The proposed framework contributes to the development of more secure and reliable IoT systems by providing an innovative solution to anomaly detection and data security challenges in IoT environments.**

***Index Terms*—Internet of Things (IoT), Deep Learning, Network Security, Blockchain, Data Integrity, Anomaly Detection.**

I. INTRODUCTION

The Internet of Things (IoT) has become a widely used technology in every aspect of daily life. IoT devices have revolutionised many different sectors, from home automation to healthcare, from energy management to smart city applications. These devices have become an important component of the big data ecosystem by continuously generating data. However, this large data ecosystem has also brought security and privacy risks. The vulnerability of IoT devices to cyber attacks has made data security and data integrity issues critical.

In this study, an innovative approach is proposed to detect anomalies in IoT network traffic and ensure data security. Network traffic obtained from IoT devices is analysed with a deep learning based model and normal and attack traffic are successfully classified. To ensure confidentiality in data transmission, encryption operations are performed with a confidentiality algorithm and data integrity is guaranteed by using a blockchain-based structure. Blockchain provides a structure that prevents unauthorised access to data while ensuring secure storage of data.

In this study, a dataset containing the network traffic of a device within the N-BaIoT dataset is used. The data is cleaned, scaled and labelled as attack/normal traffic in preprocessing steps. Then, a deep learning based model is trained to detect anomalies. In the simulation environment, the test data is encrypted at random time intervals and transmitted to a central server and data integrity is ensured by using a blockchain-based structure.

The results obtained show that the proposed system exhibits a high performance in terms of both anomaly detection accuracy and data security. This study provides an effective solution for eliminating security vulnerabilities in IoT networks and increasing data security. Future studies can investigate the applicability of the proposed system on different IoT devices and the scalability of the blockchain structure.

II. BACKGROUND

A. Key Stages of Big Data Analysis

Nowadays, it has become important to understand the meaning and importance of data and to produce and collect data. Understanding data correctly plays an important role in decision-making. Big data analysis is the process by which algorithms are applied to data to extract useful and unknown patterns, relationships, and information [1]. Extracting and analyzing useful information from big data sets requires smart and scalable analysis services, programming tools, and algorithms [2]. Also, big data analytics is used to extract previously unknown, useful, valid, and hidden patterns and data from large data sets and identify important relationships between stored variables. This section examines four main phases of big data analysis processes, whose infographic are examined. The key stages of big data analytics in Figure 1 are presented.

1) Data Collection: For big data analysis, data can be collected from semi-structured sources such as data warehouses, business transactions, sensors, log files, unstructured sources such as social media platforms, and structured sources such

Doygun Demirol is with the Department of Computer Technologies, Bingol University, Bingol, 12000 Türkiye. Email: ddemirol@bingol.edu.tr

Resul Daş is with the Department of Software Engineering, Technology Faculty, Firat University, Elazig, 23119 Türkiye. Email: rdas@firat.edu.tr

Mehmet Özdem Türk Telekom, Ankara, Türkiye. Email: mehmet.ozdem@turktelekom.com.tr

Ceren Nur Cansel is with the Industrial Engineering, College of Engineering, Koç University, Istanbul, 34450 Türkiye. Email: ccansel19@ku.edu.tr

Davut Hanbay is with the Department of Computer Engineering, Faculty of Engineering, Inonu University, Malatya, 44000 Türkiye. Email: davut.hanbay@inonu.edu.tr

Manuscript received Jan 27, 2025; accepted March 5, 2025.
DOI: 10.17694/bajece.1625208

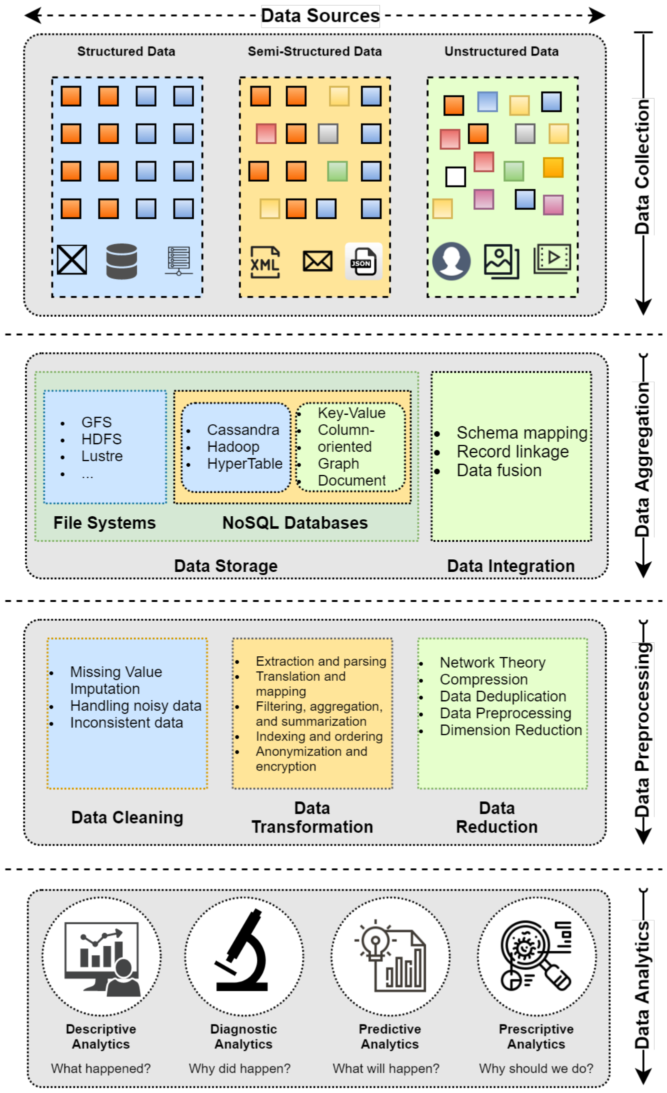


Fig. 1. Key stages of the big data analysis.

as multimedia resources. In this step, obtaining reliable data is critical for the analysis process.

2) Data Aggregation: The accuracy of useful information obtained from data analysis depends on the sufficient amount of data used in the analysis and the accuracy of high-quality data. Data aggregation is a crucial step in big data analytics, as the transform of different data collected from multiple sources into a more readable and analyzable format in the data collection step directly affects the quality and accuracy of the data to be used in the analysis.

3) Data Preprocessing: Due to the wide variety of data sources, undesirable factors may negatively affect data quality, such as noise, inconsistency, and missing information in the collected datasets. Storing these meaningless data is unnecessary both in terms of storage cost and analysis quality. Therefore, data must be preprocessed to perform effective data analysis. Pre-processing of data not only increases storage costs but also increases analysis accuracy [3].

4) Data Analytics: Analyzing big data sets directly without realizing other phases means time-consuming navigation for users across a very large search area. In addition, to analyzing big data quickly, it is also very important to analyze the data on time and to use the right type of analysis. Four types of data analytics are taxonomied for big data analysis: Descriptive, Diagnostic, Predictive, and Prescriptive.

B. Big Data Storage Technologies

Sensor networks, scientific experiments, websites, and many other applications generate semi-structured or unstructured data in various forms [4]. The transition from structured data to unstructured data renders traditional databases useless for storage [5]. This disadvantage of traditional databases contributes to the development of efficient and robust storage systems, and many new technologies are being developed. Ensuring high scalability, reliability, robust, and efficient storage for rapid growth data is the main goal of technologies designed for Big Data storage [6].

This section briefly describes, categorizes, and compares storage technologies developed for Big Data.

1) Column-Based Databases:

- **BigTable**: BigTable is a flexible, reliable and applicable storage system developed by Google for managing petabyte-scale data on thousands of machines to provide high storage capacity for huge amounts of structural data. It also provides dynamic control over data positioning, viewing, and clustering [7].
- **HBase**: HBase is a column-based data storage technology developed by Apache and designed to address the storage requirements of Big Data of the Apache project [8]. It uses HDFS storage technology to store table data and a clustering approach for data management. Therefore, it is a potential solution for projects that contain many rows of data.
- **Hypertable**: Hypertable is a distributed database that provides support for the consistency of stored data. It is designed to work with many distributed file systems such as GFS, HDFS, and CloudStore. It stores the data in tabular form and manages the tables by splitting them to ensure scalability with distribution.
- **Cassandra**: Cassandra is a decentralized, highly accessible, and structured key-value storage system [9]. It provides scalable storage and improved performance for applications that perform intensive reading and writing operations. Fault tolerance, reduced latency, scalability, clustering, segmentation and replication are the highlights of Cassandra [10]. The social media platform Facebook where scalability depends linearly on the number of users and requires intensive reading and writing operations, uses Cassandra for messaging applications [11]

2) Document-Based Databases:

- **MongoDB**: MongoDB, an open source and free to use under the GNU AGPL license is a document-based NoSQL database that stores each entry in JSON documents. It has a query language that uses JSON structure.

It is also named a grid-based file system (GridFS), which allows storage of large objects [12].

- **Terrastore**: Terrastore is a document-based, open-source, distributed storage system developed by Terracotta Inc. hlthat provides scalability, consistency, and dynamic clustering support while running. It automatically distributes data through server nodes and has the ability to redistribute data automatically when new servers are included or servers are removed. It also has replication and switching as backup system features [13].
- **DynamoDB**: DynamoDB is a distributed, schemaless NoSQL database technology used by Amazon [14]. It is widely used to store unstructured, mutable and scalable data. It provides infinite capacity for data storage and access rate. It offers updating, efficient indexing, adaptability and scalability in distributed systems. Therefore, it is suitable for rapidly growing data and scalable applications [15].
- **RethinkDB**: RethinkDB is the first open-source document-based database that efficiently supports complex queries for real-time web applications [16]. It successfully responds to real-time requests from users and also uses the JSON data structure.
- **OrientDB**: OrientDB is the first multi-model, open source, and highly scalable database technology for document data with an extended and transparently managed graph layer that provides relationships between records. Thanks to the embedded graphic layer, data circulation and data relationship management can be implemented quickly and without increasing costs. OrientDB, supports all operating systems, has features such as clustering and replication on heterogeneous servers to ensure performance and scalability [17].
- **CouchDB**: NoSQL document-based database using JSON, JavaScript for MapReduce queries, and HTTP for its API, known for replication and offline capabilities.

3) Graph-Based Databases:

- **HyperGraphDB**: HyperGraphDB is an open source graphical database system designed for artificial intelligence and web semantics projects [18].The storage infrastructure is provided by Berkeley DB. Also, it has a data mapping system among storage systems and hosts. It uses a key-value mechanism to store graph information [19]. Unlike other master-slave storage technologies, HyperGraphDB uses peer-to-peer data distribution architecture. It has features such as being able to run each peer independently and performing updates asynchronously [18].
- **Neo4j**: Highly popular, open-source, providing ACID(Atomicity, Consistency, Isolation, Durability) compliance, widely used in social networks, fraud detection, and network management.

4) Key-Value Stores:

- **Aerospike**: Aerospike is an open-source data storage technology for real-time data that offers scalability and reliability at low cost. It has a ”share-nothing” architecture that supports petabyte-scale data volumes with reliability and scalability.
- **Redis**: In-memory key-value store known for high performance and support for multiple data structures, often used for caching, real-time analytics, and message brokering.

5) Time-Series Databases:

- **InfluxDB**: Open-source time-series database designed for high write and query loads, used in monitoring and real-time analytics.
- **OpenTSDB**: Scalable, open-source time-series database built on top of HBase, optimized for storing and serving large amounts of time-series data.
- **Prometheus**: Open-source system monitoring and alerting toolkit, designed for reliability and simplicity.
- **TimescaleDB**: Open-source time-series database optimized for fast ingest and complex queries, built as an extension to PostgreSQL.

6) NewSQL Databases:

- **Google Spanner**: Globally distributed, strongly consistent database service by Google, offering the scalability of NoSQL with the transactional consistency of SQL.
- **CockroachDB**: Open-source, distributed SQL database designed for high availability and horizontal scalability.
- **VoltDB**: High-speed, in-memory, NewSQL database designed for real-time analytics and decision-making.
- **NuoDB**: Distributed SQL database providing horizontal scalability and continuous availability.

7) Object-Oriented Databases:

- **db4o**: Open-source object database for Java and .NET developers, allowing the storage and retrieval of complex data types.
- **ObjectDB**: High-performance object database for Java, providing a simple and intuitive API.
- **InterSystems Caché**: High-performance, multi-model database that supports object and relational data models.

8) Multi-Model Databases:

- **ArangoDB**: Native multi-model database supporting graph, document, and key-value data models.
- **OrientDB**: OrientDB is the first multi-model, open source, and highly scalable database technology for document data with an extended and transparently managed graph layer that provides relationships between records. Thanks to the embedded graphic layer, data circulation and data relationship management can be implemented quickly and without increasing costs. OrientDB, supports all operating systems, has features such as clustering and replication on heterogeneous servers to ensure performance and scalability [17].
- **MarkLogic**: Enterprise NoSQL database with multi-model capabilities, providing a single platform for data integration.
- **Couchbase**: Distributed NoSQL cloud database supporting multiple data models including document, key-value, and SQL for JSON.

9) Distributed File Systems:

- **Google File System (GFS)**: Proprietary distributed file system developed by Google to provide efficient, reliable access to data using large clusters of commodity hardware. Google File System (GFS) is a storage system developed by Google Inc. to manage data-intensive applications. Along with its innovative features, it is designed to meet the ever-increasing data storage requirements. For large files, GFS guarantees that writing new data simultaneously, even when reading and writing are intensive [20].

- **Hadoop Distributed File System (HDFS)**: Open-source distributed file system designed to run on commodity hardware, part of the Apache Hadoop project, and used for reliable, scalable, and distributed storage. Hadoop Distributed File System (HDFS) inspired by Google File System and meets the data storage requirements of Hadoop applications for Map-Reduce [21]. It is a suitable solution that works with high efficiency in Big Data for data-intensive applications. Although it contains a large number of components, it has fast error detection and automatic recovery features.

C. Big Data Applications

Today, there are many business and industrial areas where big data analysis is applied. These sectors generate large amounts of Big Data that must be analysed for effective and efficient decision-making processes. Some of these areas of application are shown in Fig. 2 and are explained below.

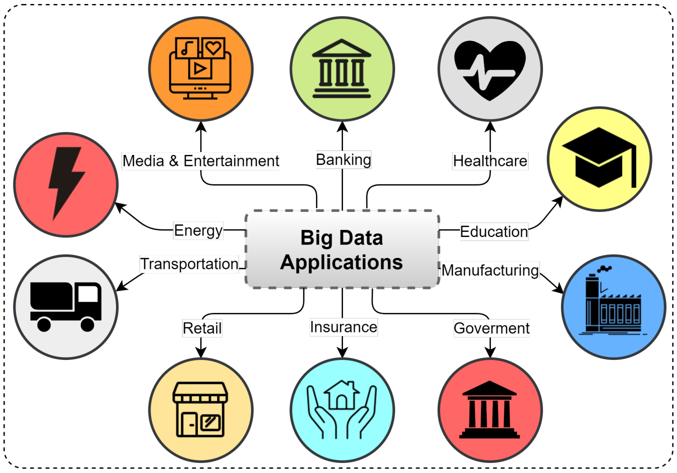


Fig. 2. Important applications fields of the big data

1) Health Applications: Big Data is a part of the healthcare industry, as it is in many other industries, and is actively used in healthcare applications. Big Data are generated from heterogeneous sources such as laboratory records, patient symptoms, devices used, and patient health record data in healthcare applications. These data are crucial to facilitating and improving decision-making for both physicians and other healthcare employees. Advanced analysis of centralized medical data provides great benefits to healthcare. For example, doctors can monitor patients' symptoms and prescribe the medication in real time [22].

Today, many applications are developed with the analysis of big data gathering from the healthcare sector. The analysis of Big Data plays a vital role in predicting the outbreak of diseases such as the Ebola virus, using patient records and sensor data to improve the quality of healthcare, and providing a feedback mechanism [23], analyzing of genome [24]. Applications for the Covid 19 virus epidemic, which is still having its effect, are also developed [25]. Through these applications, Big Data is in many steps from the identification of infected cases to analyzing the risks, storing the travel history of the people and identifying the people who may be in contact with the infected patient, keeping the data of the patients' fever and other symptoms, whether medical intervention is required, and the detection of the virus at early stages, are used [26].

2) Recommendation Systems: Recommendation systems play an important role in our daily life. These systems provide users with recommendations based on their interests. YouTube, for example, analyzes the user's Watch history and predicts in which category to watch videos in the future, and offers video suggestions accordingly. Similarly, e-commerce applications analyze users' product search and review histories and offer product recommendations to users.

3) Smart Cities: Smart Cities is a concept that encompasses the economy, governance, mobility, people, environment and life [27]. In order to increase the quality, performance, and interaction of urban services, it is necessary to use information technology in smart cities. Big Data technologies are used in smart cities in a variety of fields, including traffic statistics, smart agriculture, healthcare, transportation, and some other purposes [28]. In smart cities, data is collected from power poles, water pipes, buses, trains, and traffic lights with sensors on them [29]. With the data collected, faults in the specified systems can be prevented in advance or the faults that occur in the future can be fixed. Also, energy can be managed according to needs, and attacks on smart grids or possible attempted attacks can be prevented in advance.

4) Shipping and Logistics: Public transport companies use GPS devices in vehicles to monitor their vehicles and schedule transportation. With big data analysis, data such as using different routes, identification of optimized routes, passenger travel frequencies, and the number of passengers are collected with GPS devices and processed and analyzed with big data analysis methos. Various real-time systems not only provide advice to passengers but also provide useful information about when to wait for the next vehicle that will take them to the right destination. In India, for example, which has one of the largest railway networks in the world, the total number of reserved seats are allocated every day is around 250,000 and can be booked 60 days in advance. It is quite complicated to make predictions on such data due to factors such as weekends, festivals, night trains, starting or intermediate stations, etc. Considering the factors mentioned above, the calculation of the probability of approval of the ticket to be purchased for any train to solve the passengers' ticket purchase decision problem is performed using Big Data analysis methods [30].

5) Social Networks and Media: Social media platforms, which consist of many Internet applications such as social and business-oriented networks, online mobile photo and video sharing services, are another data source that requires real-time data processing. Analyzing and processing of Big Data from these sources requires a range of methods and algorithms such as text analysis, sentiment analysis, information fusion, and network analytics [31]. Many governments use sentiment analysis on social media data to predict election outcomes by monitoring political trends.Such systems, where a lot of social media data is aggregated and used, also provide to identify national and local problems [32].

Apart from the areas mentioned in Section II-C, there are many business sectors and industries that benefit from the possibilities of big data analysis today. These industries that use Big Data produce huge amounts of data that require Big Data analytics for effective and efficient decision making.

D. Privacy of Big Data

People around the world use various web services and networks such as social media, e-government, e-health, etc. Users save personal information such as name, surname, user name, password, contact information, address, phone number, user behavior, habits, political or religious tendencies, and locations in these systems. The use of such services increases risks such as user privacy and data disclosure. Companies use these personal, sensitive data which can be obtained by using various data mining methods to determine customer needs and customer habits and to get more benefits from customers. As a result, hackers can pose a security threat at any phase of the big data life-cycle to compromise data security or privacy.

1) Protecting Privacy of Big Data: Data analysis activities lead to data privacy issues throughout the life cycle of big data. Although there are privacy protection laws in many countries, malicious actors violate the privacy of personal information. For example, most mobile applications require access permission for contact lists, files, cameras, microphones, etc. Users also often accept all terms and conditions without reading the privacy terms. Such permissions may lead to the leakage of vital information about individuals and breach of data privacy. In this section, preservation methods used in the privacy of personal information categorized and summarized under three main headings. These categories are anonymization, encryption, and perturbation methods.

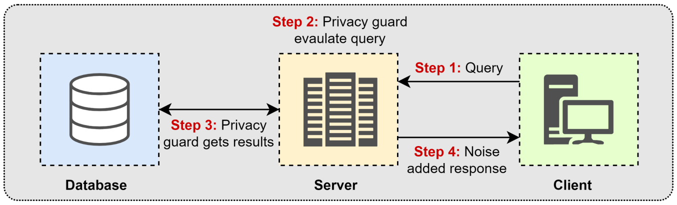


Fig. 3. Differential privacy mechanism.

2) Anonymization Techniques: Anonymization techniques preserve privacy by masking or altering personal, sensitive data. As with most data applications, raw data should first be preprocessed to secure sensitive information. In anonymization approaches, basically, data are represented in four categories. Explicit Identifiers are the set of attributes that directly concern individuals, such as name, surname, and telephone number. The Quasi Identifier (QID) is a set of attributes such as age, gender, postal code, which can identify individuals when combined with other attributes, although not enough by itself to identify individuals. Sensitive attributes (SA) is an attribute set that includes sensitive information about individuals as disease and income information. Non-Sensitive attributes (NSA) include attributes not fall into the other categories. This section discusses K-Anonymity, L-Diversity, T-Closeness, and Differential Privacy privacy-preserving methods [33].

K-Anonymity [34], is developed to hide information specific to individuals that show singular characteristics in a data set. In the K-anonymity algorithm, suppression and generalization methods are used to anonymize the data. In the suppression method, values of categorical variables such as names are completely removed from the data set. With the generalization method, quantitative variables such as age or height are expressed within a range. This makes each record in a data set indistinguishable from other records with at least (k-1). One of the biggest disadvantages of the K-anonymity method is that it is possible to extract identity from the data set if certain properties are already known.

Table I shows non-anonymized dataset of the fictional patient records. In suppression method, ”Name” and ”Religion” columns are replaced with ”*” character as its shown in Table II. In generalization method attributies are replaced a specific value with a more generic one. In Table I, values in ”age” column replaced with a wider range. As a result Table I has become 2-anonymity. Any combination of ”Age”, ”Gender” and ”Country” attributes is any row of the Table I that exists at least two rows with these attributes.

TABLE I
NON-ANONYMIZED OF FICTIONAL PATIENT RECORDS [?].

Name	Age	Gender	Country	Religion	Disease
Ramsha	30	Female	Tamil	Hindu	Cancer
Yadu	24	Female	Kerala	Hindu	Viral infection
Salima	28	Female	Tamil	Muslim	Tuberculosis
Sunny	27	Male	Karnataka	Parsi	No illness
Joan	24	Female	Kerala	Christian	Heart-related
Bahuksana	23	Male	Karnataka	Buddhist	Tuberculosis
Rambha	19	Male	Kerala	Hindu	Cancer
Kishor	29	Male	Karnataka	Hindu	Heart-related
Johnson	17	Male	Kerala	Christian	Heart-related
John	19	Male	Kerala	Christian	Viral infection

TABLE III
L-DIVERSITY ORIGINAL DATASET.

Zip Code	**Age**	**Country**	**Disease**
25044	38	Hindu	Heart-related
25059	39	Japan	Heart-related
25059	31	Chinese	Viral infection
25044	33	Japan	Viral infection
26844	60	English	Cancer
26844	65	Hindu	Heart-related
26841	67	Japan	Viral infection
26841	69	Japan	Viral infection
25044	51	Japan	Cancer
25044	57	English	Cancer
25059	56	Turkish	Cancer
25059	55	Japan	Cancer

TABLE IV
K=4 ANONYMIZATION APPLIED DATASET.

Zip Code	**Age**	**Country**	**Disease**
250**	< 40	*	Heart-related
250**	< 40	*	Heart-related
250**	< 40	*	Viral infection
250**	< 40	*	Viral infection
268**	≥ 60	*	Cancer
268**	≥ 60	*	Heart-related
268**	≥ 60	*	Viral infection
250**	≥ 60	*	Viral infection
250**	5*	*	Cancer
250**	5*	*	Cancer
250**	5*	*	Cancer
250**	5*	*	Cancer

TABLE II
ANONYMIZED METHODS APPLIED RECORDS [?].

Name	**Age**	**Gender**	**Country**	**Religion**	**Disease**
*	$20 < \text{Age} \leq 30$	Female	Tamil	*	Cancer
*	$20 < \text{Age} \leq 30$	Female	Kerala	*	Viral infection
*	$20 < \text{Age} \leq 30$	Female	Tamil	*	Tuberculosis
*	$20 < \text{Age} \leq 30$	Male	Karnataka	*	No illness
*	$20 < \text{Age} \leq 30$	Female	Kerala	*	Heart-related
*	$20 < \text{Age} \leq 30$	Male	Karnataka	*	Tuberculosis
*	$\text{Age} \leq 20$	Male	Kerala	*	Cancer
*	$20 < \text{Age} \leq 30$	Male	Karnataka	*	Heart-related
*	$\text{Age} \leq 20$	Male	Kerala	*	Heart-related
*	$\text{Age} \leq 20$	Male	Kerala	*	Viral infection

TABLE V
K=4 ANONYMIZATION AND L=3 DIVERSITY APPLIED DATASET.

Zip Code	**Age**	**Country**	**Disease**
2505*	≤ 60	*	Viral infection
2505*	≤ 60	*	Heart-related
2505*	≤ 60	*	Cancer
250**	≤ 60	*	Cancer
2685*	> 60	*	Cancer
2585*	> 60	*	Heart-related
2685*	> 60	*	Viral infection
250**	> 60	*	Viral infection
2506*	≤ 60	*	Heart-related
2504*	≤ 60	*	Viral infection
2506*	≤ 60	*	Cancer
250**	≤ 60	*	Cancer

L-Diversity [35] is a method developed to overcome the disadvantages of k-anonymity. L-Diversity model is developed to diversify SA attributes of k records in QID groups created in the k-anonymity approach. If SA attributes in QID groups are combined with other information, it is possible to extract sensitive information. To prevent this problem, the L-Diversity model is developed by diversifying the ”l” piece of the SA attributes. In Table III, patients’ private attributes(name, gender etc.) are hidden to apply anonymization method to avoid leak. But other attributes are given like zip code, age, country, and disease. So, adversaries may be possible to detect patients’ private information. In Table IV, k=4 anonymity method applied to avoid this problem. However, if the last 4 records are examined, it is seen that everyone whose zip code starts with 250 and who is in their 50s has cancer. If a malicious person has this information and additional information, it may be possible to guess that any of the patients has cancer. For this reason, the confidentiality of personal information is ensured by the l-diversity method, which is applied by creating 3 variations within each group of 4 in Table V.

T-Closeness [36] model is proposed by Li et al. because k-anonymity and l-Diversity approaches could not provide enough protection. The proposed model guarantees that the semantic closeness of SA attributes in each QID group is t-closeness to each other also in SA attributes as in the QID group. In other words, the T-Closeness approach argues that the distribution of SA attributes in the data set should also be preserved in QID groups [37].

Differential Privacy [38] minimizes the chance of personal information being identified. It also maximizes the accuracy of queries from databases. With this technique, researchers can make queries from databases containing personal information and protect the privacy of personal information. As seen in Fig. 3, in the first step, the client sends a query. In the second step, the privacy protection tool between the client and the Database evaluates this query for risk. In the third step, the privacy protection tool retrieves the results from the database, and in the fourth step, the server sends the noise added results to the client. The amount of noise appended to pure data is directly related to the assessed privacy risk. If the risk of data privacy is low, the added noise is too small to influence the quality of the result set. However, in the opposite case, the amount of noise to be added is large enough to protect individual privacy.

3) Data Perturbation Techniques: Data perturbation techniques protect the privacy of personal data by applying systematic changes to the records in the database. Data perturbation can be defined in three categories as input, output, and algorithm perturbation according to the phases to which noise is added. The phase where noise is added before performing any computation is called input and the phase where noise is added to an algorithm result is called algorithm, and the phase where the noise is added directly to the data during the algorithm and computations is called the output. Data perturbation techniques apply certain changes to the original data, such as adding randomization and noise to protect privacy. *Randomization* is a technique for privacy-preserving data generally defined as adding noise to the data to mask the attribute values of records with the probability distribution [39]. This technique is generally used in surveys, emotion analysis, etc. Records in the data set are evaluated independently from each other. The method can be implemented during data collection and preprocessing. It is stated in [40], that the randomization technique is proved to be not suitable for large data sets due to its time complexity and data utility. These techniques are known to have lower computational complexity than cryptographic methods to protect privacy. The dataset to which the technique is implemented is generally not distinguishable from the original dataset [41]. The most major disadvantage of data perturbation techniques is that they cannot process high volumes of data effectively. This technique provides sufficient privacy on the data, but it takes a significant amount of time to get excellent results [42], [43].

4) Cryptographic Techniques: In the context of big data, the most crucial issue to preserve the privacy of data is to ensure communication security. Encryption methods meet communication security to a great extent [44]. However, it is very difficult to encrypt large-scale data using traditional encryption techniques. So, privacy protection methods based on data encryption are mostly used for distributed applications. Ensuring the security of sensitive private data in the cloud and other big data storage platforms is generally possible today with methods such as homomorphic encryption, attribute-based encryption, and image encryption. Although homomorphic encryption allows some operations without decrypting the encrypted data, the computational efficiency and scalability of this encryption method need improvement to handle large data [45], [46]. In order to address the complex access control mechanism and fine-grained data sharing issue over encrypted data, Sahai et al. [47] introduced an attribute-based encryption (ABE) concept. Next, ABE is categorized into Key-Policy ABE (KP-ABE) [47] and the Ciphertext Policy ABE (CP-ABE) [48]. KP-ABE technique uses identifier attributes to encrypt data and can be accessed if the user's key contains an access policy that includes which user is allowed to access the data [49]. In the CP-ABE technique, users' secret keys are associated with their attributes, access control policies are defined by the data owner, and ciphertext is generated under the access structure [50]. An image encryption algorithm transforms users' private images into texture-like or noise-like encrypted images, namely, incomprehensible form. In this way, encrypted images become withstand various attacks by hackers. So, they can not obtain the original image without the correct secret keys.

III. Literature Review

Anomaly detection and data integrity in IoT systems are of critical importance in the field of cyber security. In this context, the integration of machine learning and deep learning-based methods with blockchain technology offers promising approaches to improve the security of IoT networks. There are various studies in this direction in the literature. In this study, the literature is reviewed and categorised under three headings: machine and deep learning based anomaly detection, ensuring data integrity with blockchain technology and integration of machine learning, deep learning and blockchain.

Anomaly detection in IoT networks plays a critical role in terms of security and performance of the systems. For this purpose, various machine learning and deep learning models are used [51]. Satılmış et al. [52] investigated machine learning and deep learning models proposed to detect anomaly-based attacks in IoT networks. In their study, the advantages and disadvantages of the models used are discussed. Gokdemir et al. [53] investigated the effectiveness of deep learning methods for anomaly detection in IoT time series data. In their study, it was shown that LSTM (Long Short-Term Memory) networks can successfully detect anomalies in time series data. In another study [54], the authors propose a deep learning based model to detect and classify anomalies in IoT networks. The proposed model is based on Residual Networks and Bi-directional GRU architectures and can fully utilise the spatial and temporal characteristics of network traffic data. In addition, the attention mechanism is used to extract key features to improve the performance of the model. Experimental results show that the proposed model performs well in anomaly detection and classification. Ferrag and Maglaras [55] developed a blockchain-based collaborative model for anomaly detection in IoT networks. The study provides a low-cost and highly accurate approach for resource-constrained IoT devices while improving data integrity and inter-device security with blockchain technology. Golomb et al. [56] developed a lightweight framework called CIoTA for anomaly detection in IoT devices. This framework provides a reliable collaboration between IoT devices and a continuously updated anomaly detection model using blockchain technology. The study shows that devices with limited resources can perform anomaly detection securely within a distributed structure. Diro et al. [57] extensively investigated machine learning-based solutions for anomaly detection in IoT networks. While addressing the advantages and limitations of existing algorithms, the study emphasises that blockchain-based anomaly detection systems can be effectively used by integrating with machine learning models. Emec and Özcanhan [58] developed a BLSTM-GRU based hybrid deep learning model for intrusion detection in IoT networks. The model was tested on CIC-IDS-2018 and BoT-IoT datasets and showed superior performance with 98.78% and 99.99% accuracy rates, respectively. Dongxing et al. [59] developed a blockchain-based mechanism for authentication and security of IoT devices. In the study, a unique ID is created for each device and stored on the blockchain, eliminating the dependency on centralised authorities. The proposed system has been tested using the Hyperledger Fabric platform and shown to provide a successful solution for ensuring the data integrity of IoT devices. In another study [60] developed a blockchain-based data protection framework to secure IoT data in untrusted storage. The proposed framework aims to reduce storage overhead and increase security through lightweight verification structures and smart contracts on the blockchain. Simulation results show that the system effectively reduces the storage overhead on the blockchain. Türker et all. [61] employs blockchain technology to secure data obtained from IoT devices in smart home systems. Blockchain is implemented as a distributed database for data privacy and integrity, with results indicating preserved data integrity and no data loss. This paper [62] explores the potential of blockchain technology to address critical security and trust challenges within the rapidly growing IoT ecosystem. It examines how blockchain's decentralized, immutable, and transparent features contribute to enhancing security and trust in IoT networks. This study [63] proposes a scalable blockchain-based framework for efficient data management in IoT networks accommodating a large number of devices. By utilizing the Delegated Proof of Stake (DPoS) consensus algorithm, performance and efficiency in resource-constrained IoT networks are enhanced. This paper [64] explores the pivotal role of artificial intelligence in enhancing network security and privacy within blockchain-enabled IoT systems. Blockchain provides a decentralized and immutable ledger for secure management of device identities and transactions in IoT networks. When combined with artificial intelligence, these systems gain the ability to adaptively respond to new and evolving cyber threats, thereby enhancing the resilience of networks against cyber-attacks.

Literature reviews show that the combination of machine learning, deep learning and blockchain technologies can be useful for anomaly detection and data security in IoT networks. In addition to these studies, it is thought that the proposed system can contribute to anomaly detection and data integrity in IoT network traffic. It is also considered to provide a solution for preventing data manipulation and protecting data confidentiality by using data privacy methods in data transmission. The system has been tested with experiments on the N-BaIoT dataset and is considered as an approach that can address security vulnerabilities in IoT networks.

Unlike previous studies that focus solely on anomaly detection, our approach integrates blockchain to ensure data integrity. This combination enhances security by preventing data manipulation, which is a limitation in traditional IDS approaches.

IV. Proposed System

In this study, a system is proposed to detect anomalies in IoT network traffic and improve data security. The proposed system includes a blockchain-based data storage structure for processing data collected from IoT devices, using a deep learning model for anomaly detection, and securely storing the data.

A. Dataset

In this study, the dataset used to detect anomalies in IoT network traffic is obtained from the N-BaIoT dataset, which contains the network traffic of the Philips_B120N10_Baby_Monitor device and is available on the Internet. The dataset includes normal traffic data of the device and various types of attacks (e.g., Mirai and Gafgyt attacks). In the data preprocessing stage, the raw data is cleaned, scaled and labelled as attack/normal traffic. This was done in order to ensure that the model works correctly in the training and testing processes. The data set is divided into 80% training and 20% testing. The training data is further divided into a subset to evaluate the validation performance of the model. This structure is used to objectively evaluate the performance of the model and increase its generalisation ability.

B. System Architecture

The proposed system consists of two main layers to analyse and securely manage network traffic from IoT devices: anomaly detection and blockchain-based data management. In the first layer, IoT network traffic is analysed using a Fully Connected Neural Network (FNN). The model consists of two hidden layers and contains 128 and 64 neurons, respectively. ReLU activation function and Dropout regularisation are used to increase the generalisation capability of the model. Adam algorithm and binary cross-entropy loss function were used to optimise the model. High accuracy was achieved throughout the training process.

In the blockchain layer, the data analysed by the model is stored by encrypting the hash functions of each data block. This SHA-256 based structure guarantees the reliability of the chain and the immutability of the data. Fernet algorithm was used to encrypt the data and privacy was ensured by this method. This approach made it possible to manage data from IoT networks both securely and transparently.

C. Implementation

The implementation of the proposed system involves a series of steps for the detection of anomalies in IoT network traffic and secure management of data. The dataset used in the study is obtained from the N-BaIoT dataset available on the Internet and the network traffic of the Philips_B120N10_Baby_Monitor device is analysed. The dataset contains both normal traffic and various attack types such as Mirai and Gafgyt. The data was preprocessed to make it suitable for training the model. In this process, normal traffic is labelled as ‘0’ and attack traffic is labelled as ‘1’; unnecessary columns are removed and only features that will facilitate the learning of the model are left. Then, the dataset was split into 80% training and 20% testing, and the training data was further split into a subset for validation. All features were scaled using StandardScaler, which ensured that the model evaluated each input on the same scale.

The model is structured as a fully connected neural network. In the model, there are two hidden layers following the input layer. There are 128 neurons in the first layer and 64 neurons in the second layer and ReLU activation function is used in both layers. These layers are designed to learn the complex relationships between the data and improve classification accuracy. To prevent overfitting, dropout was applied at each fully connected layer. The application of dropout effectively mitigated overfitting, as demonstrated by the high accuracy and stable validation results obtained during training. No significant performance degradation was observed, indicating that the chosen dropout rate was sufficient to maintain model generalization. The output layer is structured with sigmoid activation function for classification of anomalies. The model uses Adam algorithm for optimisation and binary cross-entropy function for loss computation. During the training process, the model was optimised for 10 epochs and the accuracy and loss values were monitored at the end of each epoch. The results show that the training accuracy is 99.11%, the test accuracy is 99.79% and the low test loss (0.0006) values prove the generalisation capacity of the model.

During the simulation process, 10% of the test data was encrypted with the Fernet algorithm at random time intervals and sent to the central server. This method aims to protect data confidentiality. On the server, the data after decryption was analysed by the model. These analysis results integrated into the blockchain structure enabled each data block to be encrypted and stored with SHA-256 hash function. This structure of the blockchain prevented the manipulation of data and made it possible to store it securely. In addition, the transparency and immutability of the data within the chain increased the reliability of the system.

In conclusion, the proposed application effectively uses a deep learning model to detect anomalies in IoT networks, while utilising blockchain technology to ensure data confidentiality and integrity. The high accuracy of the model and the security contributions of blockchain are promising for the usability of the system in IoT networks.

V. Results and Discussion

In this section, the performance of the proposed system is discussed in detail and the accuracy metrics of the model and blockchain-based data management results are evaluated. The integration of the deep learning model that detects anomalies in IoT network traffic and the blockchain-based data storage structure shows that the proposed system offers a powerful solution in terms of both technical and security.

A. Model Performance Metrics

The deep learning model was trained and tested using N-BaIoT dataset. The dataset consisting of normal traffic and various attack types of ‘Philips_B120N10_Baby_Monitor’ device was used in the training of the model. During the training process, the dataset was labelled as ‘0’ for normal traffic and ‘1’ for attack traffic, thus increasing the classification capacity of the model. The training of the model was optimised for 10 epochs and the accuracy and loss values were evaluated at the end of each epoch.

The results show that the model performs well on both training and test data. The training accuracy increased with each epoch and reached 99.11%. In the evaluation results on the test data, the accuracy of the model was recorded as 99.79% and the test loss value was recorded as 0.0006. These results show that the model learnt the data effectively and the classification performance is quite high. Figure 4 shows that the training and validation accuracy of the model shows a parallel increase throughout each epoch, reflecting a stable learning process.

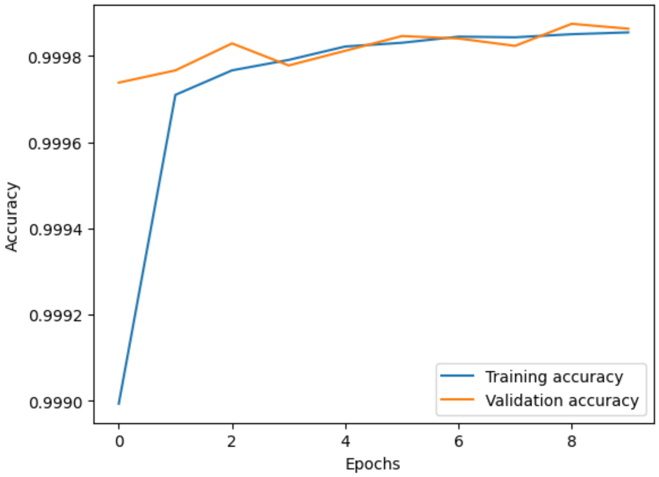


Fig. 4. Train and Validation Accuracy

The loss values decreased steadily during training and validation. As can be seen in Figure 2, the training loss of the model decreased at the end of each epoch and the validation loss also decreased steadily. This shows that the model is optimised and does not overfit the data. This high accuracy and low loss of the model is due to the significant differences between normal traffic and attack traffic in the dataset. In particular, the clear distinction between the characteristics of Mirai and Gafgyt attacks led to the successful classification of the model.

B. Blockchain Data Results

The blockchain-based data storage structure ensured secure storage of data from IoT devices and protection of data integrity. In this system, the data generated by the deep learning model that detects anomalies are added to the blockchain structure in blocks. Each block is associated with the hash function of the previous block, thus ensuring the integrity and security of the chain.

During the simulation process, 10% of the test data was encrypted at random time intervals and transmitted to the central server. The encryption process was performed using the Fernet algorithm and the data was protected against unauthorized access. After the decryption process was completed on the central server, this data was integrated into the blockchain structure. The blockchain system ensured that the data was not only stored with accuracy, but also managed transparently.

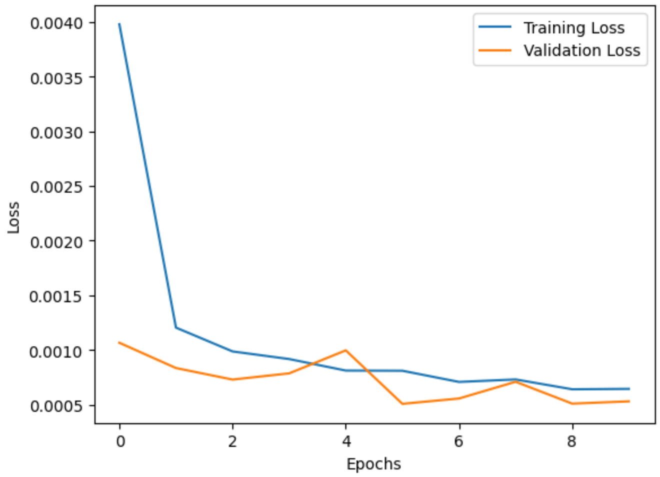


Fig. 5. Train and Validation Loss

As a result, the blockchain-based data storage system guaranteed the accuracy of each block on the chain and made data manipulation impossible. During the simulation, it was observed that data encryption and decryption processes were performed with high performance. Storing data on the chain with hash values proved to be an effective method to secure the large data flow from IoT devices.

The impact of blockchain on processing time, energy consumption, and scalability has been discussed. While blockchain improves data security, it introduces computational overhead, which must be considered in real-time IoT applications. Future studies will explore optimizations for lightweight blockchain implementations.

VI. Conclusion and Future Work

In this study, a system is proposed to detect anomalies in IoT network traffic and ensure data integrity. The proposed system combines a deep learning based anomaly detection model with a blockchain based data storage structure. In this study, network traffic data of Philips_B120N10_Baby_Monitor device is analysed using N-BaIoT dataset. The deep learning model used for anomaly detection provided an effective classification in IoT networks with a high accuracy rate (99.11%) and low loss value. The blockchain structure guarantees data integrity by securely storing and preventing data modification.

The results show that the proposed system provides a powerful solution for the analysis and security of data from IoT devices. Deep learning-based anomaly detection offers the opportunity to prevent potential threats by detecting attacks early in IoT networks. The integration of the blockchain structure not only prevented the manipulation of data, but also provided transparency and traceability. The encryption methods used in the simulation process have increased the effectiveness of the system in terms of data privacy and security.

However, the clear class distinctions in the dataset used in this study played an important role in the high performance of the model. Future work should include testing the proposed system on more complex and noisy IoT datasets to increase its generalisability. Furthermore, analysing data from different IoT devices and the compatibility of the system with these devices is another important issue that needs to be investigated.

Blockchain data structure is an effective solution to ensure data security and integrity in IoT data and critical network infrastructures. With its decentralised and immutable structure, it prevents data manipulation and unauthorized access, but has limitations such as processing speed and energy consumption. In this study, blockchain was used for a relatively limited data flow. However, the performance and energy efficiency of blockchain for large volumes of IoT data should be further investigated in the future. In addition, the integration of more complex encryption algorithms and privacy enhancing technologies can further increase the security level of the system.

The N-BaIoT dataset includes data from nine different IoT devices; however, this study used data from only one device. Future work will expand the evaluation to multiple IoT devices to assess the model's robustness in diverse environments.

In conclusion, this paper presents an innovative approach to provide anomaly detection and data security in IoT network traffic. The integrated use of deep learning and blockchain has made significant progress in IoT data management and security. Future work can enhance the scalability and generalisation capabilities of the proposed system, allowing it to find a wider application in the IoT ecosystem.

References

[1] N. M. Adams, "Perspectives on data mining," *International Journal of Market Research*, vol. 52, no. 1, 2010.

[2] D. Talia, "Clouds for scalable big data analytics," *Computer*, vol. 46, no. 5, pp. 98–101, 2013.

[3] D. Demirol, R. Das, and D. Hanbay, "Büyük veri üzerine perspektif bir bakış," in *2019 International Artificial Intelligence and Data Processing Symposium (IDAP)*, 2019, pp. 1–9.

[4] A. Abouzeid, K. Bajda-Pawlikowski, D. Abadi, A. Silberschatz, and A. Rasin, "HadoopDB: An architectural hybrid of mapreduce and DBMS technologies for analytical workloads," *Proceedings of the VLDB Endowment*, vol. 2, no. 1, pp. 922–933, aug 2009.

[5] V. Subramaniyaswamy, V. Vijayakumar, R. Logesh, and V. Indragandhi, "Unstructured data analysis on big data using map reduce," *Procedia Computer Science*, vol. 50, pp. 456–465, 2015.

[6] S. F. Oliveira, K. Fürlinger, and D. Kranzlmüller, "Trends in computation, communication and storage and the consequences for data-intensive science," in *Proceedings of the 14th IEEE International Conference on High Performance Computing and Communications, HPCC-2012 - 9th IEEE International Conference on Embedded Software and Systems, ICESS-2012*. IEEE, jun 2012, pp. 572–579.

[7] F. Chang, J. Dean, S. Ghemawat, W. C. Hsieh, D. A. Wallach, M. Burrows, T. Chandra, A. Fikes, and R. E. Gruber, "Bigtable: A distributed storage system for structured data," *ACM Transactions on Computer Systems*, vol. 26, no. 2, pp. 1–26, 2008.

[8] L. George, *HBase: The Definitive Guide*. O'Reilly Media, Inc., 2016.

[9] A. Lakshman and P. Malik, "Cassandra - a decentralized structured storage system," *Operating Systems Review (ACM)*, vol. 44, pp. 35–40, 2010.

[10] V. Abramova and J. Bernardino, "Nosql databases: Mongodb vs cassandra," in *C3S2E '13: Proceedings of the International C* Conference on Computer Science and Software Engineering*, 2013, pp. 14–22.

[11] M. Armbrust, A. Fox, D. Patterson, N. Lanham, B. Trushkowsky, J. Trutna, and H. Oh, "Scads: Scale-independent storage for social computing applications," in *CIDR 2009 - 4th Biennial Conference on Innovative Data Systems Research*, January 2009.

[12] K. Pattnaik and B. Mishra, "Introduction to big data analysis," in *Techniques and Environments for Big Data Analysis*, 2016, pp. 1–20.

[13] R. Cattell, "Scalable sql and nosql data stores," *SIGMOD Record*, vol. 39, pp. 12–27, May 2010.

[14] S. Sivasubramanian, "Amazon dynamodb: A seamlessly scalable non-relational database service," in *Proceedings of the 2012 International Conference on Management of Data - SIGMOD '12*, 2012, pp. 729–730.

[15] U. Vyas and P. Kuppusamy, *DynamoDB Applied Design Patterns*. Packt Publishing Ltd., 2014.

[16] R. Paul, "An introduction to building realtime apps with rethinkdb," March 2018, accessed: 2025-01-15. [Online]. Available: https://jaxenter.com/building-realtime-apps-rethinkdb-115254.html

[17] OrientDB, "Orientdb nosql models," 2021. [Online]. Available: http://orientdb.com/docs/3.0.x/gettingstarted/

[18] B. Iordanov, "Hypergraphdb: A generalized graph database," in *Web-Age Information Management. WAIM 2010 Workshops*, ser. 6185 LNCS, 2010, pp. 25–36.

[19] D. Dominguez-Sal, P. Urbón-Bayes, A. Giménez-Vañó, S. Gómez-Villamor, N. Martínez-Bazán, and J. Larriba-Pey, "Survey of graph database performance on the hpc scalable graph analysis benchmark," in *Web-Age Information Management. WAIM 2010 Workshops*, ser. 6185 LNCS, 2010, pp. 37–48.

[20] S. Ghemawat, H. Gobioff, and S. Leung, "The google file system," *SIGOPS Oper. Syst. Rev.*, vol. 37, pp. 29–43, 2003.

[21] N. Gemayel, "Analyzing google file system and hadoop distributed file system," *Research Journal of Information Technology*, vol. 8, pp. 66–74, 2016.

[22] O. Kisi, J. Shiri, S. Karimi, and R. M. Adnan, *Big Data in Engineering Applications*. Springer Singapore, May 2018, vol. 44.

[23] E. J. Khatib, R. Barco, P. Munoz, I. D. La Bandera, and I. Serrano, "Self-healing in mobile networks with big data," *IEEE Communications Magazine*, vol. 54, no. 1, pp. 114–120, jan 2016.

[24] B. Das, "A deep learning model for identification of diabetes type 2 based on nucleotide signals," *Neural Computing and Applications*, vol. 22, no. 1, pp. 1–5, 2022.

[25] Q. V. Pham, D. C. Nguyen, T. Huynh-The, W. J. Hwang, and P. N. Pathirana, "Artificial Intelligence (AI) and Big Data for Coronavirus (COVID-19) Pandemic: A Survey on the State-of-the-Arts," *IEEE Access*, vol. 8, pp. 130 820–130 839, 2020.

[26] A. Haleem, M. Javaid, I. H. Khan, and R. Vaishya, "Significant Applications of Big Data in COVID-19 Pandemic," *Indian Journal of Orthopaedics*, vol. 54, no. 4, pp. 526–528, jul 2020.

[27] J. H. Lee, R. Phaal, and S. H. Lee, "An integrated service-device-technology roadmap for smart city development," *Technological Forecasting and Social Change*, vol. 80, no. 2, pp. 286–306, feb 2013.

[28] C. L. Stimmel, *Building smart cities: Analytics, ICT, and design thinking*. Auerbach Publications, aug 2015.

[29] C. T. Yin, Z. Xiong, H. Chen, J. Y. Wang, D. Cooper, and B. David, "A literature survey on smart cities," *Science China Information Sciences*, vol. 58, no. 10, pp. 1–18, oct 2015.

[30] S. R., P. (2015, aug) 8 innovative examples of big data usage in india. [Online]. Available: https://www.dqindia.com/8-innovative-examples-of-big-data-usage-in-india/

[31] M. Assefi, E. Behravesh, G. Liu, and A. P. Tafti, "Big data machine learning using apache spark MLlib," in *Proceedings - 2017 IEEE International Conference on Big Data, Big Data 2017*, vol. 2018-January. IEEE, dec 2017, pp. 3492–3498.

[32] A. Oussous, F. Z. Benjelloun, A. Ait Lahcen, and S. Belfkih, "Big Data technologies: A survey," *Journal of King Saud University - Computer and Information Sciences*, vol. 30, no. 4, pp. 431–448, oct 2018.

[33] W. Fang, X. Z. Wen, Y. Zheng, and M. Zhou, "A Survey of Big Data Security and Privacy Preserving," *IETE Technical Review (Institution of Electronics and Telecommunication Engineers, India)*, vol. 34, no. 5, pp. 544–560, sep 2017.

[34] P. Samarati and L. Sweeney, "Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression," *Technical Report, SRI International Computer Science Laboratory*, 1 1998.

[35] A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkitasubramaniam, "l-diversity: Privacy beyond k-anonymity," *ACM Transactions on Knowledge Discovery from Data*, vol. 1, no. 1, p. 3, mar 2007.

[36] L. Ninghui, L. Tiancheng, and S. Venkatasubramanian, "t-Closeness: Privacy beyond k-anonymity and l-diversity," in *Proceedings - International Conference on Data Engineering*. IEEE, apr 2007, pp. 106–115.

[37] H. Y. Tran and J. Hu, "Privacy-preserving big data analytics a comprehensive survey," *Journal of Parallel and Distributed Computing*, vol. 134, pp. 207–218, dec 2019.

[38] C. Dwork, "Differential privacy," in *Lecture Notes in Computer Science*, 2006, vol. 4052 LNCS, pp. 1–12.

[39] C. C. Aggarwal and P. S. Yu, *A General Survey of Privacy-Preserving Data Mining Models and Algorithms*. Boston, MA: Springer US, 2008, pp. 11–52.

[40] P. Ram Mohan Rao, S. Murali Krishna, and A. P. Siva Kumar, "Privacy preservation techniques in big data analytics: a survey," *Journal of Big Data*, vol. 5, no. 1, 2018.

[41] M. A. Chamikara, P. Bertok, D. Liu, S. Camtepe, and I. Khalil, "Efficient data perturbation for privacy preserving and accurate data stream mining," *Pervasive and Mobile Computing*, vol. 48, pp. 1–19, aug 2018.

[42] K. Chen and L. Liu, "A random rotation perturbation approach to privacy preserving data classification," *International Conference on Data Mining*, 2005.

[43] Chen, Keke and Liu, Ling, "Geometric data perturbation for privacy preserving outsourced data mining," *Knowledge and Information Systems*, vol. 29, no. 3, pp. 657–695, dec 2011.

[44] M. Z. Gündüz, D. Demirol, R. Daş, and K. Hanbay, "Frameworks for smart grid cyber security analysis," in *Cyber Security Solutions for Protecting and Building the Future Smart Grid*, D. Asija, R. Viral, R. Daş, and G. Tuna, Eds. Elsevier, 2025, pp. 191–214.

[45] S. Venkatraman and R. Venkatraman, "Big data security challenges and strategies," *AIMS Mathematics*, vol. 4, no. 3, pp. 860–879, 2019.

[46] M. Zhao E and Y. Geng, "Homomorphic Encryption Technology for Cloud Computing," *Procedia Computer Science*, vol. 154, pp. 73–83, 2019.

[47] A. Sahai and B. Waters, "Fuzzy identity-based encryption," in *Advances in Cryptology – EUROCRYPT 2005*, R. Cramer, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 457–473.

[48] J. Bethencourt, A. Sahai, and B. Waters, "Ciphertext-Policy Attribute-Based Encryption," in *2007 IEEE Symposium on Security and Privacy (SP '07)*. IEEE, may 2007, pp. 321–334.

[49] T. Bouabana-Tebibel and A. Kaci, "Parallel search over encrypted data under attribute based encryption on the Cloud Computing," *Computers & Security*, vol. 54, pp. 77–91, oct 2015.

[50] M. Ali, J. Mohajeri, M.-R. Sadeghi, and X. Liu, "A fully distributed hierarchical attribute-based encryption scheme," *Theoretical Computer Science*, vol. 815, pp. 25–46, may 2020.

[51] M. Z. Gunduz and R. Das, "Cyber-security on smart grid: Threats and potential solutions," *Computer Networks*, vol. 169, p. 107094, 2020.

[52] H. Satılmış and S. Akleylek, "Iot güvenliği İçin kullanılan makine Öğrenimi ve derin Öğrenme modelleri Üzerine bir derleme," *Bilişim Teknolojileri Dergisi*, vol. 14, no. 4, p. 457–481, 2021.

[53] A. Gökdemr and A. Çalhan, "Nesnelerin interneti ortamlarında derin öğrenme ve makine öğrenmesi tabanlı anomali tespiti," *Gazi Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi*, vol. 37, no. 4, p. 1945–1956, 2022.

[54] L. E. Dai, X. Wang, and S. B. Xu, "A deep learning based anomaly detection model for iot networks," in *Proceedings of the 2nd International Conference on Internet of Things, Communication and Intelligent Technology*, J. Dong, L. Zhang, and D. Cheng, Eds. Singapore: Springer Nature Singapore, 2024, pp. 187–196.

[55] Y. Mirsky, T. Golomb, and Y. Elovici, "Lightweight collaborative anomaly detection for the iot using blockchain," *Journal of Parallel and Distributed Computing*, vol. 145, pp. 75–97, 2020.

[56] T. Golomb, Y. Mirsky, and Y. Elovici, "Ciota: Collaborative iot anomaly detection via blockchain," *CoRR*, vol. abs/1803.03807, 2018. [Online]. Available: http://arxiv.org/abs/1803.03807

[57] A. Diro, N. Chilamkurti, V.-D. Nguyen, and W. Heyne, "A comprehensive study of anomaly detection schemes in iot networks using machine learning algorithms," *Sensors*, vol. 21, no. 24, 2021.

[58] M. Emeç and M. H. Özcanhan, "A hybrid deep learning approach for intrusion detection in iot networks," *Advances in Electrical and Computer Engineering*, vol. 22, no. 1, pp. 3–12, 2022.

[59] D. Li, W. Peng, W. Deng, and F. Gai, "A blockchain-based authentication and security mechanism for iot," in *2018 27th International Conference on Computer Communication and Networks (ICCCN)*, 2018, pp. 1–6.

[60] T. Xu, Z. Fu, M. Yu, J. Wang, H. Liu, and T. Qiu, "Blockchain based data protection framework for iot in untrusted storage," in *2021 IEEE 24th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, 2021, pp. 813–818.

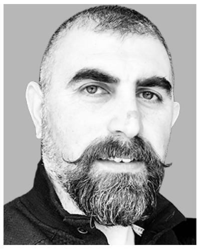
[61] G. F. Türker and K. Tanyeri, "Blokzincir teknolojisi ile nesnelerin İnterneti tabanlı (iot) sistemlerin veri güvenliğinin sağlanması," *Gazi Üniv. Fen Bilim. Derg. C Tasar. ve Teknol.*, May 2023.

[62] S. Almarri and A. Aljughaiman, "Blockchain technology for iot security and trust: A comprehensive slr," *Sustainability*, vol. 16, no. 23, 2024. [Online]. Available: https://www.mdpi.com/2071-1050/16/23/10177

[63] E. U. Haque, A. Shah, J. Iqbal, S. S. Ullah, R. Alroobaea, and S. Hussain, "A scalable blockchain based framework for efficient IoT data management using lightweight consensus," *Sci. Rep.*, vol. 14, no. 1, p. 7841, Apr. 2024.

[64] A. M. Ruzbahani, "Ai-protected blockchain-based iot environments: Harnessing the future of network security and privacy," 2024. [Online]. Available: https://arxiv.org/abs/2405.13847

Doygun Demirol received his Bachelor's and Master's degrees in Electronics and Computer Education from Firat University in 2010 and 2012, respectively. He is currently pursuing his Ph.D. in Computer Engineering at Inonu University. He works as a lecturer at Bingol University. His research interests include cybersecurity, deep learning, natural language processing, and graph data science.



Resul Das is a full professor in the Department of Software Engineering at the Faculty of Technology, Fırat University. He earned his B.Sc. and M.Sc. degrees in Computer Science from Fırat University in 1999 and 2002, respectively, and completed his Ph.D. in Electrical and Electronics Engineering in 2008. Between 2000 and 2011, he served as a lecturer in the Department of Informatics and worked as a network and system administrator at the University's IT Center. Since 2002, he has been an instructor in the Cisco Networking Academy Program, teaching CCNA and CCNP courses. From September 2017 to June 2018, he researched as a visiting professor at the University of Alberta, Edmonton, Canada, under the TÜBİTAK-BİDEB 2219 Postdoctoral Research Fellowship program. He served as Head of the Department of Software Engineering from March 2020 to April 2023.

Prof. Das has held editorial roles in several prestigious academic journals. He was an Associate Editor for IEEE Access and the Turkish Journal of Electrical Engineering and Computer Science. Currently, he serves as an Associate Editor for Elsevier journals (Internet of Things, Alexandria Engineering Journal, and Telematics and Informatics Reports), IEEE Open Journal of the Communications Society (OJ-COMS), and the International Journal of Grid and Utility Computing (Inderscience). Globally recognized for his contributions, Prof. Daş has been included in the top 2% of the "World's Most Influential Scientists" list, compiled by Stanford University researchers, for five consecutive years (2019–2023). His research interests encompass computer networks, cybersecurity, IoT and systems engineering, data science and visualization, software quality assurance, and testing.

Ceren Nur Cansel graduated from Malatya Doğa College Science High School as valedictorian in 2019. She graduated from Koç University Industrial Engineering and Economics departments with Vehbi Koç Honour Award in 2024. She has successfully completed three different specialisation programmes: Finance, Financial Engineering and Macroeconomic Policies and Financial Markets. She is fluent in English. After her university education focused on data analysis, modelling and forecasting, she continues his studies in the field of financial analysis and pricing.

Mehmet Özdem received his BS degree from Başkent University, Department of Electrical and Electronics Engineering in 2002, his MS degree from Middle East Technical University, Institute of Informatics in 2007, and his PhD degree from Gazi University, Institute of Science, Department of Electrical and Electronics Engineering in 2021. He has more than 20 years of experience in the Telecommunications industry. He worked in many groups (investment, planning, operation, quality assurance, and architecture teams). He currently works as the Network Architecture & Quality Assurance Director and Head of R&D Centers, following responsibilities in Turk Telekom. Simultaneously, he is Vice President of ITU (International Telecommunication Union) SG12 and QSDG organizations. He is also a board member of the Wireless Broadband Alliance. He lectures part-time at universities and has international certificates such as CCIE, PMP, and ITIL.

Davut Hanbay received his Bachelor's and Master's degrees in the Department of Electronics Education and Department of Electric–Electronic Engineering from Firat University, Elazig, Turkey, in 1999, and 2003 respectively. He completed his Ph.D. at Fırat University, Department of Electrical and Electronics Engineering in 2007. He is currently Professor in the computer engineering of Inonu University. His research interests include computeraided systems, operating systems, deep learning, machine learning, image processing, and signal processing.