Federated Blockchained Supply Chain Management: A CyberSecurity and Privacy Framework

Konstantinos Demertzis[1,2], Lazaros Iliadis[1], Elias Pimenidis[3], Nikolaos Tziritas[4], Maria Koziri[4], Panagiotis Kikiras[4], Michael Tonkin[3]
[1]School of Civil Engineering, Democritus University of Thrace, Kimmeria, Xanthi, Greece,
[2]Department of Physics, International Hellenic University, Kavala Campus, 65404, Greece;
[3]Computer Science and Creative Technologies, University of the West of England, Bristol
[4]Department of Computer Science, University of Thessaly, Lamia, PC 35100, Greece,
kdemertz@fmenr.duth.gr, liliadis@civil.duth.gr, {Elias.Pimenidis[3], Michael2.Tonkin[3]}@uwe.ac.uk, {nitzirit[4]; mkoziri[4]; kikirasp[4]}@uth.gr

Abstract. The complete transformation of the supply chain in a truly integrated and fully automated process, presupposes the continuous and endless collection of digital information from every stage of the production scale. The aim is not only to investigate the current situation, but also the history for every stage of the chain. Given the heterogeneity of the systems involved in the supply chain and the non-institutional interoperability in terms of hardware and software, serious objections arise as to how these systems are digitally secured. An important issue is to ensure privacy and business confidentiality. This paper presents a specialized and technologically up-to-date framework for the protection of digital security, privacy and industrial-business secrecy. At its core is Federated Learning technology, which operates over Blockchain and applies advanced encryption techniques.

Keywords: Blockchain, Meta-Learning, Federated Learning, Cyber-Security, Privacy, Industry 4.0, Supply Chain Management.

1 Introduction

Digital revolution, especially big data and artificial intelligence, offers new opportunities for the full automation of the supply chain as shaped by the Industry 4.0 standard [1]. At the same time, the complexity increases exponentially, as the number of interconnected systems which participate in continuous interconnection services and uninterrupted real-time information exchange, expands. Applications that monitor security incidents and detect digital hazards, receive a continuous unlimited stream of observations from interconnected systems. In the typical case, the latest data is the most important, as there is the concept of aging based on their timing. This data is characterized by high variability, as their characteristics can change drastically and in an unpredictable way over time, changing their typical, normal behavior [2].

In general, the ever-increasing communication, variability and chaotic planning, exposes the supply chain and the industrial environment in general to serious digital risks.

Given the inability of traditional security systems to detect serious threats, the adoption of intelligent solutions is imperative. Intelligent systems, have the ability to transform human knowledge and experience into optimal valid and timely decisions. In the industrial environment, central storage of all historical data is not appropriate. This fact requires either the retraining of the intelligent system on a subset which contains a small percentage of the total observations, or the extraction of knowledge in real time. The prospect of comprehensive retraining creates serious technical glitches, while data-driven detection raises objections to the accuracy and reliability of the methods used. In both cases, the classifiers degrade over time and become incapable of detecting serious threats [3]. The exchange of data that could create more complete classifiers to generalize, also poses risks to the security and privacy protection of sensitive industrial data. In the context of this work, a standard intelligent digital security information system has been developed and proposed, which seeks to fully upgrade the operation of passive intelligent systems.

The target is the development of an adaptive federated auto meta-learning system through blockchain technology, which ensures privacy and industrial secrecy.

2 Proposed Framework

The proposed architecture has three main principles. The sensitive data is not transmitted through communication channels. The data is not stored in a central point of attack and the learning algorithms are constantly updating their predictive power. Specifically, this research introduces an intelligent control mechanism to detect abnormalities in the Industry 4.0 communication network [4]. This is based on the automatic analysis of digital packets of network traffic. In addition, an automated intelligent neural network was developed to monitor and detect abnormalities, to train and update the model with *federal learning*, and to communicate all involved parameters through a distributed blockchain system.

The architecture of the proposed model is the following [5]: When one device wants to communicate with another, then the proposed intelligent mechanism is activated, which implements a network traffic control to detect anomalies. In the first phase, the features of network traffic are exported. They are used as input to a neural network that is automatically developed through the *Neural Architecture Search* technique. In the beginning the model is trained on the host server, based on some initial data. Then the model is encrypted with homomorphic encryption and through blockchain and it is sent to the nodes that will use it. Then the nodes take the model and improve it using their own data at their disposal. The enhanced model encrypts and returns via blockchain to the host server. The best models are aggregated and the weighted average is selected using the *Grid Search Weighted Average Ensemble* method. The obtained final model returns a blockchain medium to the final nodes. Even if the initial data is not appropriate, there is a continuous improvement of the intelligent model so that it can categorize with great accuracy the anomalies in the network’s traffic.

If the traffic is classified as normal, further communication is allowed, while otherwise, it is stopped, and an alarm is sent to the control center for further control of the transaction.

More specifically, the proposed architecture concerns the process of recording, analyzing and visualizing network traffic in the Industry 4.0 standard and the communication technologies that it integrates. The purpose is to identify abnormalities associated with digital attacks. For this reason, a hybrid technique of automated IP flow analysis was used. Its core idea of standardization and operation is based on the architecture of the open-source framework *Stream4Flow*. It uses the *IPFIXCol collector*, the *Kafka* messaging system, the *Apache Spark* and *Elastic Stack.*
The following figure 1, presents an overview of the proposed architecture.

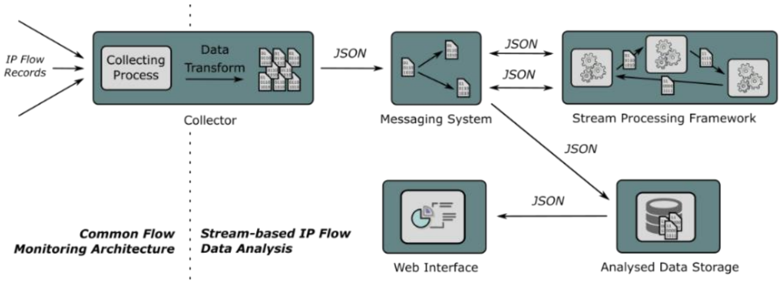

Fig. 1. Stream4Flow Architecture (https://stream4flow.ics.muni.cz/)

IPFIXCol allows the conversion of incoming IP stream records in JSON format provided on the Kafka messaging system. Kafka adds serious scalability and allocation capabilities that provide adequate data performance, and Apache Spark is used as a data processing framework for fast IP data flow. The results of the analysis are stored in an *Elastic Stack* containing *Logstash, Elasticsearch and Kibana*, which allow the results to be stored, searched and visualized. The box also contains an additional web interface to facilitate handling and to visualize the complex results of the analysis [6]. Figure 2 presents the overall architecture of the model.

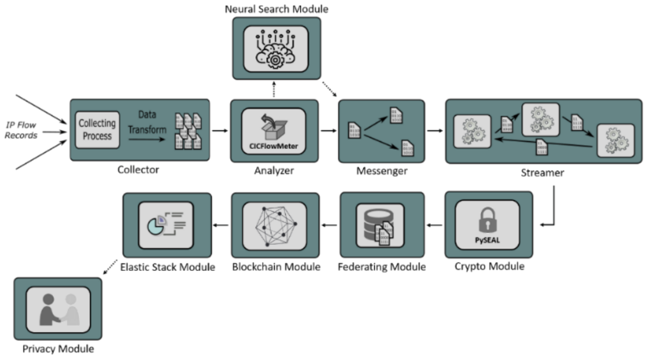

Fig. 2. The proposed architecture.

3 Methodology

It is important to define a clear and consistent description of how to support the operational capabilities of the proposed architecture. Therefore, a detailed scenario will be presented in order to clearly show the use of its mode of operation and the operational requirements that it presupposes, always based on the complexity of the Industry 4.0 ecosystem [7]. This will be achieved, taking into account the basic design principles for identifying and implementing scenarios, i.e., interconnection, access to information, technical support and decentralized decision making. The usage scenario is structured on the idea that a decentralized industry consists of three smart factories each of which has separate special sections such as vertical production and quality control. There is also a decentralized management department and a decentralized department of administrative services and product promotion. To complete the project, industry shares data with a team of scientists where there is a tacit relationship of trust between them, which if violated, can affect the final result of the project and of course it can violate the industrial secrecy [8]. The digital security of the industry infrastructure is controlled by a team of engineering information systems, which uses Machine Learning techniques to protect the infrastructure.

At the same time, it is complying with the general regulation of personal data protection GDPR and seeks to ensure privacy in every phase in which personal information is involved. There is no clear way for stakeholders to work together in a safe and secure way, just as there is no historical or complete record of how they act. Consequently, the users of the IoT environment cannot receive information that may concern them, or they receive it through methods that jeopardize privacy and industrial secrecy [9]. Respectively and in a more technical context, the machine learning algorithms used have to undergo multiple cycles of training, regularization and optimization. Through this process, the teams focus on setting the hyperparameters of the model and on collecting data for its efficient training. This is done with the corresponding cost and by depending on experts on the specific process.

It is obvious that an innovative architecture is required, which standardizes the ways of transactions, while ensuring secure network communication, privacy and industrial secrecy between the traded devices. In Industry 4.0 projects, data communication becomes important as part of the value chain. An IP communication infrastructure for the internal and external parts of the dispersed modern industry is a holistic idea that facilitates operation, expansion and troubleshooting. The communication ability of the individual departments is the basis for the effective distribution and control of the project.

The usage scenario was structured according to the standards of *Industrial Communication Network* [10] IEC 61158 and *Machine to Machine* communication (M2M) OPC UA (IEC 62541), as they are implemented based on the TCP/IP standardization. IEC 61158 is the standardization of a family of industrial network protocols called *Fieldbus* used for real-time distributed control. IEC 62541 is the standardization of OPC UA which is an industrial M2M communication protocol for full interoperability. Its goal is to provide a *Service-Oriented Architecture* (SOA) for process control, while enhancing security and providing an information exchange model.

Alignment of communication between departments based on the above specifications, leads to a higher level of interoperability, stability in processes and rapid troubleshooting with end-to-end monitoring. Based on the decentralized processing of data at the edge of the network (edge computing), it is allowed to perform operations on low cost and on low computing capacity devices, as well as the collection of data in a decentralized way for their use in various systems or applications.

In a test scenario of the proposed architecture, communication (sending, receiving, and accessing data) on sensory instruments or selectors is achieved with the widely used SCADA MODBUS TCP / IP protocol. In conclusion, this model seeks to fully upgrade the mode of operation of passive intelligent systems. In conclusion, this model seeks to fully upgrade the mode of operation of passive intelligent systems, aiming at the implementation of an advanced mechanism, offering fully automatic and personalized solutions, while ensuring the privacy of participants.

The detailed steps of the described usage scenario are as follows:

Step 1 –Collector

The collection was performed using the IPFIXcol framework, in order to edit the SCADA MODBUS TCP/IP streams in the usage scenario.

Step 2 –Analyzer

The network traffic analyzer extracts the most important features that can determine the nature of the information contained in the SCADA MODBUS TCP/IP network traffic, by using the *CIC Flow Meter*.

Step 3 –Neural Search M odule

The Neural Search Module is a holistic approach, which fully automates the creation of neural networks in a way that is not dependent on human intervention, while discovering the optimal hyperparameters, with the aim of detecting abnormalities in network traffic. It is used only once, at the beginning of the process and only for the initial development of the neural network, which is distributed to the devices that implement abnormality.

Step 4 –Messenger

The result of the export of the features performed by the analyzer is transferred to the Messenger, which, mediates the communication between the producer and the consumer using Apache Kafka. Essentially this level of architecture undertakes to route in a clear and secure way the information flows coming from the network communication. This allows partial storage of the data flow of a large number of permanent or ad-hoc consumers and the automatic recovery in case of failure.

Step 5 –Streamer

The Streamer, which has the role of the consumer at the level of Kafka, distinguishes the flows that arise from the Messenger. After determining the time period of each batch, as well as the staging time window, the flow analysis is initiated.

Streamer prepares data batches from feeds, generates RDDs, and sends them for processing at regular intervals. Streamer also undertakes the detailed monitoring of the flow evolution that comes from the Messenger and when a batch is completed, it informs the Spark driver about the flow offsets which determine the batch to be analyzed. The Driver shares the partitions of each topic. Together with the other details that determine the parts of each batch to be processed by each subsystem, they are sending them for encryption in the Crypto Module. Practically only the information of the markers that identify each batch passes through the Streamer and not the data of the flow itself.

Step 6 –*Crypto Module*

Flow encryption is performed in order to create an advanced control mechanism for possible leakage of confidential data and information in the specific architecture. It is achieved with the *Crypto Module*. The idea of its operation is based on the use of an algebraic system that allows authorized third parties to perform a variety of calculations on the encrypted data. In the described scenario, fully homomorphic encryption is used to perform calculations and analysis. The proposed architecture incorporates Microsoft's open-source *Simple Encrypted Arithmetic Library* (SEAL), which uses a uniform encryption system. It is based primarily on the cryptographic schemes of Brakerski / Fan-Vercauteren (BFV) and Cheon, Kim, Kim and Song (CKKS). SEAL initially converts the flow data into polynomials and integrates them into a defined polynomial ring.

The encrypted polynomials are then calculated by applying noisy linear transformations involving the public key for each one. Numerical operations are then performed on the encrypted flow data. If the accumulated noise in the calculations does not exceed a limit, the computational output can be decrypted with a linear transformation that includes the private key. Crypto-Module in addition to the basic encryption device, includes methods that provide optimal parameters for the initial encryption setup and the budget that reflects the noise that occurs during the execution of the given procedure. Thus, the architecture fully ensures privacy and security between the trading devices.

Step 7 – Federating Module

Centralized training is intrusive, as users essentially have to exchange their privacy or sensitive industrial data by sending it to central entities for training. With the federating module, a decentralized training approach is implemented that allows devices located in different geographical locations, to benefit from the acquisition and to make use of a well-trained and continuously upgraded, learning model. It also allows all personal data and information to be kept private, as they do not need to be sent to a central entity to be used as training data. In particular, it allows remote devices to download and operate the original machine learning model, that was initially developed by the *Neural Search Module*. This model runs on the locally available data in order to improve its accuracy and then the data are sent back to the *Federating Module*. The *Dynamic Weighted Average* approach summarizes the changes, creating a new update, which again returns to the end users through the *Blockchain Module*.

End users have constant access to an ever-upgraded neural network model. The various hyperparameters of the network are shared using *Hyperledger Fabric*, where the use of smart contracts ensures unquestionable validation and control of the process between the parties involved.

Step 8 – *The Blockchain Module*

The proposed *Blockchain Module* is based on the *Hyperledger Fabric Project* and its architecture comprises of the following levels:

Consensus Layer - It is responsible for establishing an agreement on the order and on the verification of the correctness of all transactions that constitute a blockchain.

Smart Contract Layer **-** It is responsible for processing transaction requests and approving only valid ones. In the analyzed scenario, the communication of devices which is based on *Smart Contract* considers the following parameters for the contract [5]:

a. *MachineAccount.* This is the account that represents the device in the IIoT ecosystem. Each machine has its own account to sign the transactions it needs to perform.
b. *MachineAddress.* It describes the unique address of each device in the IIoT.
c. *MachineInternals.* It is related to the monitoring of the internal parameters of the device, such as operating temperature, battery status and operating time.
d. *Publisher.* The user or the application that creates the data in the network.
e. *Subscriber.* The user or the application that is using the data in the network.
f. *Sender.* The final user that sends the data that were developed by the Publisher in the network.
g. *Receiver.* The final user that receives the data produced by the Publisher in the network.
h. *MachineStatus.* Procedures to check the status of the device, such as whether it produces data (Publisher), if a procedure is required (Sender).
i. *Session.* The time period in which a transaction takes place.
j. *SessionID.* A unique number characterizing a Session.
k. *DataStream.* The data flow that one device or application wants to transfer to another.
l. *DataStreamID.* A unique number characterizing a DataStream.
m. *TrafficFlow.* A sequence of packets from one source device to a destination, which may be another device or group of devices.
n. *TrafficFlowID.* A unique number characterizing a TrafficFlow.
o. *FeaturesOfTrafficFlow.* Features exported from web traffic. In the case of the scenario under consideration, the characteristics of the network traffic extracted by the analyst (analyzer) using the CICFlowMeter.
p. *IIoT rule.* It Describes a process by which a specific action is tested to see if it can be performed (e.g., checking the necessary rights and the ones involved in the action). For example, if the specific action triggers some additional actions such as the activation of a contract.
q. *SmartContract.* The smart contract that is activated in specific cases of transactions. The SmartContract is assigned a characteristic ID (for example it can have a value equal to ID (101)).

In the performed scenario, the *IIoT_thing_thermostat* (Sender), wants to interact by sending data (DataStream) to the device.
IIoT_thing_water_tank (Receiver). The DataStreams are assigned a *StreamID* e.g., (707), while the specific action is controlled by the IIoT rule on whether the specific transaction can take place. The IIoT rule activates the *SmartContractID*(101). The specific *SmartContractID*(101), activates the *StreamID*(707), which checks if DataStream is characterized as normal or abnormal. If the traffic is declared as normal, then there is a communication with the IIoT_thing_water_tank (Receiver), while otherwise, the communication is rejected, and an alarm is sent for further control.

***Communication Layer* -** It is responsible for the transfer of messages between peer nodes. It provides publish /subscribe functions to system resources and it is directly related to the Authorization layer. For example, a specific resource can be published to a namespace by other entities, receiving a specific URI. Subscribe permission means that a receiver_entity can receive information from the published resource, while publish permission means that this entity can publish - interact with the resource. For example, it can send a restart command, set the operating temperature of the thermostat.

The subscribe permissions are related to the “stats” of the URI, whereas the publish permissions are also related to the “cmd” permissions.

Authentication Layer **-** It is responsible for the validation of users' identities, the control of their rights and for the consolidation of trust in the blockchain. It Provides access levels expressing security policies in the ecosystem IIoT, by using *entities*, *namespaces*, *resources*, and Delegations of Trust (DoTs). An Entity is a standalone control unit that can grant, restore, or delete access rights.

It works as a username or a role, except that a username or a role exists in only one domain, while an entity is global. Anyone can create an entity that is characterized by a key pair (public and private). For example, the admin is an entity with <A_{pk}, A_{vk}> where A_{pk} is the public key and A_{vk} the private key.

A namespace is a domain that contains a hierarchy of resources. The *Raw_Materials_Utility* is a namespace and all resources inside the namespace is reported with a unique URI.

The temperature of a thermostat in Raw_Materials_Utility is described by the URI BW://Raw_Materials_Utility/thermostat/stats/temp, where “stats” is related to data describing the operational status of the resource, whereas “cmd” includes control commands like *restart, lock.* Any other entity interacting with the resources inside a namespace must be licensed by the namespace entity, directly or indirectly. This property is known as the *Delegations of Trust* (DoT). For example, for the DoT=$\langle A_{pk}^{from}, A_{pk}^{to}, URI_{rsrc}, Permissions, Metadata\rangle$ the metadata can have specific properties characterizing the resources.

Overlay layer **-** This is a level added to the *Hyperledger Fabric* architecture to map the communication network between IoT devices. This is responsible for providing the available services and applications, to the network. The nodes at this level, which essentially form an overlapping network over the existing physical one, can be thought of, as connected by virtual or logical links, each of which corresponds to a path in the underlying network.

Blockchain Layer **-** It is the level that creates and manages the blockchain. Its operation can be based on individual specifications related to the content and how to use the blockchain, while its structure can include multiple levels and architectures.

API Layer **-** It is the application programming interface that allows external applications or users to interface with the blockchain. One of the main purposes of the interface is to define and formulate all the functions-services that the system can provide to other programs, without allowing access to the code that implements these services.

Step 9 –Elastic Stack Module

The widely used system Elastic Stack that comprises of the Elasticsearch, Logstash and Kibana is used for the complete and effective real-time visualization of the results of the analysis carried out.

This Module enables the aggregation and visualization of log files from all systems and applications used in the proposed architecture.

Step 10 – *Privacy Module*

Privacy and protection of industrial confidentiality when using the *Elastic Stack* Module is ensured through the use of the Privacy Module, which implements a differential privacy system. Remote, independent observers, who want access to information or search for specific content and they want a visual output, cannot understand if this information is coming from a specific source. According to the definition of differential privacy, a randomized algorithm A with a domain defining a data set D and a value field B, presents differential privacy (ε, δ) if relation 1 is true, for all adjacent datasets D_1 and D_2 and all subsets S of the domain values $S \subseteq B$ given that (ε, δ)>0:

$$P[A(D_1) \in S] \leq e^{\varepsilon} P[A(D_2) \in S] + \delta \quad (1)$$

From the above relation it follows that the smaller the (ε, δ) the greater the security of the records in the examined data sets. For the differential privacy model (ε, δ) it is true that δ=0. The differential privacy approach (ε, δ) as a method of encryption, is achieved by adding noise to the results of queries executed in the datasets.

If q is the question and n the noise to be added to maintain data privacy, then the randomized algorithm that exhibits differential privacy (ε, δ) is described by relation 2:

$$M = f(x) + n \quad (2)$$

where f is the function of all queries q and x is the dataset. Noise is added with the Laplace mechanism.

4 Conclusion

The idea of standardizing the proposed architecture arose based on the application of a single, universal method that will cover all the industrial requirements of the new era. It combines the most up-to-date methods, and it is able to complete specialized processes for the development of modern information systems security applications. This is achieved through an adaptable, flexible and easy-to-use operating environment. The features of the proposed architecture allow the analysis, forecasting, monitoring and management of complex situations related to information systems security. This is achieved by optimally combining and implementing a hybrid system with the most technologically advanced computing methods.

Assessing the proposed architecture as a whole, a significant advantage focuses on the ability to clearly display the information transmitted between the devices involved in the Industry 4.0 ecosystem. This leads to models of trust that are based on detailed mathematical and physical frameworks for the behavior of cyber-physical systems. Another important contribution of the proposed methodology is its contribution to the assessment of uncertainty posed by digital security problems, which is a major turning point in the adoption of new technologies. Finally, an important contribution of these methods lies in the fact that generalization is experimentally ensured by statistical validation techniques that cannot be disputed. This is true even when we are using data that contains a significant percentage of noise.

Proposals for development and future improvements of the methodology should focus on the automated optimization of the appropriate parameters of the method, so that the provision of services is achieved in a simple and categorical way.
It would be important to study the expansion of this system by implementing methods of self-improvement and redefining the parameters of the overall system, so that it can fully automate the process of engagement and disengagement in the supply chain of Industry 4.0.

References

1. L. Bassi, "Industry 4.0: Hope, hype or revolution?," 2017 IEEE 3rd International Forum on Research and Technologies for Society and Industry (RTSI), 2017, pp. 1-6, doi: 10.1109/RTSI.2017.8065927.
2. A. Shobol, M. H. Ali, M. Wadi and M. R. TüR, "Overview of Big Data in Smart Grid," 2019 8th International Conference on Renewable Energy Research and Applications (ICRERA), 2019, pp. 1022-1025, doi: 10.1109/ICRERA47325.2019.8996527.
3. C. H. Li and H. K. Lau, "A critical review of product safety in industry 4.0 applications," 2017 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), 2017, pp. 1661-1665, doi: 10.1109/IEEM.2017.8290175.
4. C. H. Li and H. K. Lau, "A critical review of product safety in industry 4.0 applications," 2017 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), 2017, pp. 1661-1665, doi: 10.1109/IEEM.2017.8290175.
5. Demertzis, K., Iliadis, L., Tziritas, N. et al. Anomaly detection via blockchained deep learning smart contracts in industry 4.0. Neural Comput & Applic 32, 17361–17378 (2020). https://doi.org/10.1007/s00521-020-05189-8
6. Demertzis, K.; Tsiknas, K.; Takezis, D.; Skianis, C.; Iliadis, L. Darknet Traffic Big-Data Analysis and Network Management for Real-Time Automating of the Malicious Intent Detection Process by a Weight Agnostic Neural Networks Framework. Electronics 2021, 10, 781. https://doi.org/10.3390/electronics10070781
7. X. Chen, J. Ji, C. Luo, W. Liao and P. Li, "When Machine Learning Meets Blockchain: A Decentralized, Privacy-preserving and Secure Design," 2018 IEEE International Conference on Big Data (Big Data), 2018, pp. 1178-1187, doi: 10.1109/BigData.2018.8622598.
8. D. Das and S. Sarkar, "Machine-to-Machine Learning based framework for ad-hoc IOT ecosystems," 2018 International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS), 2018, pp. 431-436, doi: 10.1109/CTEMS.2018.8769148.
9. Tsiknas, K.; Taketzis, D.; Demertzis, K.; Skianis, C. Cyber Threats to Industrial IoT: A Survey on Attacks and Countermeasures. IoT 2021, 2, 163-186. https://doi.org/10.3390/iot2010009
10. L. Winkel, "Real-Time Ethernet in IEC 61784-2 and IEC 61158 series," 2006 4th IEEE International Conference on Industrial Informatics, 2006, pp. 246-250, doi: 10.1109/INDIN.2006.275788.