UNIVERSITATEA POLITEHNICA
DIN BUCUREŞTI

Doctoral School of Electronics, Telecommunications
and Information Technology

Decision No. 778 from 03-12-2021

Ph.D. THESIS SUMMARY

Radu Marius BONCEA

CONTRIBUȚII LA CREȘTEREA SCALABILITĂȚII ȘI FIABILITĂȚII INTERNETULUI OBIECTELOR

CONTRIBUTIONS TO IMPROVE SCALABILITY AND RELIABILITY OF THE INTERNET OF OBJECTS

THESIS COMMITTEE

Prof. dr. ing. Ion MARGHESCU Politehnica Univ. of Bucharest	President
Prof. dr. ing. Ioan BACIVAROV Politehnica Univ. of Bucharest	PhD Supervisor
Prof. dr. ing. Mircea POPA Politehnica Univ. of Timișoara	Referee
Prof. dr. ing. Gheorghe ȘERBAN University of Pitesti	Referee
Prof. dr. ing. Paul ȘCHIOPU Politehnica Univ. of Bucharest	Referee

BUCHAREST 2022

Contents

Chapter 1

Introduction

1.1 Presentation of the field of the doctoral thesis

The Internet of Things (IoT) refers to the close connection between the digital and the physical world[1, 2], evolving into a complex system that uses several technologies, from the Internet to wireless communications, from microelectromechanical systems to embedded systems. Over time, several definitions have been proposed to describe what Internet of Things essentially means, but if we were to dilute all these definitions, we would conclude that IoT is a complex system of interconnected elements such as computing, mechanical or digital devices or assets, hierarchically arranged and enabling the implementation of digital services, based on interactions and processes defined and continuously improved.

In 2018, the number of devices connected to the Internet and used worldwide exceeded 17 billion, of which the number of IoT devices is about 7 billion (excluding smartphones, tablets, laptops or landlines)[3]. The overall growth of connections is mainly driven by IoT devices - both on the consumer side (e.g. Smart Home) and on the B2B segment. The number of active IoT devices is expected to increase to 10 billion by 2020 and 22 billion by 2025. This number of IoT devices includes all active connections and does not take into account devices that were purchased in the past but are no longer actively used.

1.2 Scope of the doctoral thesis

Although the core technologies of the Internet of Things (IoT) have advanced rapidly in recent years, large-scale deployments of IoT solutions are rare occurrences in the technical and economic landscape. This is mainly due to key challenges such as energy constraints, security, scalability, interoperability and communications.

In this context, this thesis aims to identify the particular challenges that impact the adoption of the Internet of Things and the implementation of large IoT systems, identify high-level requirements and integrable solutions according to a proposed reference architecture and validate the assumptions stated by implementing and operating the identified solutions.

The scope can be summarized in the following objectives:

- Identify major challenges in implementing complex IoT systems. By complex IoT systems, we mean those systems that can be the basis for the development of open applications and services that can be integrated by other systems.
- Analysis of the current state in terms of architectures and reference models, standards and recommendations.
- Defining the high-level requirements that address the identified challenges and the reference architecture on which IoT solutions can be deployed.
- Analysis of candidate technologies and models in solving identified problems that can be used in the development of IoT solutions.
- Implementation of the reference architecture as a case study and application of the proposed models and methods .

1.3 Content of the doctoral thesis

The thesis is structured in 7 chapters and 8 annexes.

Chapter 1 introduces the doctoral thesis, which aims to analyze current trends in the adoption and implementation of IoT systems. There are also several definitions of the Internet of Things that are enumerated and underlined for the purpose of this thesis.

Chapter 2 presents a study on reference architectures for IoT systems. I analyze those models considered mature from business perspective, widely adopted and implemented by Industry 4.0. Major challenges in IoT adoption such as energy constraint, data security and privacy, communications, scalability and interoperability are also identified.

Chapter 3 sets out the high-level, functional and non-functional requirements that specifically address the identified challenges. These requirements are the basis for proposing a reference architecture for a platform that is doing automatic monitoring and operations of IoT systems, organized on 3 levels: data sources, the data retention and fusion area and the analytical level.

Chapter 4 presents an in-depth study of supporting technologies in the implementation of automated monitoring and operational systems, grouped by scope: time-oriented databases, service configuration and discovery systems, large data processing systems , message intermediation systems and resource management solutions. In the introduction section of the chapter, a comparison is made between the two methods of monitoring IT infrastructures: the traditional method which is based on the programming and hard-coding of events and monitoring processes; the *smart* method that infers events after analyzing the data collected from the system. Two methods for the optimal selection of competing technologies are also presented: the method of maturity analysis using MADM models (Multi Attribute Decision

Making), with an example provided for time-oriented databases; the comparative method based on computational performance and resource utilization, with an example provided for key-value databases.
At the end of the chapter, the architectural model based on microservices is presented and I emphasize the opportunity to integrate Blockchain technology in the AIOps reference architecture.

Chapter 5 presents a study on machine learning algorithms, supervised or unsupervised, which can be used to solve problems such as classification, grouping, regression or dimensionality reduction. For each algorithm, the study focuses on describing the mathematical model, identifying the advantages and disadvantages, and the problems it solves.

Chapter 6 presents the implementation of the AIOps solution and the integration into ICI Bucharest infrastructure with the objective of monitoring Romanian Top Level Domain infrastructure and services and the ICIPRO cloud computing platform. I provide two examples of monitored services, the Whois service and the DNS service, for which I have proposed, as a problem, the identification and cataloging of cyber attacks that take place at the level of the two services. At the end of the chapter, two problems identified from monitoring based on data collected over time are presented and solutions are proposed.

Chapter 7 concludes the doctoral dissertation by listing the results obtained and the original contributions such as scientific articles published in journals or conferences and the projects in which I was involved. In conclusion, the prospects for further development are presented.

Chapter 2

The state of play for IoT technologies

2.1 IoT reference architectures

This chapter presents an extensive study with the purpose to identify major challenges in implementing large IoT systems. In this sense, the most used reference architectures in Industry 4.0 were studied, the field with the highest maturity in terms of standardizing the processes and constituent elements of an IoT system. The studied architectures are : RAMI - *Reference Architecture Model for Industrie 4.0*, IIRA - *Industrial Internet Reference Architecture*, IDS-RAM 3.0 - *IDS Reference Architecture Model*, BDVA - *Big Data Value Association*, Edgecross, IVI - *Industrial Value Chain*, OpenFog Consortium, Ocean Protocol, X-Road and Fiware. Based on this study, I proposed a general reference architecture for IoT systems [4], which can be seen as a common denominator of the studied architectures. Thus, the proposed architecture has 4 levels: the marginal level or *Edge* is the level where IoT devices are found and where the raw data is generated; the *Gateway* level is the level where data is collected and transmitted; *Cloud Platform* is the level at which data is stored and analyzed in context; presentation level is the level at which data is presented in the form of actionable information.

2.2 Technological challenges in IoT deployment

The most important challenge identified is the energy constraint, which affects the marginal level. Energy harvesting methods, which are described in detail in this chapter, can be used to generate at most hundreds of de $\mu W/cm^2$ of energy, insufficient to support complex software or hardware architectures, such as the IP technology stack. The current evolution of energy harvesting technologies, in the absence of a technological singularity, leads to the implementation of minimally viable and reliable systems for performing discrete measurements and communicating the results over short distances using lightweight protocols, such as radio frequency communications. Communications themselves are a challenge, because, as I have shown in the thesis, there are no technologies that have low energy consumption, high coverage, high transfer rates and low cost of implementation and operation. In the current economic and technological reality, it can be seen that for large IoT systems, optimal implementation involves low financial costs and tolerance for reduced performance (reduced transfer rates, reduced radius, power consumption or, most often, a combination of the three).

Another challenge is the scalability or ability of the IoT system to continuously accommodate an increasing number of assets, regardless of the level of architecture. For example, an increase in the number of devices is reflected in an increase in gateway and cloud computing requirements. There are several tools that address this challenge, described in detail in this chapter, such as the automated bootstrapping process, the control of the pipeline system for Big Data processing, the 3-axis scaling principle, the integration of microservice-oriented architecture and the adoption of data storage technologies.

Data security and confidentiality is also a major constraint, especially if we refer to the purpose of IoT systems, which is to turn data into actionable information, and, any data corruption can result in incorrect decisions with negative effects[5, 6, 7]. Key vulnerabilities are described in the thesis and include: identification, location and tracking, profiling, breach of privacy, life cycle transitions, data inventory and data connectivity.

Interoperability is another major challenge, being associated with the lack of open or general reference standards agreed at the industry level. The best way to avoid this challenge is to prepare IoT networks for interoperability from the outset. The highly fragmented IoT landscape will address three basic rules for IoT connectivity that facilitate network design: adopting open industry standards, adopting software-focused technologies, and using open interfaces.

Chapter 3

Reference architecture

3.1 High level requirements

This chapter proposes a reference architecture model that addresses the challenges identified in the implementation of large IoT systems, such as energy constraint, communications, data security and confidentiality, scalability and interoperability. Thus, a set of high-level requirements has been formulated, which aims to guide the process of implementing the functionality of an IoT ecosystem:

1. **Protocol agnosticism** is a principle according to which the system cannot be *captive* to a particular protocol, but, instead, by using specialized methods and models such as the adapter method or the proxy method, the system natively integrates all data collection protocols, or at least the most used protocols.

2. **Semantic structuring** implies the describing of the assets and services of the IoT system using standardized ontologies, such as SSN (*Semantic Sensor Network*), IoT-Lite, IoT-Stream, or SOSA (*Sensor Observation Sampling Actuator*).

3. **Data aggregation, augmentation and correlation** is a requirement for methods of aggregating, summarizing and filtering data collected from multiple sources.

4. **Data processing and machine learning** is the process of translating large volumes of raw data collected from the system (sensors, metric data related to the state of the system, augmented and associated data) into actionable information.

5. **Automatic orchestration and provisioning** involves implementing the capabilities needed to manage and support the processes used by digital services, in an automated and secure way, humans contributing strictly to supervision, processes improvement, identify unclassified anomalies and remedy where automation fails or is incomplete.

6. **Open Architecture** facilitates the replacement or the upgrading of the technologies used.

3.2 Reference model

The reference model for an AIOps platform is implemented taking into account a set of high-level functional requirements and is based on the data sources inventoried within the company: metric data generated by IoT devices and sensors; metric data associated with system performance, collected at the operating system level; metric data associated with the performance of applications and services; technical support data (e.g. ticketing); sources of intelligence or connected and profiled data based on social media, blogs and forums.

The next level of architecture is where the data is stored. Initially, the data is saved in a database optimized for short retention (usually a few days or hours), such as InfluxDB or Prometheus. This data is used for real-time analysis. After the retention period has expired, the data is *pushed* to long-term retention databases, such as OpenTSDB. This historical data is used to support machine learning algorithms and to correlate with past events. These data are a constant source for AIOps process optimization and provide a better understanding of complex systems.

The third level is associated with the data analysis system, which has the data transformed into information, using machine learning or artificial intelligence algorithms specialized in solving problems of classification, clustering, regression and dimensionality reduction.

Chapter 4

Support technologies

4.1 Monitoring

Implementing the reference architecture for an AIOps solution involves integrating a set of specialized technologies and solutions. The final solution requires the provision of intelligent monitoring processes, in order to integrate automated IT operations, based on machine learning models. AIOps platforms combine BigData technology and machine learning models to make it easier to operate core processes in large IT infrastructures. The AIOps platform works by scalable ingestion and ever-increasing data volume analysis. These data are characterized by a wide variety and speed of generation. The platform allows the simultaneous use of multiple data sources, data collection methods and analytical and presentation technologies.

4.2 Timeseries oriented databases

In this section I studied the most used timeseries oriented databases, such as Prometheus, InfluxDB, OpenTSDB, TimescaleDB. The solutions were compared based on performance and, additionally, I proposed a method for assessing technological and business maturity based on MADM model(*Multi Attribute Decision Making*) [8], which takes into account a wide range of qualitative and quantitative attributes, such as: market interest quantified by the number of Git stars, the number of appearances on StackOverflow, GoogleTrends score; active development (number of iterations on the source code, number of contributors to the code); pace of development (average time to fix bugs or new capabilities, pace of versioning; documentation and tools. The method of assessing the maturity of TSDB solutions is useful for reaching technical and business consent when deciding on the technological stack to be used. The method could also be profiled for different technologies and services[9, 10].

4.3 Service discovery and configuration solutions

Service discovery is the automatic process of detecting services , devices and other digital assets in a network using specific protocols. This capability has become critical for complex systems and in an omniscient computing environment or pervasive computing. I have studied several solutions, such as Etcd, Apache Zookeeper and

Consul and I have measured and compared their performance in terms of computational resources: processor utilization, memory, network, latency and write and read rates. Two implementation models are proposed, the server-centric model and the client-centric model and I presented the optimal implementation conditions as well as the high-level requirements associated with complex systems: sustainable and consistent storage, choosing the leader, atomicity of operations.

4.4 Data processing solutions

By definition, data processing is the automated process of converting data, which come in various formats, into actionable information. The process involves recording, analyzing, sorting, summarizing, performing calculations, presenting and storing. Because processed data is useful when it is presented to the user in an intuitive and useful way as information, data processing systems are known as information systems. Two processing methods are studied, batch processing and stream processing and the integrated technologies of Apache Spark, Apache Storm and Apache Flink are evaluated according to the methods of processing presented.

4.5 Message brokers

Communication and exchange of information between applications, systems, and other services is done through a software called a message broker, also known as a Message Oriented Middleware (*MOM*). Message brokers rely on a component called message queue, which stores the message until the applications are ready to process it. Messages are stored in the exact order of entry and remain in the queue until an acknowledgment is received. The message queue can also have managers who coordinate the interactions between multiple queues, message translation, services that provide data routing, but also the persistence and functionality of managing the client's state. The performance of Apache Kafka, Apache ActiveMQ and RabbitMQ technologies was studied and analyzed.

4.6 Resource management solutions

One of the high-level requirements for ensuring the horizontal scalability of IoT systems is the ability to automatically provide new computing resources when required. We can imagine the scenario in which the AIOps platform, following inference based on metric data collected and stored in time-oriented databases, inference which makes use of processing systems and specialized machine learning algorithms, issues alerts about high system utilization, which alerts, are routed by message brokers to resource management systems that can add new resources depending on the type of alert and parameters provided. The technologies studied are Kubernetes and Apache Mesos.

4.7 Microservices

Microservice-oriented architecture is an important and relative new approach in software architecture, and it aims at provide guideline on developing applications as a suite of services focused on distinct functionalities or domains, where each service is executed independently in its own process and where communication is supported by simple mechanisms such as HTTP protocol. The development of services itself can be done independently, both in terms of positioning in time and as requirements and workflow. We could say that microservices are an architectural style or pattern that structures an application as a collection of services that are easy to maintain and test, are decoupled, can be implemented independently and are organized around business capabilities [11].

A web service is usually associated with a software component or set of features, which, in collaboration, achieves a capability. A web service is uniquely identifiable and can be accessed via HTTP protocols (SOAP, Restful). The difference between a web service and a web microservice is related to the granularity of the component, the microservices being concentrated on limited sets of functionality. An important concept of microservice-oriented architecture is the composition of services, which can be achieved in two ways: by orchestration or by choreography.
Orchestration involves active control of all elements and interactions, while choreography involves establishing a pattern or routine that microservices follow, without the need for supervision and instruction.

4.8 Blockchain

Data security and confidentiality are the second most important obstacle to the deployment of IoT infrastructures, through challenges that include identification, location and tracking, profiling, privacy, life cycle transitions, data inventory and data connectivity. Although there are solutions in web application software architectures to address these challenges, in the M2M space, where data access and processing is done automatically, these solutions cannot be applied efficiently and evenly, requiring human adjustments, a process that is slow and leads to blockages. Given these constraints, this section presents a solution, that is the implementation of a smart contract (*smart contract*) and the integration into a blockchain ledger [12].

Chapter 5

Information intelligence

We define information intelligence as a technical method used to turn large volumes of complex data into relevant and actionable information, in order to better manage risks and increase productivity and thus profitability. The methods used are based on machine learning algorithms and data analysis tools and are used to solve a set of classical problems, such as: classification, grouping or clustering, regression, dimensionality reduction and noise elimination.

In this chapter, I study the main algorithms and models that can be used in solving the enumerated problems. The algorithms are compared in terms of complexity and performance and are associated with optimal application areas and issues. The algorithms were tested using data sets collected by the AIOps solution and grouped into:

1. Supervised machine learning algorithms: linear models (least squares method, Tikhonov regularization, Lasso model, Multi-task Lasso, Elastic-Net, Least Angle Regression, Orthogonal Matching Pursuit, Bayesian Regression, Logistic Regression, Stochastic Gradient Descendant , Huber Regression , Polynomial Regression), analysis of quadratic linear discriminants, Support Vector Machines, method of k-nearest neighbors, decision trees.

2. Unsupervised machine learning: Gaussian blended models, Manifold learning, isometric association, Locally Linear Embedding, Modified Locally Linear Embedding, Hessian Eigenmapping, Local Tangent Space Alignment, Multi-dimensional Scaling, t-distributed Stochastic Neighbor Embedding - t-SNE.

Information intelligence can bring great benefits and results even in the less visible areas of the proposed reference architecture. An example is in the area of cybersecurity, where we want to identify abnormal behaviors at the level of operator-system interaction [13, 14]. It can also be used to strengthen the knowledge of DevOps teams by providing automated documentation and testing tools [15, 16].

Chapter 6

Experimental implementation

This chapter describes the implementation of an AIOps platform, using the reference architecture and technologies described in the previous chapters. The initial purpose of this platform was to monitor the IT infrastructure of the Romanian .ro domain names registry, ROTLD. The implementation of the platform started in 2015 and, from the very beginning, a continuous iterative development was considered, with minimal interference at the level of existing infrastructure or applications. The goal was extended a year later by including monitoring of the ICI PRO data center and services, in particular IaaS services. The resulting system can be described technologically as having a converged infrastructure (hyperconverged from 2020), virtualized both at the network level and at the level of provisioning and administration of virtualized computing machines, using VMWare hypervisors and consisting of about 200 hosts and over 1000 virtual machines.

In terms of services and applications monitored and which are specific to the activity of the .ro domain registry and cloud computing services, the system consists of DNS services at the top domain level, Whois service, DAS (*Domain Availability Service*), EPP (*Extensible Provisioning Protocol*), programmatic interfaces for registrars, open data services and customer applications (databases, web applications). The data thus collected is provided to the analytical platform where it is processed and evaluated in the context of anomalies and events detection. The information obtained is passed on to decision levels, such as SOC (*Security Operation Center*) - the level that addresses the security policy, the DevOps level - the level of incident control and remediation, and the level of orchestration, where control and remediation are automatically enforced.

The selection of support technologies used for the implementation of the AIOps platform was made based on an analysis of maturity and technical performance, as follows:

1. The maturity analysis is based on a multi-attribute model [8], in which quantitative and qualitative attributes of the proposed solutions are evaluated.

2. Performance analysis is based on testing the main functions by simulating an execution environment under stress, following indicators such as resource utilization and processing capacity.

The system to be monitored is implemented using the *middleware* model as an architectural style. Middleware is a component that aggregates system functions into a set of new high-level functions, so that client applications strictly implement integration with middleware, without the need for integration with every element of the system. For example, in our case, the middleware provides access to the registration function of a .ro domain name. The client application will send all the data necessary to register a domain name to the middleware, and the middleware will execute the logic of this process: it will connect to the database, query and register the domain name with the associated data, will issue an invoice if necessary, will record the event in the log system, etc.
Thus, it is the middleware level where the metric data publishing interface associated with the applications has been implemented and used by Prometheus for data scrapping.

However, there are also situations where, for performance or security reasons, applications and services are implemented directly using the client-server model, such is the case with Whois and DNS services. Whois is a service that accepts search criteria (eg .ro domain name), performs database searches, and provides domain availability information or domain information. The DNS service provides customers with the ability to register domain name-specific attributes, using protocols associated with dynamic changes, or provide domain information based on search criteria.

The Whois service was implemented based on RFC3912 specifications and is used to provide domain name information. A typical process of interacting with the Whois service involves a client application submitting a search criterion, such as a domain name, and the service will query the database and provide the information according to the criteria. Rotld has 4 servers running associated with whois, and another 4 servers in hot standby that will start automatically when the number of queries increases or the capacity of the original 4 servers is exceeded.
For any request received by Whois application, the existence of an associated response in the cache database is checked. If there is a cached answer, the application will send that answer to the customer. If there is no response or the response was too older, the Whois application will send the request to a load balancer of the database with domain registrations. Like Whois applications, there are several instances of the database.

Some clarifications are required. There are two possible types of answers offered by the Whois service: if the domain exists, the service returns information about it; if the domain is not registered, the service will return a specific error. By implementing a counter based on Prometheus specifications, we will collect the number of responses for each response category.

For this service, we intend to identify, based on the metric data collected from the system and the answers provided, the state of the system from a security perspective. By state we understand one of the following situations:

- The system is under bruteforce attack, which is characterized by a large number of requests with many error-type responses (querying non-existent domains).

- Targeted attack, characterized by a large number of requests for existing domains.

- Combined attack, characterized by a large number of requests for both existing and non-existent domain names. This type of attack usually involves generating domain names based on a dictionary.

- The system is in the normal state if it is not in any of the previous states.

In the 6.1 table we have an example of data collected that is necessary for identifying the status of the Whois system. Supplementary, in addition to the number of *Found* and *Not Found* responses, we also collected CPU utilization. By CPU utilization, I mean the CPU load for the whole system by aggregating the metrics associated with each processor on each server of the system. The correlation of the data in the table can be seen in the figure 6.1, where we can see the following:

- The bruteforce attack correlates with a large number of *Not Found* responses and a relatively low CPU utilization. The low CPU utilization is explained by the fact that no additional domain query queries are performed for *Not Found* answers.

- The targetted attack correlates with a large number of *Found* responses and high CPU utilization, while the combined attack can be said to have a large number of *Found* and *responses. Not Found* and a medium-high CPU.

- The normal state consists of a small number of requests and therefore low CPU.

- In the data set provided we notice two outliers, the green dot should have been yellow and one of the yellow dots should be green.

Other correlations can be found. For example, based on the metric data collected, a *slowloris* attack can be identified. A slowloris attack involves opening a large number of connections to the server, without sending data or formulating the request, thus exhausting the pool of connections that the server may have. There are, of course, protection mechanisms, such as the implementation on the server side of an automatic connection shutdown mechanism that waits for data for no more than t seconds. But an attacker can establish hundreds of thousands of such connections using a simple computer in a very short time. And if the attack is distributed, with thousands of machines establishing tens of thousands of connections, the system may become unavailable for a long time.

Found Hits	Not Found Hits	CPU	Label	Label Name
1.3	0.19	1.5	0	NORMAL
1.6	0.9	1.8	0	NORMAL
0.90	0.23	1.2	0	NORMAL
1.1	0.19	1.3	0	NORMAL
0.83	0.23	1.1	0	NORMAL
1.12	0.26	1.2	0	NORMAL
0.56	0.22	0.89	0	NORMAL
2.31	0.98	1.82	0	NORMAL
7.52	1.22	7.42	0	NORMAL
16.12	1.9	18.2	1	TARGETTED
14.32	6.42	16.32	1	TARGETTED
12.37	1.08	11.28	1	TARGETTED
9.12	0.98	11.23	1	TARGETTED
1.23	0.24	1.23	1	TARGETTED
2.05	0.14	2.17	1	TARGETTED
3.9	0.38	4.1	1	TARGETTED
4.3	1.2	4.9	1	TARGETTED
5.32	0.76	5.6	1	TARGETTED
0.53	11.21	3.8	2	BRUTEFORCE
1.3	12.65	4.5	2	BRUTEFORCE
0.48	18.32	5.45	2	BRUTEFORCE
0.65	12.14	3.83	2	BRUTEFORCE
0.49	11.40	3.57	2	BRUTEFORCE
0.98	4.7	4.23	2	BRUTEFORCE
12.21	11.59	13.99	3	COMBINED
11.20	13.19	12.95	3	COMBINED
9.83	12.97	11.74	3	COMBINED
11.42	11.46	12.67	3	COMBINED
11.20	16.32	13.02	3	COMBINED
12.35	11.21	12.98	3	COMBINED

Table 6.1: Example of a training dataset used for machine learning when monitoring the status of the Whois service.

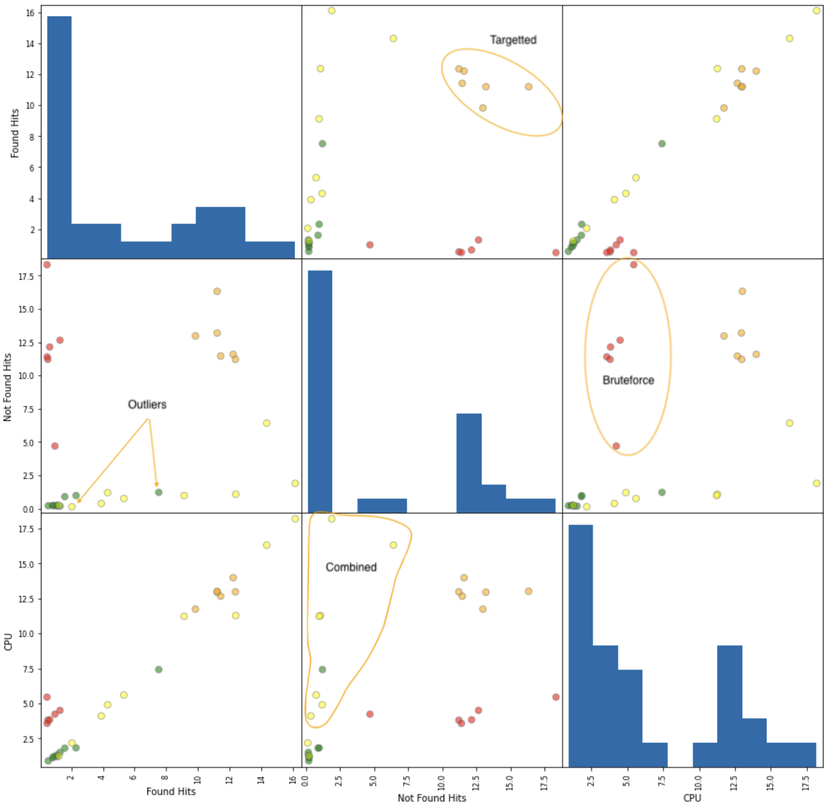


Figure 6.1: Correlation of responses and CPU utilization to determine system status

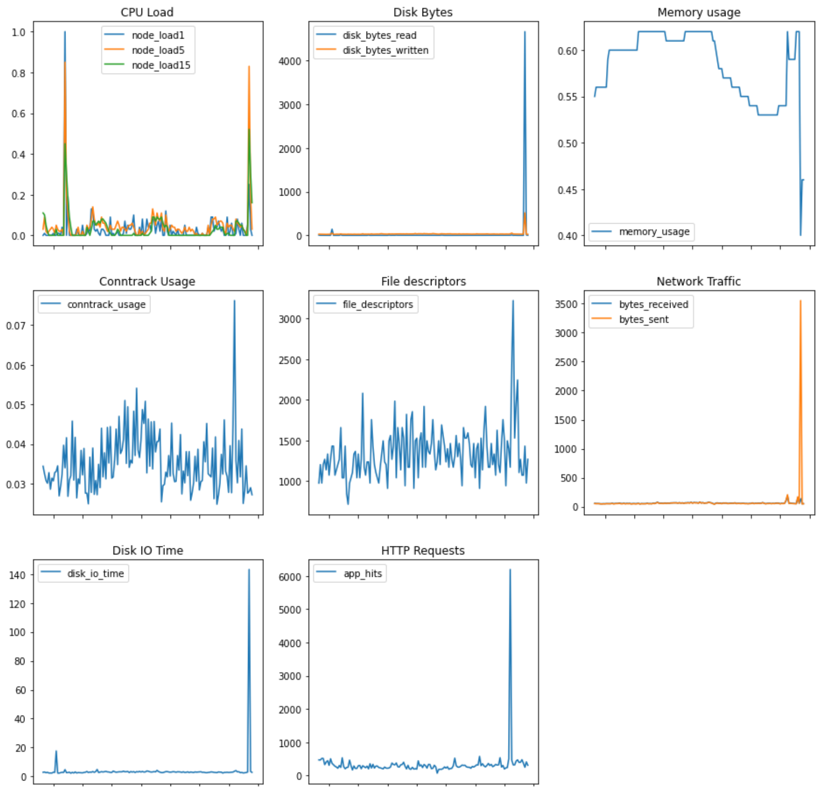
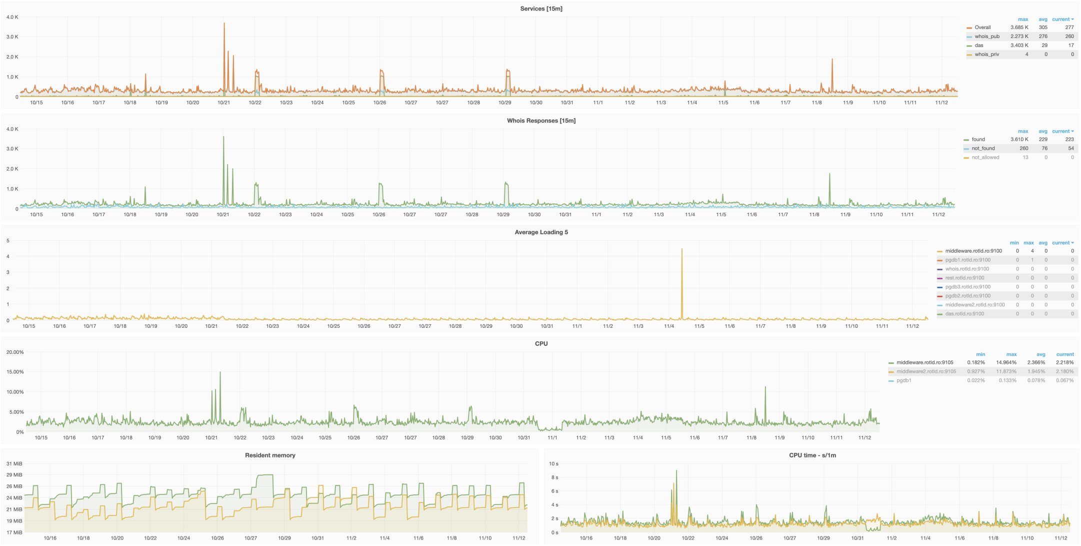


Figure 6.2: Different metric data associated with running the Whois service



	max	avg	current
Overall	3.685 K	305	277
whois_pub	2.273 K	276	260
das	3.403 K	29	17
whois_priv	4	0	0

	max	avg	current
found	3.610 K	229	223
not_found	260	76	54
not_allowed	13	0	0

	min	max	avg	current
middleware.rotld.ro:9100	0	4	0	0
pgdb1.rotld.ro:9100	0	1	0	0
whois.rotld.ro:9100	0	0	0	0
rest.rotld.ro:9100	0	0	0	0
pgdb3.rotld.ro:9100	0	0	0	0
pgdb2.rotld.ro:9100	0	0	0	0
middleware2.rotld.ro:9100	0	0	0	0
das.rotld.ro:9100	0	0	0	0

	min	max	avg	current
middleware.rotld.ro:9105	0.182%	14.964%	2.366%	2.218%
middleware2.rotld.ro:9105	0.927%	11.873%	1.945%	2.180%
pgdb1	0.022%	0.133%	0.078%	0.067%

Figure 6.3: Time series associated with Whois monitoring, viewed in Grafana and stored in Prometheus and OpenTSDB

Chapter 7

Conclusions

7.1 Obtained results

Currently, the implemented AIOps platform has a sufficient degree of maturity to allow the development of commercial monitoring services. Thus, the solution was implemented, tested and launched for the Ministry of Foreign Affairs, to monitor the platform cariera.mae.ro in the initial phase.

Also, in 2015 I had the chance to win and lead as a project manager, a project worth about USD 10 million, within the *textbf "Operational Program Competitiveness 2014-2020 Priority Axis 1 "*, project entitled ***"Creation of large data research laboratories to drive the development of innovative products and services for the Internet of Future"***, project code ID _34_462, SMIS 2014+, with Anagrama SRL as beneficiary. One of the objectives of the project was to demonstrate the capabilities of the infrastructure developed by implementing and defining as products or services, four smart-city mobile software applications: Smart City Map, Buy Local, City Drop and Jobs Nearb. Thus, I had the opportunity to implement the architecture described in this thesis in order to collect data from mobile devices and perform data analyzis in order to improve improve the services and the applications.

7.2 Original contributions

The general purpose of this thesis was to identify the challenges and obstacles to the implementation of large IoT systems, to identify the high-level requirements that address these challenges, to develop a reference architecture for the implementation of an IoT system, to develop an architecture reference of an IoT ecosystem monitoring system and to demonstrate these concepts and models in an industrial environment. Demonstration and validation of the described models have been carried out in some research projects that I won in the national competitions, which are listed in section 7.3.2. A "RDI projects won and managed as project manager". My original contributions are aimed at achieving the proposed goal, as follows:

1. I conducted a study on the current state of adoption of IoT technologies, standards and reference models for the implementation of complex and multilayered IoT systems. I have placed particular emphasis on the study of refer-

ence architectures developed by major industrial actors with a dual purpose: the generalization of a single reference model and the identification of areas that present particular challenges.[C1]

2. I proposed a generalized reference model for the implementation of IoT systems that can be seen as a common denominator of the studied models, a model that has 4 layers: the marginal layer that contains most IoT devices, the intermediate level or gateway, where data is aggregated and filtered, the cloud platform layer, where the data is stored, processed and analyzed, and the presentation layer.[C2, B3]

3. I conducted a detailed study of the major challenges in implementing IoT systems. The most important challenge identified, energy constraint, has a major impact on architecture. Current methods of energy harvesting and storage do not allow the implementation of complex communication protocols or methods of monitoring the status of IoT assets. This is why I conducted a comparative study of communication protocols from the perspective of energy consumption versus cost, coverage area and transfer rates, using the Z score method. [A3,A4]

4. I conducted a study on data security and confidentiality, in which I identified and ranked, by impact and relevance, the security issues associated with IoT systems, both from the perspective of the IoT service consumer and the provider.[C2,A3]

5. I have formulated high-level, functional requirements for the implementation of an IoT system that addresses the identified challenges, such as communication protocol agnosticism, semantic structuring, data aggregation and augmentation, data processing and machine learning, orchestration and machine provisioning and the need for open architecture. Based on these requirements, we have proposed a reference architecture for the implementation of an AIOps platform that aims to automatically monitor IoT assets in order to increase the scalability, availability and reliability of the IoT system. [A1, B3]

6. I have identified and tested the technologies that can be used in the implementation of the AIOps platform, with focus on those technologies that facilitate the transition to hyperconvergent infrastructures. [C4, B9]

7. I have developed and proposed a model for analyzing the maturity of solutions that can be used in strategic decisions. The algorithm is based on the multi-attribute decision models MADM (Multi Attribute Decision Making) and I used it in selecting the optimal solution for time-oriented databases. However, the model can be generalized and used for any other IT technologies. [A2,B1]

8. I coordinated the research and development activities within a work package of Cloud for Europe C4E project, activities that aimed at developing the technical specifications (reference architecture and requirements) necessary to implement a secure and legal-rules-based cloud storage solution , textit "Secure Legislation Aware Storage". [B2,B4]

9. Based on the solution for maturity analysis of technologies, I contributed to the development of two MADM models for selecting cloud service providers.[A10,A11]

10. I performed a comparative analysis of the performance of key-value database solutions, by running stress tests and measuring performance in different scenarios. The method can also be applied to other solutions and can be a quantitative evaluation criterion for the maturity analysis method.[C4]

11. I tested the opportunity to integrate Blockchain technology as a method of providing and authenticating IoT services, using Blockchain transactions to register IoT devices and services, and using *smart contracts* to validate access to services and devices. I used Ethereum and Arwen WASM virtual machines (Elrond).[B9]

12. I have contributed to the development of methods for automating the code testing processes of continuously implemented applications, according to the "*agile*" working methodology and "*DevOps*" operational environment. [B6]

13. I have conducted a study of the main machine learning algorithms that can be used to solve problems associated with grouping, classification, regression, dimensionality reduction and noise elimination, in order to compare the advantages, disadvantages, complexity and optimal areas of application. [C3, A5]

14. I have adapted the reference model for the AIOps platform in order to develop a distributed system for scanning common cyber vulnerabilities at national level by collecting data to identify the deployment technologies of .ro websites, a method that is known as *webserver fingerprint.* The solution was implemented at the level of the .ro domain register as a concept. [B5]

15. I have implemented, tested and operated the AIOps platform based on the reference architecture described, under industrial conditions, currently the platform being operational at the .ro domain name registry (RoTLD). The developed solution was profiled for commercial purposes, being currently used by certain government institutions. [C4]

16. Using the data collected by the AIOps platform from the two data centers, ICIPRO and RoTLD, during approximately 3 years of continuous operation, I have contributed to the development of new methods and models for consolidating data centers in order to optimize energy consumption. [A9].

17. I investigated the possibility of detecting abnormal behavior by users of a monitored AIOps application by recording user-web interface interactions in the AIOps platform. If the subject of the interaction can be viewed as a node in a graph and the interaction itself as an edge, then the sequence of interactions over a period of time can be modeled as an oriented graph. [A6]

18. As a project manager of the project "Creation of large data research laboratories to drive the development of innovative products and services for the Internet of Future", I contributed to the implementation of BigData computing infrastructure, using methods and methodologies described in this thesis,

such as the selection of technologies based on maturity analysis and the implementation of the AIOps reference architecture.[B1]

7.3 List of original publications

7.3.1 Scientific articles

A. Scientific articles in ISI indexed publications

A1 Boncea, R., A. Zamfiroiu, and I. Bacivarov, *New method for monitoring microservices in a federated and distributed architecture*, Proceedings of the IE 2018 International Conference, 17-20 May, Iasi, Romania, 2018.

A2 I Petre, R Boncea, CZ Radulescu, A Zamfiroiu, I Sandu, *A Time-Series Database Analysis Based on a Multi-attribute Maturity Model*, Studies in Informatics and Control, ISSN 1220-1766, vol. 28(2), pp. 177-188, 2019. WOS:000473284800006

A3 Alin Zamfiroiu, Ionut Petre, and Radu Boncea. 2019. *Cloud Computing Vulnerabilities Analysis.* In Proceedings of the 2019 4th International Conference on Cloud Computing and Internet of Things (CCIOT 2019). Association for Computing Machinery, New York, NY, USA, 48–53. DOI:https://doi.org/10.1145/3361821.3361830. WOS:000526710900008

A4 Carmen Elena CÎRNU, Carmen Ionela ROTUNĂ, Adrian Victor VEVERA, Radu BONCEA, *Measures to Mitigate Cybersecurity Risks and Vulnerabilities in Service-Oriented Architecture*, Studies in Informatics and Control, 08 2018, ISSN: 1220-1766 eISSN: 1841. WOS:000447079500011

A5 Boncea, R., A. Zamfiroiu, and E. Mitan, *Proposing algorithm to improve student evaluation process*, 10th Annual International Conference on Education and New Learning Technologies, Palma de Mallorca (Spain), 2nd - 4th of July, 2018. WOS:000531474300023

A6 A. Zamfiroiu and R. Boncea, *Modelling the users' profiles based on their behaviour in social applications*, 2018 10th International Conference on Electronics, Computers and Artificial Intelligence (ECAI), Iasi, Romania, 2018, pp. 1-6, doi: 10.1109/ECAI.2018.8678990. WOS:000467734100060

A7 Petre, I., Cohal, A.M., Boncea, R., *e-Participation Platform for Facilitating Citizens Involvement in Smart City Initiatives*, Revista Română de Informatică şi Automatică, Volume: 28 Issue: 2 Pages: 5-14, 2018. WOS:000455837200001

A8 Zamfiroiu, A., Boncea, R., Petre, I., *Quality of mobile applications based on their development*, Romanian journal of information technology and automatic control, Volume: 28 Issue: 1 Pages: 35-46, 2018. WOS:000455836300003

A9 Delia Mihaela Radulescu, Constanta Zoie Radulescu, Gheorghe Lazaroiu, Radu Boncea, *Binary programming models for the server consolidation problem in data centers*, The 18 International Multidisciplinary Scientific

GeoConference SGEM 2018, 30 June - 9 July 2018, Albena, Bulgaria, The accepted article will be published in the Conference Proceedings (ISSN 1314-2704) and will be submitted for evaluating and indexing by ISI Web of Knowledge, Web of Science, Thomson Reuters

A10 C. Z. Rădulescu, I. C. Rădulescu, R. Boncea and E. Mitan, *A group decision approach based on rough multi-attribute methods for Cloud Services Provider selection*, 2018 10th International Conference on Electronics, Computers and Artificial Intelligence (ECAI), Iasi, Romania, 2018, pp. 1-6, doi: 10.1109/ECAI.2018.8678966. WOS:000467734100036

A11 Constanţa Zoie RADULESCU, Marius RADULESCU, Radu BONCEA, Ionut PETRE, Ionut-Eugen SANDU, Mihail DUMITRACHE, *A Multicriteria Framework for Cloud Service Providers Selection Based on the Matter Element Extension Method*, Studies in Informatics and Control, ISSN 1220-1766, vol. 30(1), pp. 77-87, 2021. https://doi.org/10.24846/v30i1y202107

A12 Boncea, R., Petre, I., Vevera, V., Gheorghiţă, A. *Machine Learning Based Methods Used for Improving Scholar Performance.* In The International Scientific Conference eLearning and Software for Education 2019 (Vol. 2, pp. 471-478). Carol I National Defence University. WOS:000473324400065.

A13 BANCIU, D., PETRE, I., BONCEA, R. *Information and Documentation through New Technologies in E-Learning Process.* In The International Scientific Conference eLearning and Software for Education 2019 (Vol. 2, pp. 465-470). Carol I National Defence University. WOS:000473324400064.

A14 Alin Zamfiroiu, Radu Boncea. *Using Decision Tree and Machine Learning to recognize users by their behaviour.* Proceedings of the 16th international conference on informatics in economy (IE 2017). WOS: 000418463600015.

B. Scientific articles in BDI indexed publications

B1 Radu BONCEA, Ionuț PETRE , Dragoș-Marian SMADA, Alin ZAMFIROIU, *A Maturity Analysis of Big Data Technologies*, Informatica Economică vol. 21, no. 1/2017, DOI 10.12948, ISSN 14531305

B2 Alin Zamfiroiu, Carmen Elena Cîrnu, Radu Boncea, Carmen Rotună, Monica Anghel, *Principii de proiectare, securitate și administrare a soluțiilor de stocare în cloud*, Revista Română de Informatică și Automatică, vol. 25, nr. 2, 2015

B3 Radu BONCEA, Ioan BACIVAROV *System Architecture for Monitoring the Reliability of IoT*, PROCEEDINGS of the 15th International Conference on Quality and Dependability, pp. 143-149, 2016

B4 Radu BONCEA, Carmen Elena CÎRNU, *Cloud for Europe Project: New Solutions for Addressing Cloud Security Issues*, PROCEEDINGS of the 15th International Conference on Quality and Dependability, PROCEEDINGS of the 15th International Conference on Quality and Dependability, pp.156-160

B5 Eugenie STĂICUȚ, Radu BONCEA, Carmen ROTUNĂ, *A Reliable Architecture for a Massive and Continuous Scanner of Web Vulnerabilities in Internet*, PROCEEDINGS of the 15th International Conference on Quality and Dependability, pp. 176-180

B6 Dragoș SMADA, Carmen ROTUNĂ, Radu BONCEA, Ionuț PETRE. *Automated Code Testing System for Bug Prevention in Web-based User Interfaces*, Revista de Informatica Economică, vol. 22, no. 3/2018

B7 Alin ZAMFIROIU, Radu BONCEA, Ionuț PETRE, *Quality of mobile applications based on their development*, Romanian Journal of Information Technology and Automatic Control, ISSN 1220-1758, vol. 28(1), pp. 35-46, 2018

B8 Badea, V.E., A. Zamfiroiu, and R. Boncea, *Big Data in the Aerospace Industry*, Revista Română de Informatică şi Automatică, vol. 22, no. 1, pp. 17-24, 2018

B9 Radu BONCEA, Ionut PETRE, Victor VEVERA, *Building Trust Among Things in Omniscient Internet Using Blockchain Technology*, Romanian Cyber Security Journal, Spring 2019, No. 1, vol. 1.

B10 Gabriel Neagu, Ionut Petre, Radu Boncea, Mihail Dumitrache, *Building a business model for service offer integrator in case of cloud-iot based monitoring*, Conference: 18th International Conference on INFORMATICS in ECONOMY. Education, Research and Business Technologies, May 2019, DOI: 10.12948/ie2019.02.05.

C. Scientific reports within the doctoral program

C1 The scientific report nr.1 "*State of Art Analysis of the Internet of Things*"

C2 The scientific report nr.2 "*IoT reference architectures*"

C3 The scientific report nr.3 "*Applications of artificial intelligence algorithms in wireless sensor networks to detect system errors and anomalies*"

C4 The scientific report nr.4 "*Integrating machine learning into data-driven infrastructures monitoring architectures*"

C5 The scientific report nr.5 "*AIOps solution reference architecture*"

7.3.2 Research, development and innovation projects

A. RDI projects won and managed

D1 **PN 19370401 "New solutions to complex problems in current ITC research areas based on modeling and optimization"**, Programul Nucleu 2019-2022. The project focuses on developing a collection of new solutions to complex problems in current areas of ITC research, based on modeling and optimization. New solutions mean new methods and models, methodologies, algorithms and software, as a result of modeling complex problems. The complex issues addressed by the project relate to: (a) increasing the reliability and security of complex systems, (b) the acquisition and analysis of quantitative data through integrated solutions, (c) improving energy efficiency in data centers, (d) optimal selection of cloud, product and energy services suppliers, (e) complex issues of NP hard graph theory and (f) data security in software systems for mobile devices.

D2 **PN18190101 "New research in modeling and optimizing of complex systems with applications in industry, business and cloud computing"**.The project proposed and developed new solutions for complex systems with applicability in industry and business. Methods, models and optimization software tools usable in the industrial context, multi-criteria decision-making methods, algorithms and software with cloud computing and business applications as well as graph theory algorithms for NP-hard problems with real applications have been developed.

D3 **"Creation of large data research laboratories to drive the development of innovative products and services for the Internet of Future"**, cod de proiect ID_34_462, SMIS 2014+, Operational Program Competitiveness 2014-2020 Priority Axis 1. The project aims to develop BigData labs in order to drive the development of innovative products and services for the Future Internet. The project objective was to prove the labs capabilities by developing 4 smart-city applications.

D4 **"Study on adaptive systems for early-stage recognition of cyber attacks on state resources"** within the MCSI Sector Plan 2018-2020. The project aimed at developing a reference architecture for a cyber attack detection system using machine learning models, algorithms and the analysis of metric data collected from the system.

B. Other projects

E1 The European project **"EuroCC - National Competence Centers in HPC (High Performance Computing)"** aims to establish, connect and operate a number of 33 national competency centers in HPC in order to facilitate access to HPC technologies, knowledge, expertise, aligned to specific national needs and depending on the level of maturity of the HPC of each state.

E2 The European project **"Cloud for Europe - C4E"** addresses the objectives of the European Cloud Partnership and helps partners to adopt a well-defined European strategy for cloud computing technology for the public sector. The objectives of the project are: to identify obstacles to the use of Cloud Computing technology in the public sector, to define the services that will overcome these obstacles, to facilitate research by industry and to identify innovative solutions for Cloud services.

E3 Core pogram project **"Research on advanced policies and solutions to secure critical infrastructure against cyber attacks"** focuses on the area of industrial control systems, the identification of vulnerabilities at organizational structure level, industrial control systems and networks, as well as on applied security procedures that can be exploited involuntarily or voluntarily by one or more attackers thus generating cyber events with strong impact on business, critical infrastructure and national security.

7.4 Prospects for further development

One of the major challenges identified in the implementation of large IoT systems is cyber security, especially data security. A solution for data security through access authorization and auditing methods is the integration of Blockchain technology, a technology tested in laboratory conditions by defining a catalogue of IoT devices and services, where the registration of an asset implies commiting a Blockchain transaction that contains information (*payload*) associated with asset identity. Various methods of authenticating and authorizing access to devices and services were also tested using *smart contracts* and Ethereum and Arwen acrshort wasm virtual machines. But the integration of Blockchain in order to operate in industrial conditions has not been tested.

Thus, a development prospect is the testing under operational conditions, by verifying the performances and the associated costs, as follows:

– compare transaction costs for different Blockchain solutions;

– compare the performances associated with the average ledger access times, latencies, etc;

– studies on new models of information compression stored in the ledger;

– studies on the opportunity to develop dedicated and independently operated subchains similar to the IoTex architecture;

– studies on use cases and examples of decentralized applications (DApps).

Another direction of development is the simulation of IoT systems, both at the level of network or communication protocols, and at the level of generated data, by implementing similar tools to *ns3* (Network Simulator 3). The implementation of such a simulator is in itself a large-scale project that involves the development of a set of tools, each instrument being specialized. Following a brief market study I conducted, I concluded that there are specialized simulators on certain capabilities (NetSim, Proteous, CupCarbon, Bevywise, NS3, Cooja, MANET), but there is no simulation environment that integrates simulators on all architectural levels.

Advanced embedded systems for edge computing are also an area of interest, as major industry players, such as NVidia, have developed hardware and software solutions that can run on complex models of neural networks, in domains such as video or image processing, with low power consumption. An example is the Nvidia Jetson product portfolio, which, integrated at the gateway level into the IoT reference architecture described in this thesis, can substantially change the data flow between the cloud platform and the gateway. We can imagine a scenario in which there is a video camera located in traffic junction and we aim to count the cars that pass through the intersection every t minutes. By integrating a solution like the NVidia Jetson Nano as a gateway located next to the camera, the counting process can be performed by Jetson, which has implemented real-time object recognition models, including cars recognition. This avoids sending the video stream to the cloud platform for processing, instead we will strictly send the status of the counter, saving significant network and communication resources.

As the head of research department "Software and complex systems engineering" within ICI Bucharest, I set out to establish an innovation laboratory in the field of the Internet of Things within the institute. In this regard, I have made initial agreements with various stakeholders, such as companies whose object of activity is research and innovation, universities and local authorities interested in implementing smart city technologies. The aim of the laboratory is to provide the material and knowledge base for young ICI researchers or interns, through their involvement in research, development and innovation projects with the participation of stakeholders. Another objective of the laboratory is to spin-off solutions that reach the level of technological maturity at least demonstrable in simulated or laboratory conditions and to ensure their transition to commercially valid products and services.

Bibliography

[1] L. Atzori, A. Iera, and G. Morabito, “The internet of things: A survey,” *Comput. Netw.*, vol. 54, pp. 2787–2805, Oct. 2010.

[2] B. Sterling, *Shaping Things (Mediaworks Pamphlets).* The MIT Press, Sept. 2005.

[3] K. L. Lueth, “State of the iot 2018: Number of iot devices now at 7b – market accelerating,” tech. rep., 2018.

[4] R. Boncea and I. Bacivarov, “A system architecture for monitoring the reliability of iot,” in *Proceedings of the 15th International Conference on Quality and Dependability*, pp. 143–150, SOCIETATEA ROMÂNĂ PENTRU ASIGURAREA CALITĂȚII, 2016.

[5] R. Boncea and C. E. Cîrnu, “Cloud for europe project: New solutions for addressing cloud security issues,” in *Proceedings of the 15th International Conference on Quality and Dependability*, pp. 156–160, SOCIETATEA ROMÂNĂ PENTRU ASIGURAREA CALITĂȚII, 2016.

[6] A. Zamfiroiu, I. Petre, and R. Boncea, “Cloud computing vulnerabilities analysis,” in *Proceedings of the 2019 4th International Conference on Cloud Computing and Internet of Things*, CCIOT 2019, (New York, NY, USA), p. 48–53, Association for Computing Machinery, 2019.

[7] C. CÎRNU, C. I. ROTUNĂ, A. V. VEVERA, and R. BONCEA, “Measures to mitigate cybersecurity risks and vulnerabilities in service-oriented architecture,” *Studies in Informatics and Control*, vol. 27, no. 3, pp. 359–368, 2018.

[8] I. PETRE, R. BONCEA, C. Z. RADULESCU, A. ZAMFIROIU, and I. SANDU, “A time-series database analysis based on a multi- attribute maturity model,” *Studies in Informatics and Control*, vol. 28, no. 2, pp. 177–188, 2019.

[9] C. Z. Rădulescu, I. C. Rădulescu, R. Boncea, and E. Mitan, “A group decision approach based on rough multi-attribute methods for cloud services provider selection,” *2018 10th International Conference on Electronics, Computers and Artificial Intelligence (ECAI)*, pp. 1–6, 2018.

[10] I. Z. RADULESCU, M. RADULESCU, R. BONCEA, I. PETRE, I.-E. SANDU, and M. DUMITRACHE, “A multicriteria framework for cloud service providers selection based on the matter element extension method,” *Studies in Informatics and Control*, vol. 30, no. 1, pp. 77–87, 2021.

[11] R. Boncea, A. Zamfiroiu, and I. Bacivarov, “New method for monitoring microservices in a federated and distributed architecture,” in *Proceedings of the 17th International Conference on Informatics in Economy*, pp. 13–18, Bucharest University of Economic, 2018.

[12] R. BONCEA, I. PETRE, and V. VEVERA, “Building trust among things in omniscient internet using blockchain technology,” *Romanian Cyber Security Journal*, vol. 1, no. 1, pp. 17–24, 2019.

[13] A. Zamfiroiu and R. Boncea, “Modelling the users’ profiles based on their behaviour in social applications,” *2018 10th International Conference on Electronics, Computers and Artificial Intelligence (ECAI)*, pp. 1–6, 2018.

[14] A. Zamfiroiu and R. Boncea, “Using decision tree and machine learning to recognize users by their behaviour,” in *PROCEEDINGS OF THE 16TH INTERNATIONAL CONFERENCE ON INFORMATICS IN ECONOMY (IE 2017)*, pp. 90–95, Bucharest University of Economic Studies, 2017.

[15] R. Boncea, A. Zamfiroiu, and E. Mitan, “Proposing algorithm to improve student evaluation process,” in *EDULEARN18 Proceedings*, 10th International Conference on Education and New Learning Technologies, pp. 5799–5805, IATED, 2-4 July, 2018 2018.

[16] R. Boncea, V. Vevera, I. Petre, and A. Gheorghita, “Machine learning based methods used for improving scholar performance,” in *The International Scientific Conference eLearning and Software for Education*, vol. 2, pp. 471–478, Carol I National Defence University, 2019.