PRIVACY AND SECURITY FEDERATED REFERENCE ARCHITECTURE FOR THE INTERNET OF THINGS

By

Musab Kamal

A thesis submitted to the faculty of Information Security Department, Military College of Signals, National University of Sciences and Technology, Islamabad, Pakistan, in partial fulfillment of the requirements for the degree of MS in Systems Engineering.

JUNE 2021

ABSTRACT

Physical objects are getting connected to the internet at an exceptional rate making the idea of the Internet of things a reality. The advent of the internet of things ecosystem is everywhere in our daily lives in the form of smart homes, smart healthcare systems, smart wearables, smart connected vehicles, and industries. This has given rise to the risks associated with the privacy and security concerns of the users. The potential growth of the IoT products might get hindered due to the increasing amount of cyber-attacks on IoT devices due to deficiencies in its architecture. To counter it we need to implement privacy and security rights from its building blocks. This can be achieved through the reference architecture. A reference architecture is a recommended structure a building block that integrates products and services. It provides a framework for the domain. There has been an evolution of IoT architecture's over the years improving the stack of architecture with new solutions like Scalability, Management, Interoperability, and Extensibility. This gives us more responsibility and the need to standardize and organize IoT reference architecture in federation with privacy and security concerns. In this thesis, we propose the core IoT requirements extracted through the standards considering the quantifiable metrics that integrate privacy and security as well. These requirements are divided into functional and non-functional requirements. We surveyed and analyzed twelve existing IoT reference architectures based on these requirements. Shortcomings were identified through the analysis, we then proposed privacy federated IoT security reference architecture that addresses privacy and security concerns and is a step towards the standardization of concrete domain architecture. Finally to validate our proposed reference architecture we used the industry-recognized scenario-based approach known as the Architecture Tradeoff Analysis Method (ATAM).

DEDICATION

This thesis is dedicated to

MY PARENTS, FAMILY, AND TEACHERS

for their love, endless support, and encouragement

ACKNOWLEDGEMENTS

I am grateful to ALLAH Almighty who has bestowed me with the strength and the passion to accomplish this thesis and I am thankful to Him for His mercy and benevolence. Without His consent, I could not have indulged myself in this task.

TABLE OF CONTENTS

TABLE OF FIGURES

LIST OF TABLES

ACRONYM

Internet of Things	IoT
Internet Protocol Version 6	IPv6
Machine to Machine	M2M
Radio Frequency Identification	RFID
Distributed Denial of Service	DDOS
Domain Name System	DNS
International Organization for Standards	ISO
International Electro technical Commission	IEC
World Standards Cooperation	WSI
Electronic Product Code Global	EPC Global
China Electronics Standardization Institute	CESI
National Institute of Standards and Technology	NIST
Internet Architecture Board	IAB
Near Field Communication	NFC
Vehicular ad hoc Networks	VANETs
Open Systems Interconnection	OSI
Wireless Sensor Network	WSN
Low Power and Lossy Networks	LLN
International Telecommunication Union	ITU
Physical Unclonable Functions	PUF
Special Working Group	SWG
Ad Hoc Group	AHG
European Union	EU
Institute of Electrical and Electronics Engineers	IEEE
Privacy Enhancing Technologies	PET’s
Business to Business	B2B
Application Programming Interface	API
Architectural Reference Model	ARM
Privacy Enhancing Technologies	PETs
Software-Defined Network	SDN
Transport Layer Security	TLS
Secure Socket Layer	SSL
Virtual Private Network	VPN
Java Script Object Notation	JSON
Customer Relationship Management	CRM
Enterprise Resource Planning	ERP
Enterprise Application Integration	EAI
Business to Business	B2B
Software Development Kit	SDK
Structured Query Language	SQL
Extract, Transform, Load	ETL
Lightweight Directory Access Protocol	LDAP
Secure and Safe Internet of Things	SerIoT
Single Sign-On	SSO
Return on Investment	ROI
Privacy Control Record	PCR
Privacy Validation Chain	PVC
Architecture Significant Requirement	ASR

CHAPTER 1

1. INTRODUCTION

In this technology-driven era where everything is interconnected with each other. We can communicate with each other irrespective of the distance, can see and hear each other despite being at a distance with each other at thousands of miles. This usually refers to the Internet of things where everything is connected. Internet of things are smart lightweight devices that consist of embedded processors, sensors, actuators, and communication hardware that intelligently acquire collect, and sends data from their respective environments. These IoT devices share the collected data through the gateway or other edge devices where the data is being analyzed on the cloud or locally. These are called smart and intelligent devices because they do all the work without human intervention although people can interact with them. IoT has evolved over the years according to the statistics the number of IoT devices will certainly increase from 20.35 billion in 2017 to 75.44 billion in 2025 [1]. It is expected by 2022 the M2M traffic flows will constitute up to 45% of the whole internet traffic [2]. The market share and the economic impact of the internet of things is expected to be between $2.7 trillion to $6.2 trillion by 2025[2]. Internet of things has evolved over the years and it's been the center of attention for quite a while now in research and development. It constitutes the platforms that use RFID for the traceability of goods, algorithms for new solutions. Ipv6 and novel protocols for resource-constrained devices. It promises to evolve further in cloud computing, big data, networking, and social networks.

This evolution of the connected things devices with each other and the internet has brought up the heterogeneity in the ecosystem of the IoT. This has given rise to security and privacy concerns for the users. According to [3] there were distributed denial of service (DDOS) attacks to the DNS servers of PayPal, Twitter, Visa, etc. Several Vulnerable IoT devices like printers, Ip cameras, residential gateway, and baby monitors were affected by Mirai malware. A load of this attack was 1.2 terabits per second experts labeled it to be the largest DDOS attack on record [3]. Not much work has been done on the privacy of the users in the IoT ecosystem. According to [4] very limited amount of work has been done on the privacy and data security of the sensitive sensors and actuators as the readings from the sensors can obtain the habits and patterns of the end-users. Like when they were present at home when they leave when the guests arrive this could also be perceived as a violation of privacy. There remains a big question on the data of the users as to whether it is being profiled based on identities, where the data is being stored on the cloud under what defined purpose.

To deliver quality products to the consumers in the market IoT needs standardization in its architecture. To meet the challenges of the IoT its architecture should be revised and this can only be achieved through refined reference architecture. Many organizations are working on the building blocks of the IoT following different standard bodies like International Organization for Standards (ISO), the International Electro technical Commission (IEC), Institute of Electrical and Electronics Engineers (IEEE). Worlds Standards Cooperation (WSI), Electronic Product Code Global (EPCglobal), China Electronics Standardization Institute (CESI), and National Institute of Standards and Technology (NIST) [5]. This leads to the vendors, research, and development to follow the different standardization bodies there is a need to converge the standard bodies as well to address the IoT ecosystem as a whole. Standardization leads to interoperability which enhances the integration and exchange of information between distributed systems [6].

The Internet Architecture Board (IAB) has defined the communication models as device to device, device to the cloud, device to gateway, and backend sharing model. Each model has its way of communication. IoT devices are usually resource-constrained containing limited processing, power, and storage capabilities. But things these days consisting of devices, sensors, actuators are having increased processing, power, and storage capabilities. The combination of multifunctional devices and sensors is extremely effective for communication with each other and the internet. This physical world utilizing these smart devices is connected to cyberspace and the Internet of things. These physical objects are equipped with Radio Frequency Identification (RFID) tags, Near Field Communication (NFC) tags, and electronic bar codes that can be scanned by smartphones, tablets, and other smart devices integrated with RFID/NFC readers [7]. The devices that don't usually come equipped with the interface for the connection can also be integrated with the IoT systems. The platforms of such specifications are Arduino, .Net Gadgeteer, and even Lego Mindstorms.

Internet of things will contribute a major part to the economy by having significant applications like having a smart home that can automatically open the garage when reaching home, prepare the coffee, control the climate systems, smart TVs, etc. [2]. Vehicular ad hoc networks (VANETs), mission-critical applications, and control systems are also the applications of the IoT. There is a need for a modular and interoperable architecture that should lead us towards the standard architecture for the IoT just like the OSI layer model for network communications.

Systemic privacy flaws found in popular IoT devices from manufacturers like iHome, Merkury, Momentum, Oco, Practecol, Tplink, Wyze, and Zmodo. These devices are purchased from popular retailers Walmart, Best Buy, and Amazon [8]. IoT reference architecture is a recent topic for research not much work has been done on it. There is a need to fully develop the privacy federated IoT security reference architecture that could help in developing a standard and to start implementing the privacy and security metrics from the root in the architecture of the internet of things. The diverse data generation and the utilization of the applications performing data collection, analysis, and prediction have increased the rate of privacy issues. There should be privacy by design framework embedded with the IoT architecture to address all the concerns relevant to it.

One of the major building blocks for IoT devices is the WSNs. These are ad hoc networks. The data is being gathered from the surroundings to deliver to its users. These consist of nodes that can detect, compute and communicate with the devices. Low power and Lossy Networks (LLN) are used in the IoT network. These are networks for constrained environments such as IoT which possesses the constraints of memory, energy, processing power. Hence Lightweight encrypted algorithms are used for securing the IoT ecosystems. These aspects are not used in conventional wireless networks [9].

This Internet of things has the potential to transform connectivity at any time to anyone from anywhere. These can connect to real-time environments and can process smart and intelligent communication and can make autonomous decisions. This IoT has the potential to assist our economies, transportation, environment, and health in a way that we never expected before [7].

1.1. PROBLEM STATEMENT

The Internet of things architecture has evolved over the years considering the critical issues and improving the stack of the architecture. A reference architecture is a path towards concrete architecture. A reference architecture is a recommended structure, a building block that integrates products and services. There are vulnerabilities in the IoT ecosystem considering the security and privacy issues. Considering the privacy of the IoT devices as the incidents arise in the form of hidden CCTV camera recordings in the houses of people, Systematic privacy flaws in the form of missing encryption certificates validations. This is due to the lack of standardization, there is no standard architecture for the Internet of things it is still in its infancy. The heterogeneous IoT network has many considerable issues such as the confidentiality of information and safety of user data. The use of a multitude of languages, protocols, and standards. The main problem is that we don’t know where all the data collected from these sensors, actuators lightweight IoT devices are stored. Whether it is being profiled or not? The heterogeneous nature of the IoT is a challenge to secure them. There has not been much work done on the privacy federation to the IoT reference architecture.

1.2. OBJECTIVES AND CONTRIBUTION

- Identifying the core requirements for the Internet of things.
- Breaking down the identified core requirements into quantifiable metrics.
- Identifying the privacy and security requirements through standards.
- Analysis of existing reference architectures based on identified metrics.
- Shortcomings of the existing reference architectures are identified.
- Federating privacy and security into the reference architecture.
- Proposed privacy federated IoT security reference architecture.
- Validating the proposed privacy federated IoT security reference architecture.

1.3. THESIS OUTLINE

The thesis is divided into the following chapters:

- Chapter 1: This chapter consists of the introduction to the topic, problem statement, objectives, and contribution.
- Chapter 2: This chapter consists of the background study, literature review, and a detailed description of the recently proposed and existing IoT reference architectures.
- Chapter 3: This chapter consists of the identified requirements for the IoT reference architectures, quantifiable metrics, and analysis of existing reference architectures based on these metrics.
- Chapter 4: This chapter consists of the proposed privacy and security federated reference architecture along with its validation.
- Chapter 5: This chapter consists of the conclusion and future work directions.

1.4. SUMMARY:

This chapter consists of a detailed introduction to the IoT and business drivers in the industry. It highlights the importance of IoT in the future in terms of capital. It highlights the evolution of the IoT and raises questions on its vulnerabilities and risks associated with the users in terms of privacy and security. It gives an overview of the standardization bodies working on the building blocks of the IoT to homogenize the heterogeneous nature of the IoT and come up with the standard architecture. This chapter also describes the problem statement and its objectives and contribution. It gives the outline of the thesis. In the next chapter, we will discuss the preliminaries, background study, and detailed literature review elaborating the existing reference architectures in detail.

CHAPTER 2

2. BACKGROUND STUDY AND LITERATURE REVIEW

This chapter will cover two broad categories of the research consisting of the preliminaries, background study, and literature review done. The background study and literature review will highlight the domain-specific knowledge. This study will help to identify the research gap in the area. A detailed description of the recently proposed and existing reference architectures is given in the chapter.

2.1. BACKGROUND STUDY & LITERATURE REVIEW

This paper gives a Systematic literature review (SLR) on the existing IoT architectures their evolution and concerns regarding security and privacy. This explains the evolved phase right from its initial phases of 2008 to 2018. This comparison amongst the evolved architecture until 2018 defines the architectural stack, challenges or covered issues, the techniques used, and consideration of critical issues of security and privacy. This review elaborated the findings that initial architectures did not convey a comprehensive meaning of IoT that should describe its nature. The recent architectures give a comprehensive meaning of IoT explaining the data transmission, data collection, data processing, and data dissemination. It defines that the architecture stack has improved addressing the challenges like scalability, interoperability, extensibility, and management. Findings in this work also disclose the research gap about privacy that none of the evolved IoT architectures addresses the privacy concerns in detail which are considered to be a critical factor in its sustainability and success [10].

This work analyzes the IoT reference, architecture models. It highlights the importance of a comprehensive architecture model that should homogenize the heterogeneity in the IoT. Division of the functionality to the elements and data flow is known as a reference model. These requirements are controlled by the reference architecture to form the superset of functionalities, structures, mechanisms, and protocols. The requirements on which the analysis is done are defined by different consortia and manufacturers. The ITU-T reference model and some areas have been highlighted that need to be addressed in upcoming work [11].

There has been very little work on how to protect the sensor data after the transmission. To protect such data against malicious attacks and unauthorized access there is a need for a privacy-preserving mechanism. To do that we need privacy-aware IoT frameworks to ensure privacy and security of data collection, transmission, and usage. This work introduces the privacy-preserving architecture for the IoT and also converges it to cloud computing. It proposes an efficient privacy-preserving deep learning mechanism in a privacy layer and uses Physical Unclonable Functions (PUF) for identity management and authentication [12].

IoT standardization bodies like ISO/IEC and SWG5 of the joint technical committee 1 (JTC1) submitted the report on the IoT reference architectures and frameworks. AdHoc Group (AHG) was created by SWG5. This report resulted in the proposal of a layered IoT reference architecture by the Korean Study Group [13].

The standardization efforts started in china in 2010. China Communications Standards Association (CCSA) is the main organization for standardization. This paper gives a holistic view of work on IoT development, research, and development, policies, and applications. It proposes a reference model for IoT which consists of the sensing layer, network, business, and application layer [14].

In general, IoT devices have limited computation power and storage capacity but on the contrary, cloud computing (CC) has virtually managed unlimited computational power and storage capacity which rely on the sharing of resources. Therefore the integration of Cloud computing and IoT seems to be a promising solution. This paper does the comparison based on performance on three main cloud platforms (Google Cloud Platform, Amazon Web Service, and Microsoft Azure). These cloud platforms have proposed the reference architectures of the IoT integrating the cloud computing therefore the analysis based on metrics is also done. This work does not declare a winner amongst the three but provides a tool for the developers [15].

Defines the strategies that we can apply for data-driven IoT architectures. These strategies guide us towards the development, complexity and ensure that the IoT solutions remain scalable, robust, and flexible. These strategies adopt a layered architecture, security by design, automate operations and follow a reference architecture [16].

For the construction of concrete architecture the seventh framework program (FP7) research project EU has reinforced the reference architecture proposed by Martin Bauer et al. For the construction of a solid building block architecture this reference architecture gives us a high-level perspective and views. Perspectives are a set of qualities and views that are the vision of architecture from different angles during design and implementation [17].

This work analyzes how the legal principles support the implementation of Privacy Enhancing Technologies (PETs) at the layer of the IoT architecture model to fulfill the requirements of the individuals who will interact with the IoT ecosystem. This demonstrates to us the privacy legislation mapping with the principles of privacy which drives the design of important privacy-enhancing technologies to be incorporated in the IoT architecture stack [18].

2.2. EXISTING REFERENCE ARCHITECTURES

2.2.1. INTEL: [19]

The reference architecture of Intel is shown in the figure below. It is layered where the yellow blocks are user layers and the blue blocks are the major runtime layers. The light blue layer is for the developers.

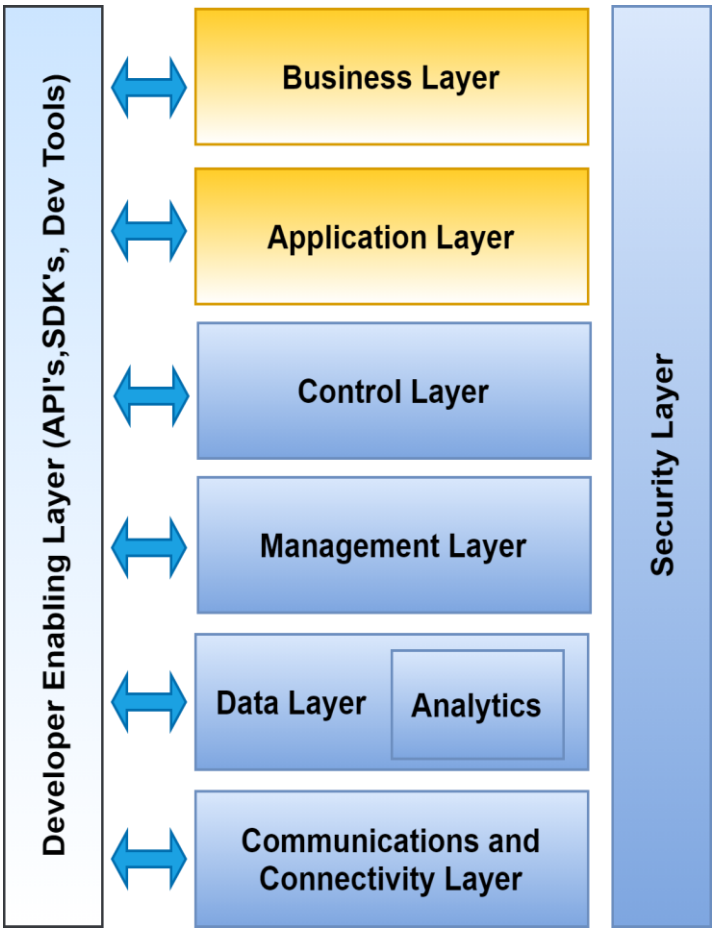

Figure 1 Intel IoT Reference Architecture

2.2.1.1. *Communications and Connectivity Layer:* For the data ingestion and device control the Intel IoT reference architecture uses broad protocol normalization and control systems. It uses multi-protocol data communication between the gateways and devices. It uses three types of networks Proximity networks (PAN) Local area networks (LAN) and Wide area networks (WAN). PAN/LAN usually connects to the edge nodes of the sensors, actuators, devices, control systems, and assets. PAN's are more constrained in comparison with the LAN by antenna distance and battery life. WANs can be the corporate networks for the internet, 4G/5G mobile networks, or satellite networks.

2.2.1.2. *Data Layer with Analytics:* It provides customer value through data analytics and controlled closed-loop systems. These analytics are distributed across the cloud, gateways, and smart end point devices. The advantage of this distribution is that it provides the flexibility to provide time-critical and computation-intensive applications.

2.2.1.3. *Management Layer:* This management layer consists of the managed devices which consist of a management agent that executes management in the device. This is managed by a web-based user interface. A device cloud is a system that manages a small to a very large number of end-point devices. Device cloud consists of the main management functionality which consists of update applications and operating systems, discover registers, and provision new devices, manage data flow i.e destination and storage policy, upload or stream data, define and manage alarms and notifications, manage organizations users and access rights, etc.

2.2.1.4. *Control Layer:* This layer separates the management layer into the management plane and control plane. This includes control objects, policies, and API. This layer can move-of the device for the cloud or remote control which is one of the main requirements for the software-defined network (SDN).

2.2.1.5. *Security Layer:* Both software and hardware level of security is important to achieve the desired level of security. Security is a process, not a product. It provides the end to end protection. This Intel IoT reference architecture provides a security software product portfolio for the developers to deliver interoperable and scalable solutions. This security is implemented at three phase's endpoint device level, network level, and cloud level. The end point device level protects the identities. The privacy of the users should not be breached and the devices should guarantee authentication. The network-level security should secure the traffic, application, and data through the wired and wireless network similarly the cloud-level security should secure the data centers and public cloud environments.

.

2.2.2. MICROSOFT AZURE: [20]

The Microsoft Azure IoT reference architecture is shown in the figure below. This architecture presented is based on the cloud-native, micro service, and serverless base. These IoT subsystems should be independently deployable and be built as discrete services. This allows us greater scale, flexibility in updating the systems and gives us the flexibility to choose the right technology on a sub-system basis.

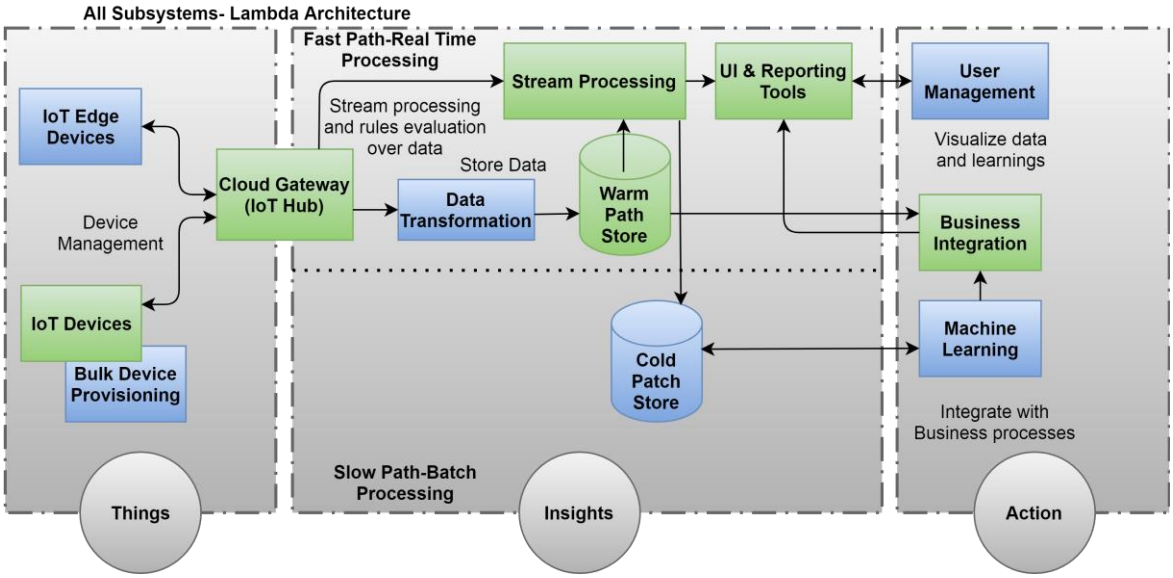

Figure 2 Microsoft Azure IoT Reference Architecture

2.2.2.1. *Devices, Device Connectivity, Field Gateway (Edge Devices), Cloud Gateway:* The IoT edge devices are connected through the field gateway. This connection results in edge intelligent capabilities. The raw telemetry and the aggregation of the data are enabled. The connectivity patterns are direct devices connected to the cloud gateway, connection via field gateway. This option is very useful for the devices that are using industry standards like Constrained Application Protocol (CoAP), resource-constrained devices not capable of hosting TLS/SSL stack, and short-range devices like Bluetooth, ZigBee. Connectivity via a custom cloud gateway requires some form of custom processing for the devices that need a translation of the protocol. Connectivity via the field gateway and custom cloud gateway. Some situations lead and require the integration of field and cloud gateways using VPN's network tunnel or application relay service.

2.2.2.2. *Data Transformation:* It manipulates and aggregates the telemetry stream either before or after it is received by the cloud gateway, IoT hub. This is done by converting the binary stream data to JSON. We suggest the IoT hub integration with Azure functions for the translation of the telemetry data before its receiving to the IoT hub.

2.2.2.3. *Machine Learning:* This subsystem in the architecture is intelligent and learns from data and experiences to respond without explicit programming. Predictive maintenance is programmed through machine learning. Azure ML fulfills all such requirements.

2.2.2.4. *User Management:* This layer subsystem allows the user management and capabilities for the users like command and control, upgrading the firmware, and user application capabilities.

2.2.2.5. *Data Flow and Stream Processing:* According to the scenarios the data records go through different stages which are processed by concurrent tasks. The stages are storage, routing, analysis, and action/display. Memory caches, temporary queues, and permanent archives include in storage Routing involves the dispatching of data records to the end points analysis and actions. The analysis put the data records through certain conditions that can result in different output data. For example, the input of telemetry Avro returns the output in encoded JSON format. These records are available for display and actions like emails, instant messages, incident tickets, CRM tasks, and device commands.

2.2.2.6. *User Interface and Reporting:* It is basically for the reporting and user interface that includes a website, mobile, or desktop app. This UI can provide access and visualization for the data analysis, discovery through registry and command and control capabilities. It provides interaction with the live dashboards.

2.2.2.7 *Business System Integration:* This layer is responsible for the downstream business like CRM, ERP, and Line of Business (LOB) applications. Service billing, customer support, dealers, service stations, third-party data sources, time, and job tracking all include in this. The IoT collaborates with the standard software solutions through business connectors or EAI/B2B gateway capabilities. The end-users will interact through this layer in B2B or B2C scenarios.

2.2.2.8 *Warm Storage, Cold Storage:* The data should be available in the database within seconds when the data is absorbed into the cloud from the device. Warm storage stores the easily accessible data to the last known state per device. The data stored in the database may be in the raw form or aggregated form or may be both. If the ingestion rate is high then a high ingestion database may be required. Keeping all the data warm storage with low latency, high throughput, and query capabilities Microsoft IoT azure split the data into cold and warm storage paths. This optimizes the lower storage costs. The cold database storage might not be as quick or frequent but can be very helpful for reporting, analysis, and machine learning.

2.2.3. MONGO DB: [21]

Apart from the databases, storage, pre-aggregation, and advanced analytics using aggregation framework. Mongo DB plays an essential role in the IoT solution and presents the reference architecture for the IoT. The architecture is shown in the figure below.

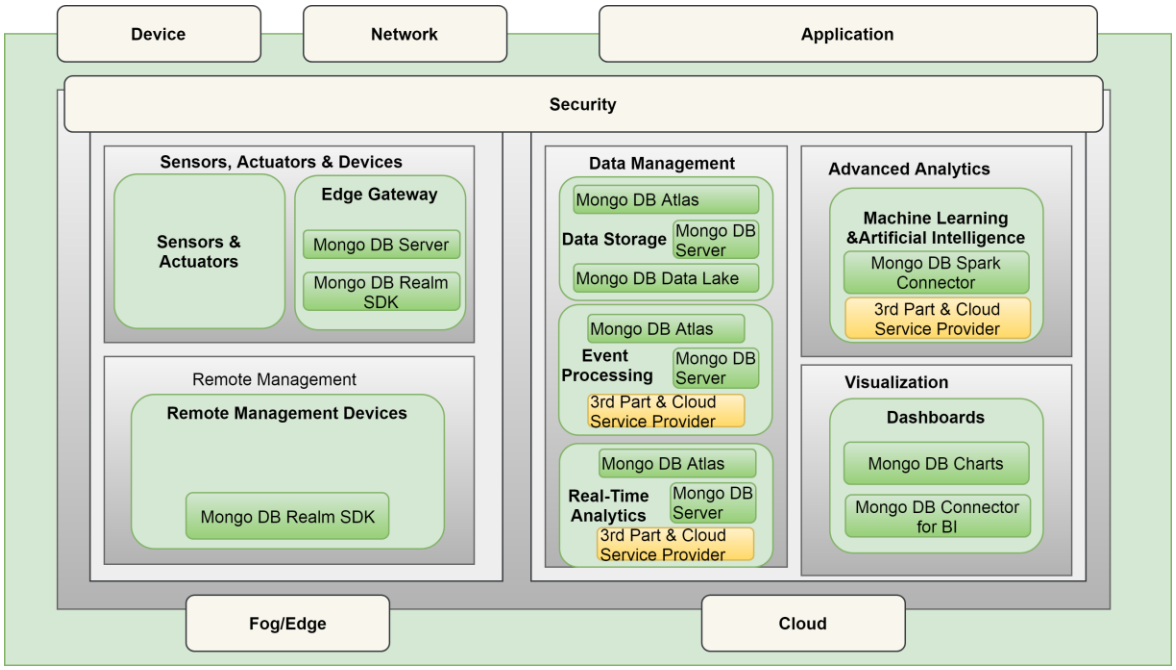

Figure 3 Mongo DB IoT Reference Architecture

2.2.3.1. *Edge Gateway:* These are high-powered devices based on the same network as the sensors and communicate with them. These edge gateways are used for data collection, filtering, offline data storage, analysis, and local aggregation. These can also communicate with the backend systems for analytics and data storage. Mongo DB gives us two options Mongo DB realm SDK and Mongo DB server to accomplish the edge needs. Mongo DB realm is a development library that consists of a lightweight realm database and supports both 32 bit and 64-bit architectures. This realm SDK allows the uni and bi-directional sync between the edge gateway and Mongo DB realm. If the edge gateways are 64 bit and conform to the requirements then we can use the Mongo DB server directly on the edge gateway.

.

2.2.3.2: *Remote Management:* This remote management in IoT is used to monitor and manage the environment. NoSQL database is used for application development and the object model. The management devices are up to date with the processed events so that the end-users can see the alerts on the mobile devices and respond to them in a real-time environment.

2.2.3.3: *Data Storage:* Things in the IoT generate a huge amount of data that need to be stored for both real-time and analysis. This reference architecture by Mongo DB terms Mongo DB as the best platform for IoT data storage. Mongo DB gives access to both real-time and batch-based workloads against the MongoDB cluster. The benefit is that we don't need to ETL the data into another system to do batch analysis. The Mongo DB ATLAS is a service that allows storing data in S3 buckets. S3 is basically for the archived data. We have to query the S3 buckets just like the Mongo DB database running Mongo DB query.

2.2.3.4: *Real-Time Analytics:* It involves the processing of high volumes of data connected to the assets in real-time. This type of analytics helps us to take immediate action within seconds or minutes so do the responses. This allows the organization to take immediate action or flag the event and follow up later whenever it is urgent.

2.2.3.5: *Stream Analytics and Event Processing:* Stream analytics perform queries and actions to the real-time data. Mongo DB can be used as the data source and data destination for the streaming platforms like Apache Spark, Mongo DB. We don't have to query the entire data set. Mongo DB enables the applications to use event-driven processing to respond to the changes.

2.2.3.6: *Advance Analytics:* For cost-saving and to prevent the system from failure whether it is a production unit this layer is very important. Machine learning which includes advanced analytics can predict when the component will fail which could result in system failure therefore one can take the preventive measures right on time. Apache Spark is a cluster computing system. It provides us with the APIs in Java, Scala, Python, and R which supports libraries like MLlib used for machine learning, Graph X, and Spark for graph processing and streaming respectively.

2.2.3.7: *Visualizing IoT Data:* Mongo DB provides us custom dashboards as well as third-party platforms for the visualization. These can be helpful in the form of reports and graphs. These can be used to indicate the performance. Mongo DB charts can visualize the Mongo DB data, we can enable the charts in the web portal and add a data source to visualize charts in the UI. These can represent complex data, arrays, and subdocuments. Business Intelligence tools are also used for reporting and analytics like Tableau, Qlik View, and Microsoft Excel. Mongo DB BI connector is used as a translation layer to receive queries from the reporting tools and pushing down these SQL queries to Mongo DB Query Language.

2.2.3.8. *Security:* Mongo DB Atlas has been incorporated and audited to meet the privacy and compliance standards like SOC type 2, Privacy shield, etc. This supports authentication mechanisms like SCRAM, X.509 authentication, LDAP proxy, and Kerberos. For access control, it follows the role-based access control. For network protection and encryption it uses and supports TLS/SSL network encryption.

2.2.4. IBM: [22]

The figure below shows the Reference architecture presented by IBM with cloud components. This is three-tier architecture consisting of edge, platform, and enterprise tiers. The edge deals with data collection and transmission. The platform tier deals with analysis, API management, and visualization. The enterprise tier deals with enterprise data, enterprise user directory, and applications.

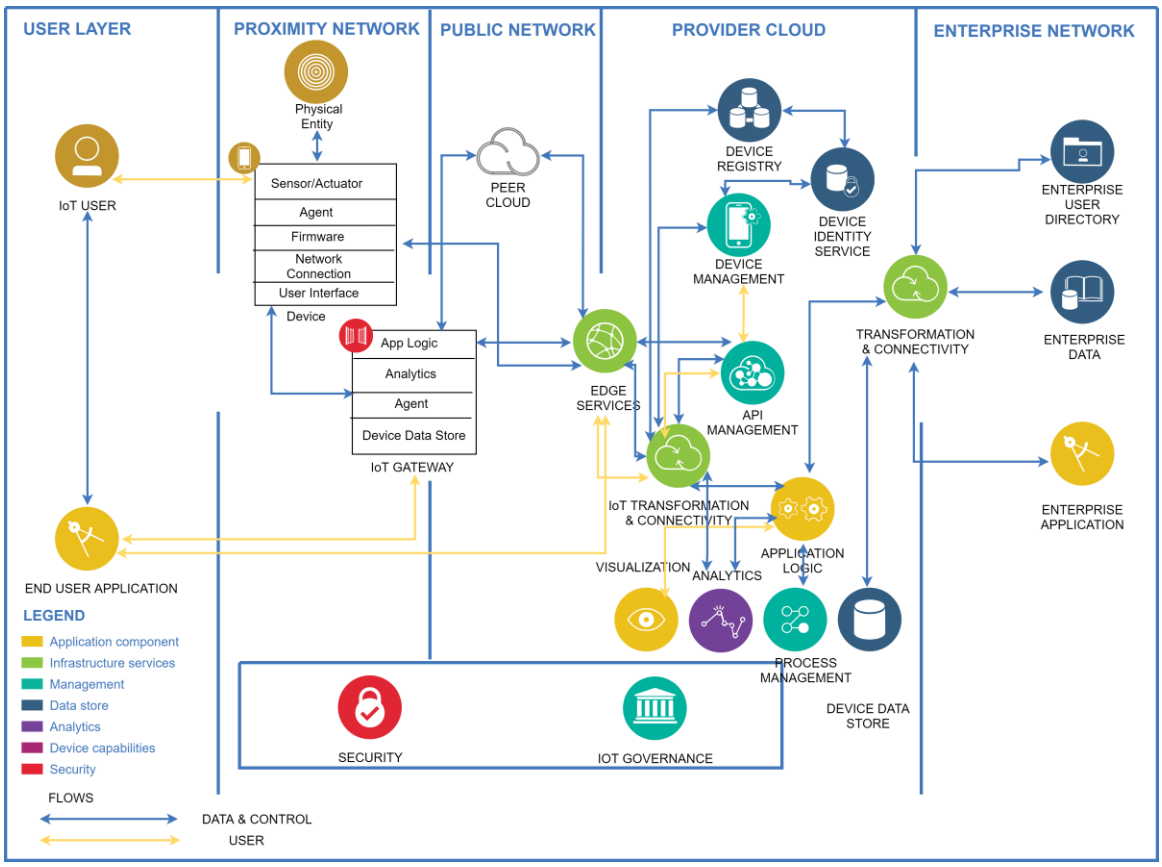

Figure 4 IBM IoT Reference Architecture

2.2.4.1. *User Layer:* This layer consists of two types of users the IoT users and end-users. The IoT users are persons or automated system that allows the user applications to achieve the goal. The end-user application is that a user uses on smartphones, tablets, and specialized IoT devices.

2.2.4.2. *Physical Entity:* These are the things that subject to sensor measurement and actuator behaviors. This layer differentiates the entities and devices that sense them and act on them.

2.2.4.3. *Device:* The device layer consists of sensors, actuators, firmware, network connection, and user interface. This also includes Agent that supports the device management protocol that gives the remote management capabilities. The software that gives the control, monitoring, and data manipulation is known to be firmware.

2.2.4.4. *IoT Gateway:* Gateway is an essential and decoupling element. It connects one or more devices with the network and the internet. The IoT devices have usually low power and computational resources which results in limited network connectivity. The local network allows the devices to communicate with the local IoT gateway. The gateway also provides us the operational efficiency.

2.2.4.5. *Peer Cloud:* It's a third-party cloud system that provides services to bring data to the IoT platform. These peer clouds can contribute to the IoT systems and also provide the capabilities in the IoT architecture.

2.2.4.5. *Edge Services:* These edge services include a Domain Name System that translates the URL of the web resource to the IP address of the system that can deliver the resource. Content Delivery Network (CDN) supports the end-user applications to make sure that the content is available to the users having low latency. The servers are deployed to minimize the response time for geographically distributed users. A firewall controls and filters the communication allowing only traffic to pass that meets the set of policies and blocking others. This can be implemented in the form of separate hardware. Load balancers are used to provide the maximum throughput, minimum response time, and increase the reliability of applications. These can balance the loads locally and globally.

2.2.4.6. *IoT Transformation and Connectivity:* This enables secure connectivity from the IoT devices. Its job is to route the high volumes of messages to the right components. The key capabilities in this domain are secure connectivity, scalable messaging, and scalable transformation.

2.2.4.7. *Application Logic:* This is an event-based model that includes trigger, action, and rules-based programming IoT application logic. It controls the workflow. It can be written in many languages but the IBM blue mix allows Node.js, Java, Websphere Liberty Profile, Swift, and Python.

2.2.4.8. *Analytics:* The discovery and communication of meaningful patterns found in the IoT data to predict and improve business performance. This includes an Analytics data repository that supports legacy, new, and streaming sources as well as output from the streaming analytics. Cognitive capabilities create intelligent systems that the self learns and adapts for augmented human intelligence. Actionable insight drives actions that are used by the business applications stored in the data repositories. Streaming computing processes the time-sensitive and continuous data streams from sensor-based monitoring devices and messaging systems.

2.2.4.9. *Transformation and Connectivity:* It enables secure connections to enterprise systems. It can filter, aggregate, and modify the data as it moves between cloud, IoT, and enterprise systems. This includes Enterprise secure connectivity, Transformation, and Enterprise data connectivity.

2.2.4.10. *Enterprise Data:* Consists of the metadata about the data and system of record for enterprise applications. This sort of data flows directly to data integration or the repositories providing the feedback loop to the analyzed IoT system. These IoT systems store raw, analyzed, and processed data in the enterprise data elements.

2.2.4.11. *Enterprise Applications:* To address the business goals the enterprise applications consume cloud data and analytics. These can be updated from the enterprise data or applications. These consist of customer experience, new business model, financial performance, risk analytics, economics, operations, and fraud.

2.2.4.12. *Security:* This also addresses the importance of the security layer in the reference architecture the areas to consider are identity and access management, data protection, security monitoring, analysis, response, system application, and solution lifecycle management.

2.2.5. SECURE AND SAFE INTERNET OF THINGS SERIoT (ISO/IEC 30141): [23]

The SerIoT is a project funded by European Union's Horizon 2020 Research and innovation program. The reference architecture presented by SerIoT is followed by the ISO/IEC 30141 standard. The architecture is shown in the figure below. This architecture targets security-driven solutions to address the threats.

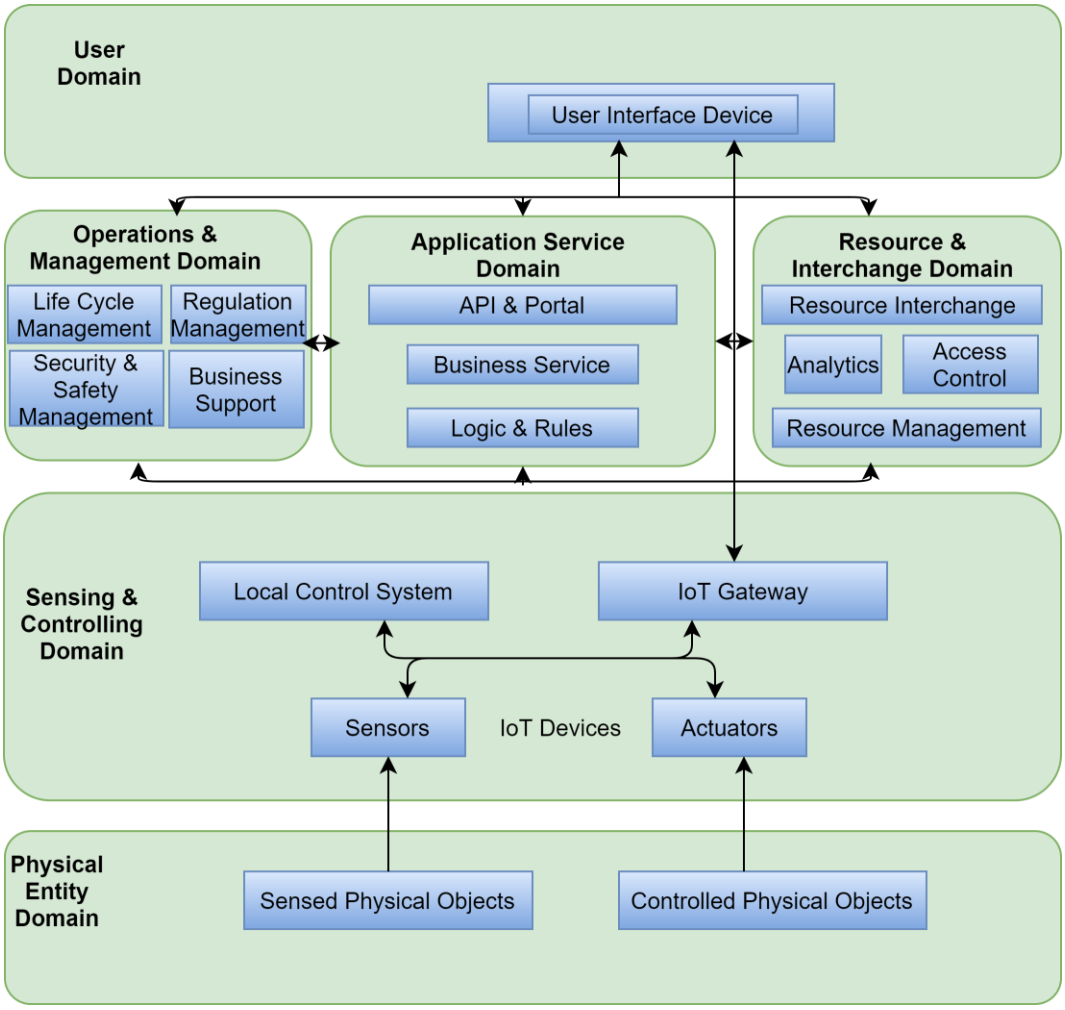

Figure 5 SerIoT Reference Architecture

2.2.5.1. *The Physical Entity Domain:* This domain consists of the sensed and controlled physical objects in the IoT system. It consists of the physical and virtual entities which are responsible for monitoring, sensing, and controlling.

2.2.5.2. *The Sensing and Controlling Domain:* The SCD provides critical information about the environment to the other domains in the IoT system. Such type depends upon the network communications also known as the proximity networks. These proximity networks use specialized protocols. The wide area networks (WAN) connect these proximity networks with the internet. This needs communication like transmission and receiving with the software services. A local control system should be deployed to perform time-sensitive critical data processing to control the objects.

2.2.5.3: *The Operations and Management Domain:* It contains a set of functions responsible for managing, monitoring, and optimization systems and their performance in real-time. Managers and system operators maintain the health of the system.

2.2.5.4: *The Resource and Interchange Domain:* In terms of resources it interacts with entities, applications, services, and systems. This resource can be physical or monetary. This includes the processing of data which includes data assurance, quality, transformation, distribution, and storage.

2.2.5.5: *The Application Service Domain:* It consists of business services and service providers. These service providers interact with the users as well as with the sensors and actuators to gain the data from physical objects.

2.2.5.6: *The User Domain:* This consists of the stakeholders and actors in the IoT system. It can also be an individual, household, society an organization, or government department.

2.2.6. CISCO: [24]

Cisco proposes the reference architecture which consists of seven layers. They have proposed the reference architecture which can lead to standardization worldwide. The architecture is shown in the figure below.

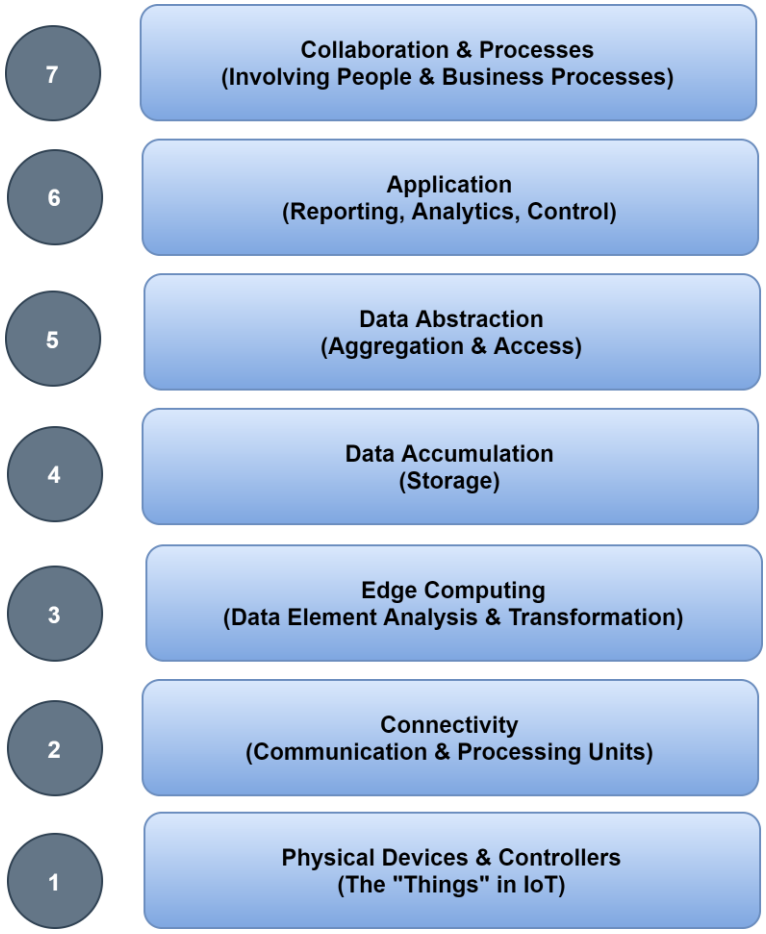

Figure 6 Cisco IoT Reference Architecture

2.2.6.1. *Physical Devices & Controllers:* This is layer one that consists of sensors, devices, machines, and things in IoT. These devices can be diverse like some will be like the size of the silicon chip and some might be very large. The IoT should be able to support the entire range and not be limited to a specific range. These devices are capable of analog to digital conversion, generating the data, and being queried.

2.2.6.2. *Connectivity:* This layer 2 connectivity function is to transmit the information timely and reliably. This transmission can be between the devices and the network, across and between the networks. The connectivity consists of protocols, switching, and routing. Security and network analytics.

2.2.6.3. *Edge Computing:* This layer focuses on high-volume data analysis and transformation. This layer involves data evaluation, formatting, expansion, distillation, and assessment. It also deals in packet and content inspection, thresholding, and event generation.

2.2.6.4. *Data Accumulation:* The applications do not need to process the data at network wire speed. The data is at rest in the memory or disk. This layer captures the data and puts it on the rest. These applications usually access the data when necessary. The event-based data is converted to query-based for processing. It also reduces the data through filtering.

2.2.6.5. *Data Abstraction:* It abstracts the data interface for applications. This layer creates schemas and views of the data according to the application's needs. Combines the data from multiple sources. To fulfill the client applications it filters, projects, and reformat the data also it also reconciles differences in data shape, semantics, access control, and security.

2.2.6.6. *Application:* The reference architecture of the IoT does not strictly define the application. It varies based on device data and business needs. The example of the applications can be ERP or business applications, mobile applications, business intelligence reports, and analytic applications.

2.2.6.7. *Collaboration and Processes:* This Internet of Things includes people and processes. People should be able to collaborate and communicate to make use of the information.

2.2.7. IOT ARCHITECTURAL REFERENCE MODEL (ARM): [25]

The representation of the IoT ARM and its functional view is given in the figure below. This proposed architecture in the seventh framework program a research project by EU helps us towards the construction of concrete architecture.

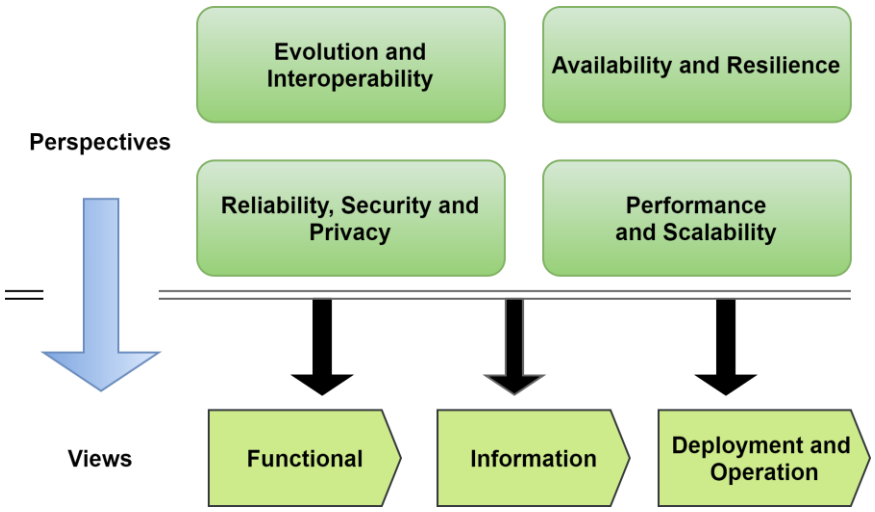

Figure 7 IoT Arm View

2.2.7.1. *Functional View:* The functional view of the IoT ARM reference architecture is shown in the figure below. It consists of nine functional groups and components.

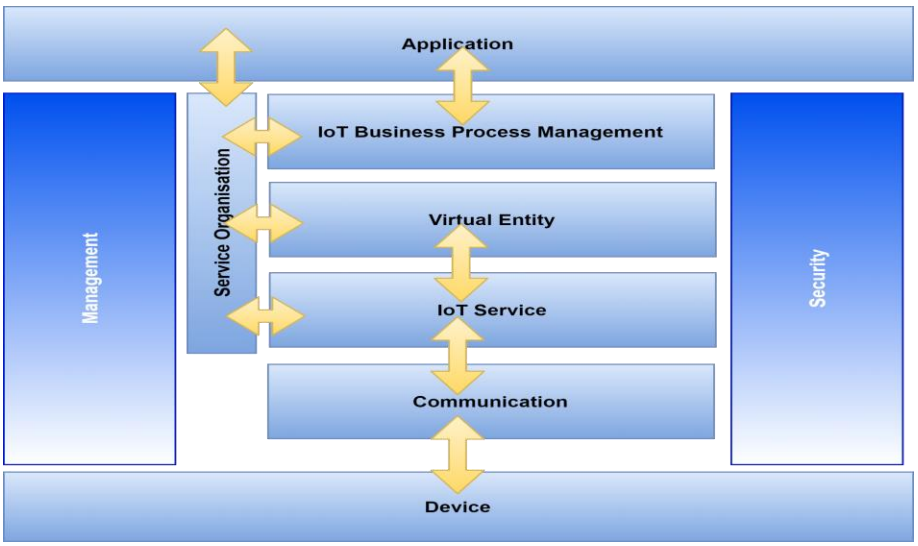

Figure 8 IoT Arm Functional View

2.2.7.2 *Information View:* To exchange the information amongst the external entities the smart objects interact with each other. Information between the entities is also handled and also keeps the track of the lifecycle.

2.2.7.3. *Deployment and Operation View:* This investigates how components communicate with each other that encircle quality, requirements, applicability, and architectural tactics.

2.2.8. KOREAN STUDY GROUP (KSG): [13].

The Korean study group has presented this reference architecture for the IoT from two view point that is functional and communication. Six blocks are present in the functional representation of the architecture. The figure below represents the functional view of the IoT reference architecture proposed by KSG.

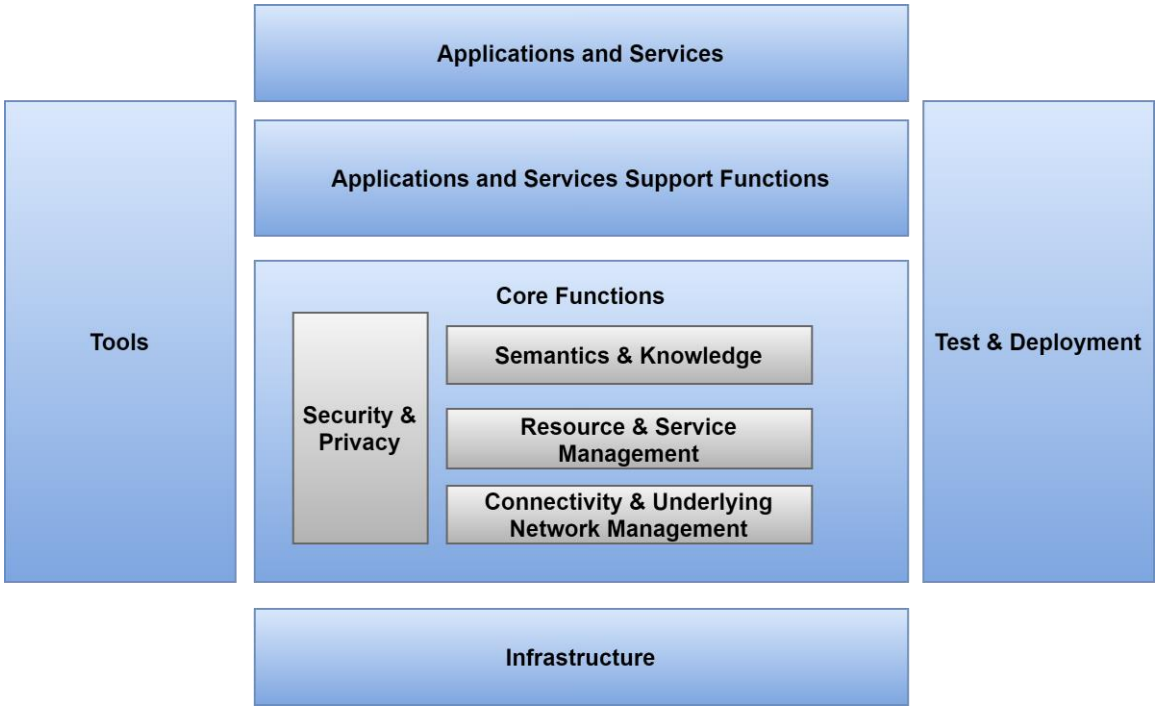

Figure 9 Korean Study Group IoT Reference Architecture

2.2.8.1. *Communication View Point:* This viewpoint consists of the interoperability and connection method. The connection method consists of the IoT devices which are directly connected with the internet and some of them are connected through the gateway. Due to the implementation issues, some are connected through the gateways to avoid such errors even with the ability to connect directly. The other devices communicate indirectly through intermediate nodes. This method also uses IoT platforms which are a layer between the devices and the services. The IoT services use the objects physical and virtual to provide different services.

The interoperability view point consists of the device-to-device communication which connects directly, a device to the platform allows the devices to migrate to any other platform without incorporating the changes in the functionality, platform to platform works together to provide services and cross-domain data to provide the services.

2.2.8.2 *Functional View Point:* This viewpoint consists of six blocks.

2.2.8.2.1 *Infrastructure:* This consists of the basic structure containing hardware, network, and system resources that are necessary for the core operations.

2.2.8.2.2 *Core Functions:* As shown in the figure above this contains knowledge, semantics, resource management, connectivity, and network management integrating security and privacy concerns.

2.2.8.2.3 *Application and Services Support Functions:* This layer provides an abstraction to the components and their core functions making it easy for the upper layer.

2.2.8.2.4 *Tools:* For the development of new applications this layer provides the tools.

2.2.8.2.5 *Test and Deployment:* This layer deals with the testing of the developed IoT system before becoming available for the users.

The detailed core function representation of the reference architecture by KSG is shown in the figure below.

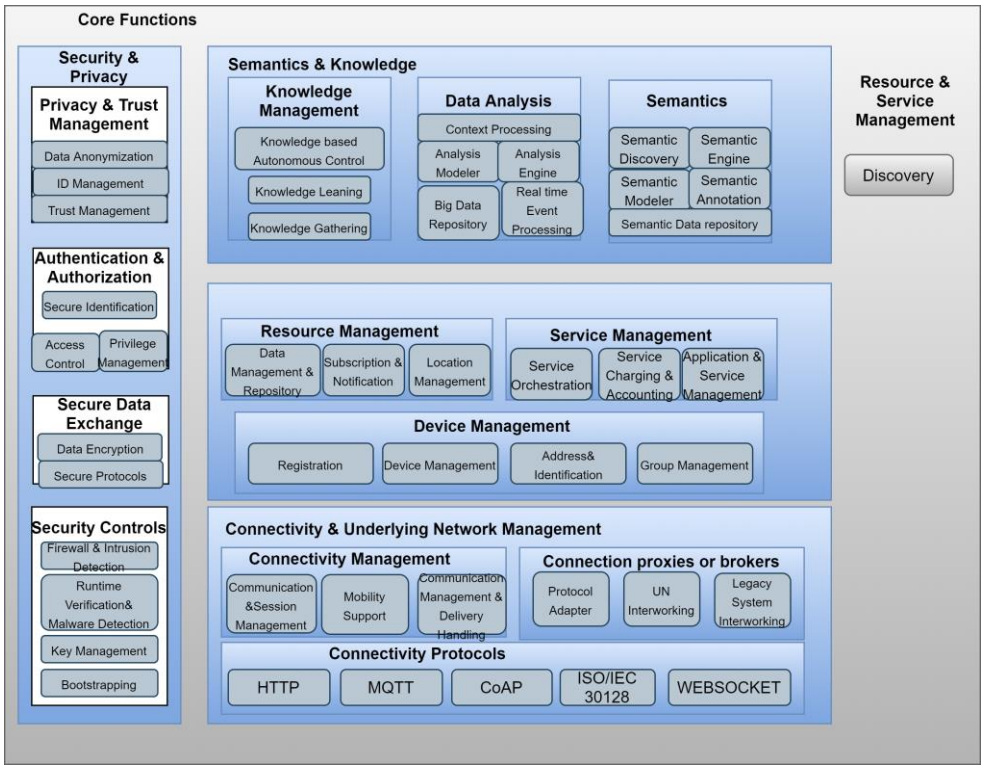

Figure 10 KSG Detailed Core Functions

2.2.9 CHINA COMMUNICATIONS STANDARDS ASSOCIATION (CCSA): [26]

The representation of the IoT reference architecture proposed by the china communications standards association is shown in the figure below.

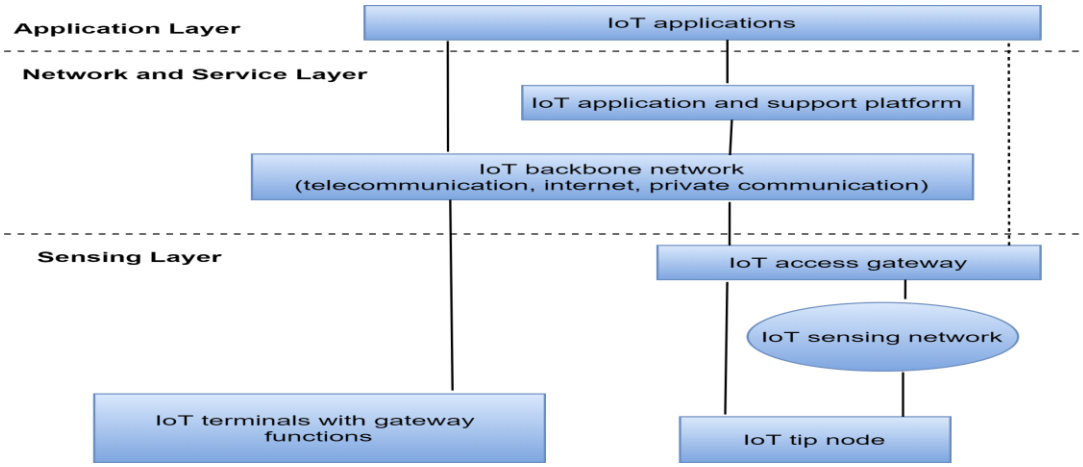

Figure 11 CCSA IoT Reference Architecture

2.2.9.1 *Sensing Layer:* This layer consists of the sensors, controllers, RFID readers, and location-sensing devices to the network layer. Modularization is supported by this layer. The components in this layer can self-adapt, operate intelligently and configure by themselves.

2.2.9.2 *Network and Service Layer:* This layer consists of the resource administration platform, application and support platform, and backbone network. This layer also supports the control functions like access control, authorization, authentication, and mobility.

2.2.9.3 *Application Layer:* This layer deals in the modularization of common functions which can be used in the development of the applications by the developers.

2.2.10. WSO2: [27]

The reference architecture of the IoT presented by the WSO2 consists of five horizontal and two vertical layers. The cross-cutting vertical layers consist of device manager, identity, and access management.

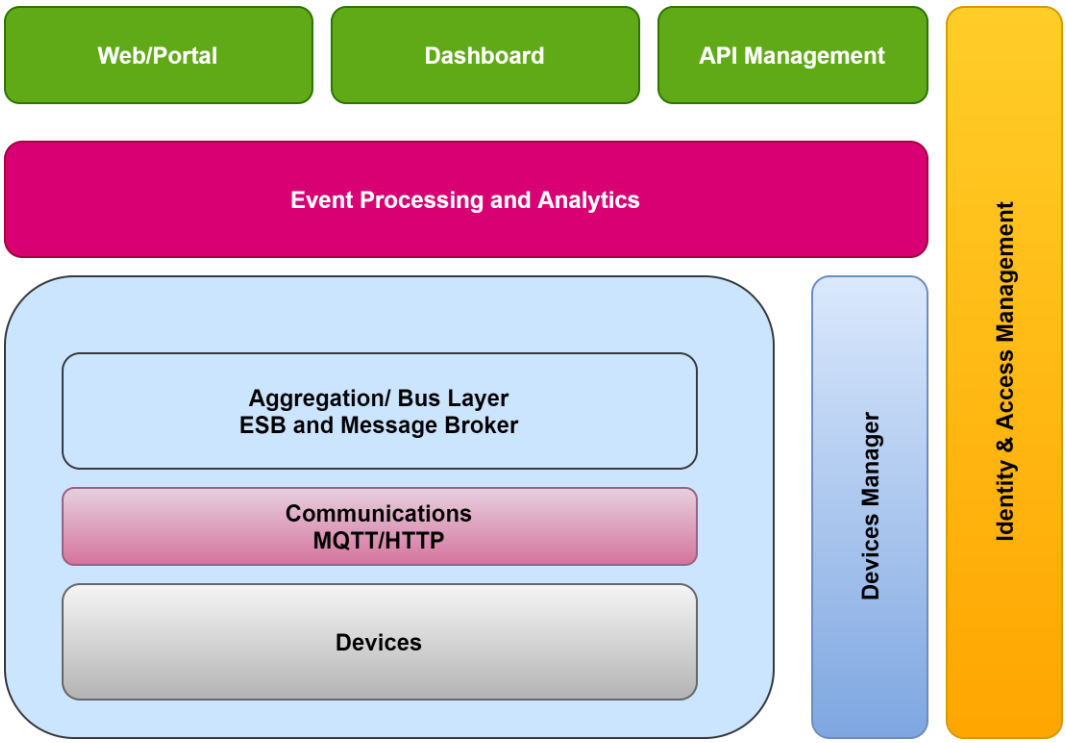

Figure 12 WSO2 IoT Reference Architecture

2.2.10.1. *Device Layer:* The device layer consists of the devices that can communicate with the internet whether it should be a direct or indirect connection. The connection of Arduino with Arduino Ethernet, Raspberry pi connected to the Wi-Fi or Ethernet, or Intel Galileo connected to the Wi-Fi or Ethernet are all examples of the direct connections. While the connection of the ZigBee through the gateway or Bluetooth connection through a mobile phone are examples of indirect connections. This architecture suggests having a unique identifier that should not be able to get modified as well as OAuth2 Refresh and Bearer token stored in EEPROM.

2.2.10.2 *Communications Layer:* The connectivity of the IoT devices is managed by this layer. The most commonly used protocols for communication are HTTP/HTTPS, MQTT 3.1/3.11, and Constrained Application Protocol (CoAP). These protocols have their strengths and weaknesses.

2.2.10.3 *Aggregation/Bus Layer:* This layer is known to be important as it brokers and aggregates the communication. This layer can incorporate the legacy protocols. The bus layer helps in the correlation and mapping of the device id to the owner's id. This layer incorporates the policy enforcement point (pep) for policy-based access.

2.2.10.4 *Event Processing and Analytics Layer:* The events are taken from the bus layer and are processed. The data is stored in the database. It also does analytics on the data that is coming from the aggregation layer.

2.2.10.5 *Client/External Communication Layer:* This layer utilizes all the functionalities like web-based portals to communicate with the devices, dashboards, and APIs that need to communicate with the systems outside the network.

2.2.10.6 *Device Management:* This layer has two components the server-side that communicate with the devices through protocols and give control of devices at both individual and bulk levels. This layer must work with the identity and access management layer and also maintain the identities of the devices to map them to their owners.

2.2.10.7 *Identity and Access Management:* This layer provides the following services Oauth2 token issuing and validation, the identity services like SAML2, SSO, and OpenID. LDAP, policy management, and access control.

2.2.11 GOOGLE: [22]

To connect, store, process, and analyze the data both at the edge and cloud Google Cloud Platform (GCP) possesses the tools. It has three essential components device, gateway, and cloud.
In this reference architecture, the device can be hardware or software and can be able to connect directly or indirectly to the internet. The reference architecture of google is shown in the figure below.

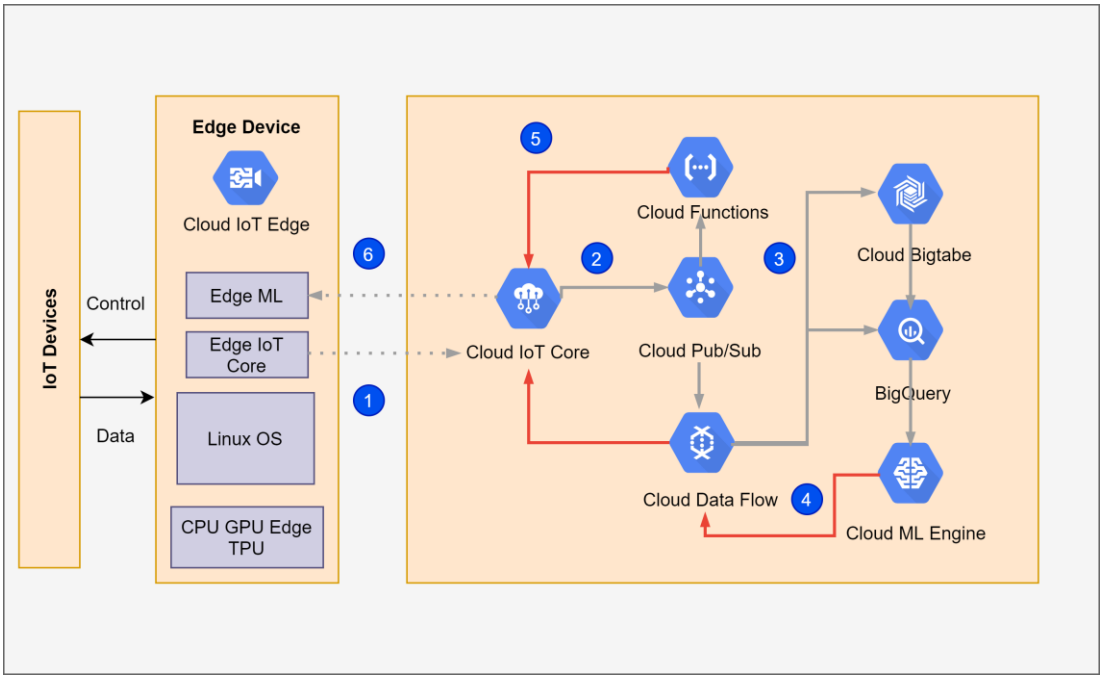

Figure 13 Google IoT Reference Architecture

The responsibility of the gateway is for the devices that are not directly connected to the internet for the cloud services. The gateway processes the data on behalf of a group of devices. The data is collected by the devices and sent to the cloud platform by the gateway.

The data is transmitted to the cloud IoT core. The devices that are using the MQTT protocol send the data to the same global endpoint regardless of the source region or location.

The data is sent to the Cloud Pub/Sub after receiving through Cloud IoT core. Cloud pub/Sub is a message queue and event broker. The data processed through Cloud IoT core or from Cloud Pub/Sub can take several different paths.

The Cloud Machine learning engine is used to anonymize the data storage on google cloud storage, the training data is used to refine the models and keep the fine-tuning.

The control configuration in the google IoT reference architecture allows the data to be sent back to the IoT devices by the cloud IoT core.

This reference architecture incorporates edge computing. The benefits of edge computing consist of fast response times reducing the latency and roundtrips. Unconstrained by connectivity limitations, regardless of the limitations in the connectivity edge devices can locally store and process the data to maintain reliability in the operations. Compliance with strict privacy requirements. It is very difficult to avoid the data that is being sent from IoT devices to the cloud but this edge technology can send only needed data through filtering of sensitive information.
Cost-effectiveness, the cost of network bandwidth, data storage, and computational power can hinder the customers from deploying solutions. The use of edge computing can help businesses to spread the computational load to the cloud and edge devices for cost-effectiveness and good ROI. Interoperability, these edge devices can communicate between legacy and modern systems to capture the benefits from both the new and legacy systems.

Cloud IoT core in the context of google cloud consists of subsystems like protocol bridge and device manager. The data is transmitted to the cloud IoT core using TLS and protocol bridge using secure MQTT port and HTTP/S port.

2.2.12 AMAZON WEB SERVICES (AWS): [22]

The IoT reference architecture presented by the amazon web services is shown in the figure below.
This architecture provides secure bidirectional communication between the internet and the devices like sensors, actuators, microcontrollers, and appliances.

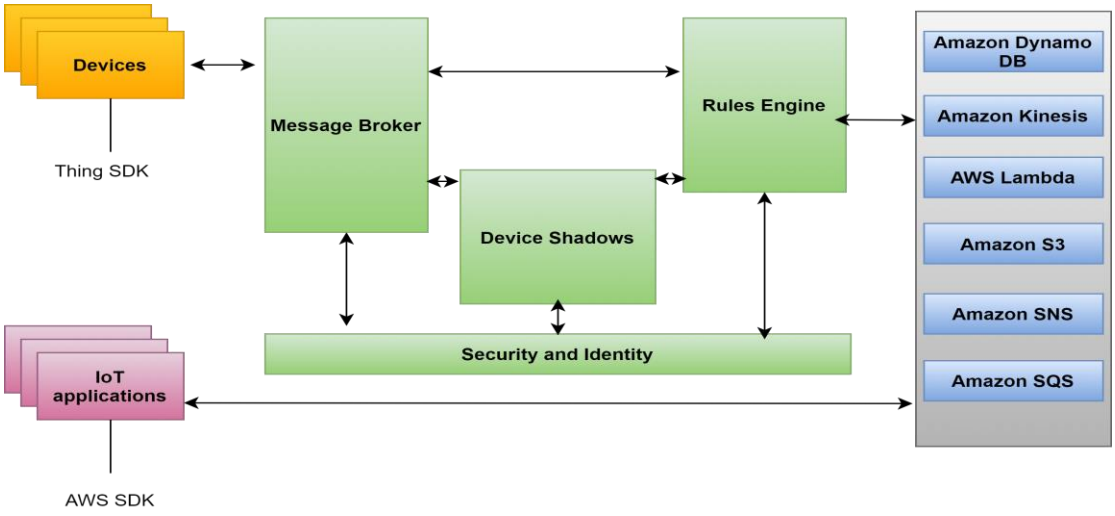

Figure 14 AWS IoT Reference Architecture

2.2.12.1 *Device Gateway:* This layer helps the devices to securely and efficiently communicate with AWS IoT.

2.2.12.2 *Message Broker:* The communication between the devices and AWS IoT is usually done by a message broker. The data is distributed to the devices and core AWS services through the message broker.

2.2.12.3 *Device Shadow:* The purpose of this layer is to maintain the state of the device. Online or not the applications should communicate with the devices. The data is maintained for the connected applications when offline and synchronizes back to its state when online to the device shadow service.

2.2.12.4 *Rules Engine:* For storage and processing the data is connected from message broker to AWS services through the rules engine. The expressions defined in the rules engine can be used to update, insert or query a Dynamo DB table.

2.2.12.5 *Security and Identity:* The communication is secured by X.509 certificates for authentication. The credentials should be secured. Both message brokers and rules engines use the AWS security and identity layer to send the data securely to the devices and AWS services.

The core IoT rules engine can connect to these AWS services.

2.2.12.6 *Amazon Dynamo DB:* This is a scalable and NoSQL database service that gives us fast and predictable database performance.

2.2.12.7 *Amazon Kinesis:* It collects, process, and analyzes the streaming data to get to know the new information. This layer uses the audio, video, and application logs for machine learning, data analytics, and applications.

2.2.12.8 *AWS Lambda:* This helps us to execute the code without managing the servers. The mobile application and web can be used to directly execute the code from AWS IoT data automatically.

2.2.12.9 *Amazon Simple Storage Service:* In Amazon S3 the data can be stored and retrieved at anytime from anywhere through the web. This data can also be sent for storage purposes.

2.2.12.10 *Amazon Simple Notification Service:* Amazon SNS is a web service that enables applications, users, and devices to send and receive information from the cloud.

2.2.12.11 *Amazon Simple Queue Service:* This is a message queuing service used to decouple and scale the services, distributed systems, and applications.

Table 1. Summary of Literature Review

No	Author	Year	Title	Source	Findings
1.	Fatma Alshohoumi.	2019	Systematic Review of Existing IoT Architectures Security and Privacy Issues and Concerns.	International Journal of Advanced Computer Science and Applications	This SLR gives the evolution of the IoT reference architectures about the architecture stack, challenges, and covered issues relevant to security and privacy concerns. The findings highlight that the initial IoT architectures do not convey the comprehensive meaning of the IoT. It highlights that none of the evolved architectures elaborate the privacy concerns in detail in the architecture.
2.	Atefeh Torkaman	2016	Analyzing IoT Reference Architecture Models.	International Journal of Computer Science and Software Engineering	This research work only analyzes four reference architecture models. The findings suggest that the requirements of the projects need to be improved in terms of reliability, management of big data, and security.
3.	Ismini Psychoula	2019	A Privacy-Aware Architecture for IoT Enabled Systems.	IEEE Smart World, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People & Smart City Innovation.	This research work proposes the privacy-aware architecture that mitigates the privacy risks associated with the IoT. This work preserves the privacy of the users by controlling data anonymization and authentication. It highlights that very little amount of work is done on how to protect the data of the sensors after its transmission.
4.	Kate Grant	2014	Study Report on IoT Reference Architectures/Frameworks.	ISO/IEC	The standardization bodies of ISO/IEC and SWG5 presented the report on the IoT reference architectures and frameworks. A layered reference architecture was proposed by the Korean Study Group.

5.	Shanzhi Chen	2014	A Vision of IoT Applications, Challenges, and Opportunities with China Perspective.	IEEE Internet of Things Journal	This research work is a step towards the standardization of the IoT reference architecture in China. It highlights the policies, research, and development plans. It also proposes a reference model for IoT.
6.	Paola Pierleoni	2020	Amazon, Google, and Microsoft Solutions for IoT Architectures and a Performance Comparison.	IEEE Access	This research work does a detailed performance comparison between the reference architectures of IoT integrated with cloud computing. It integrates the virtually unlimited capacity of cloud computing to the resource-constrained IoT devices in terms of storage, resources, and processing power. This work does not declare a winner amongst the three but helps developers to come up with a useful tool.
7.	Anna Gerber	2020	Simplify the development of your IoT solutions with IoT architectures.	IBM Developer Accessed online Mar 2021	This article highlights the security by design and provides a layered architecture that ensures that IoT architecture remains scalable, flexible, and robust.
8.	Alessandro Basi	2013	Enabling Things to Talk, Designing IoT Solutions with IoT Architectural Reference Model	Springer-Verlag	This work proposes the construction of a concrete Architectural Reference Model by the IoT-A project team. This project is incorporated by the EU seventh framework program FP7.
9.	Chao Li	2019	Privacy in Internet of Things from Principles to Technologies.	IEEE Internet of Things Journal	This research review the state of art principles of privacy laws and privacy-enhancing technologies (PETs) in the IoT. It follows the general data protection regulation GDPR.

2.3. SUMMARY:

This chapter gives a detailed description of the background study and literature review about IoT reference architectures. It describes the evolution of reference architectures from the start and the work done on them. The improvements in the architecture in terms of stack covered issues and challenges. Its findings include major research gap areas like the work on the sensor data after its transmission and lack of integration of privacy in the architecture stack of IoT. It highlights the IoT standardization bodies. The literature review includes a detailed description of reference architectures of the IoT. In the next chapter, we will discuss the requirements and quantifiable metrics extracted through the standards to analyze the reference architectures.

CHAPTER 3

3. IDENTIFIED REQUIREMENTS, METRICS, AND ANALYSIS OF IOT REFERENCE ARCHITECTURE

3.1. REQUIREMENTS FOR IOT REFERENCE ARCHITECTURES

3.1.1. Device:

The hardware or equipment connected to the things includes in the device requirements. This is a functional requirement and the proposed reference architecture must support these requirements related to things in the IoT.

3.1.1.1 *Device Connectivity:* There should be reliable connectivity between things and the identifier of the internet of things.

3.1.1.2 *Device Control and Configuration:* The devices should be able to remotely access, monitor, and control for the management of the devices. They must be easy to install and configure, supporting plug and play capability for ease of use. This should also be for the semantic configurations for the integration of the things with each other.

3.1.1.3 *Device Monitoring:* The devices should be monitored through automatic notification of things and changes in them.

3.1.1.4 *Device Mobility:* Mobility is required so that the connected devices get support in connectivity in IoT.

3.1.1.5 *Device Integrity:* There should be the integrity of the devices for the availability of the devices.

3.1.2. Security:

The IoT reference architecture should possess the requirements related to security this includes the functional requirements that should capture, store, transfer, aggregate, and process data. The security feature can be further elaborated like communication security that can prevent unauthorized access to the data. The data should be secured in its transmission and receiving. Data should also be secured while storing and processing. The authentication should be ensured between the user and the device of IoT in conformance to the security policies. Security audits must also be incorporated to ensure the transparency that proper laws and regulations relevant to security policies are followed. Encryption algorithms and techniques can further improve the security of the IoT reference architecture.

3.1.3. Modularity:

This is when a distinct unit combines with other components. To form systems the modularity helps the components to combine in different organizations. The design of the components has the flexibility by focusing on only interfaces and not on the internal working of the components.

3.1.4. Identification:

To trace and identify the entities of the IoT system unique identification plays an important role. Entities include the software components, sensors, actuators, and network components. To communicate and monitor specific entities the unique identification is very important. Identification schemes can be incorporated to meet the requirements.

3.1.5. Network Connectivity:

The IoT systems communicate through network links. The medium of this connectivity is either wired or wireless. IoT devices that route and terminate the communication are known as nodes. Different network topologies are followed for networking. This network structure can be static or dynamic.

3.1.5.1. *Communication Control:* Control in network communication is required to minimize communication errors. To provide in-time message handling and delivery time-critical communication is also required.

3.1.5.2. *Intelligent Communication:* The intelligent communication requirements include autonomic networking content-aware and location-based communication.

3.1.5.3. *Heterogeneous Communication:* Support for heterogeneous network communication is required. This communication can take place in the technologies like controller area network (CAN) bus, ZigBee, Bluetooth, Wi-Fi, etc.

3.1.6. Risk Management:

The IoT reference architecture should be adaptable in every condition. The environment should not have an impact on its performance.

3.1.7. Awareness:

The Architecture must possess the requirements relevant to time, location, content, and context.

3.1.7.1: *Time Awareness:* Just like the domain server's time synchronization is important to its clients to perform correctly. The IoT environment consists of different components that also need time awareness and synchronization of components with each other to perform and communicate correctly.

3.1.7.2: *Location Awareness:* Location parameters are required in many IoT systems to perform correctly such systems need awareness of the location. The accuracy of the location parameters depends on the application which is going to utilize them.

3.1.7.3: *Context Awareness:* It is the property of the IoT device which enables the service to monitor the operating environment. It also measures the order of the events that are occurring in the physical world.

3.1.7.4: *Content Awareness:* This property is basically of being aware of the data in the IoT component. This type of awareness helps the devices and services to adapt interfaces, application data and improve the precision of the information.

3.1.8. Support for Legacy Components:

The components that get old or outdated should get incorporated when needed by the IoT system. The reference architecture should provide relevant support to such components. This support for the legacy components should not hinder the reliability and performance of the system.

3.1.9. Confidentiality:

Confidentiality is one of the key aspects of security which means not to disclose information without authorization. It prohibits the users to read data and control the information they are not authorized for. To protect the private information of the individuals like personal information regarding financial records or medical history confidentiality is used. For the validation of the data, integrity is checked and life cycle management should be supported so that the IoT system should become more reliable. There will be a need to manage the large volumes of the data as well when big data is incorporated due to a large number of connected devices receiving and transmitting data in the IoT system.

3.1.10. Heterogeneity:

The IoT system is heterogeneous it consists of diverse components that communicate in different ways. Due to this the IoT reference architecture and system should support the heterogeneity. The components of devices should work together in diverse environments.

3.1.11. Programmable Interface:

The open-access of the application can only be provided by the interfaces through standard programs. API's also includes in this that is used in the interactions between different software's. These programmable interfaces should support interoperability and collaboration.

3.1.12. Promptitude:

The reference architecture proposed must support the time constraints. Time should be the priority during providing the services to the users.

3.1.13. Virtual Storage and Processing:

To store and process large volumes of the data which includes big data, virtual storage and processing are required.

3.1.14. Compliance:

The IoT reference architecture should comply with the regional, organizational, and regulations of the standards.

3.1.15. Service Tracking:

The requirements that are related to the services which include mobility, autonomic, management, and discovery should be incorporated in the reference architecture. These should be relevant to service providers. Mobility services are required so that they can support mobility. It should also support user and device mobility. Autonomic services are required to enable the automatic capture, communication, and processing of the data. The discovery is also required so that the users can discover the IoT service providers which leads to the requirement of the service management as well.

3.1.16. Autonomous:

Working of things that include devices, sensors, actuators operate in diverse environments which include the risk of being getting faulty. So they should be autonomous to be able to self-rectify, troubleshoot the issues and configure by themselves. They should heal by themselves. This requirement, the capability should be integrated into the architecture.

3.1.17. Power and Energy:

There are resource constraints of the power and energy in the IoT components as they are low-powered devices. They have constrained resources so the reference architecture should provide a provision of energy and power, data processing, and storage for such devices. They should be able to harvest energy efficiency and utilize it for their benefit.

3.1.18. Privacy:

Privacy of the users is one of the important aspects according to the nature of the IoT. The protection of the privacy of the users should be guaranteed. The private information should be hidden of things. The identity of the users should not be located or traced back to them. Privacy is a basic right of the individual that the information which is related to them is stored and processed under what defined purpose. These privacy principles should be applied in data collection, storage, processing. Data anonymization, minimization techniques should be incorporated in the privacy federated reference architecture. Authentication, encryption, access control, and authorization also include in the privacy protection of the users. Privacy can be achieved through confidentiality. To prevent the leakage of the data privacy requirements should be applied for data removal, requisition, and encryption.

3.2. QUANTIFIABLE METRICS FOR REQUIREMENTS

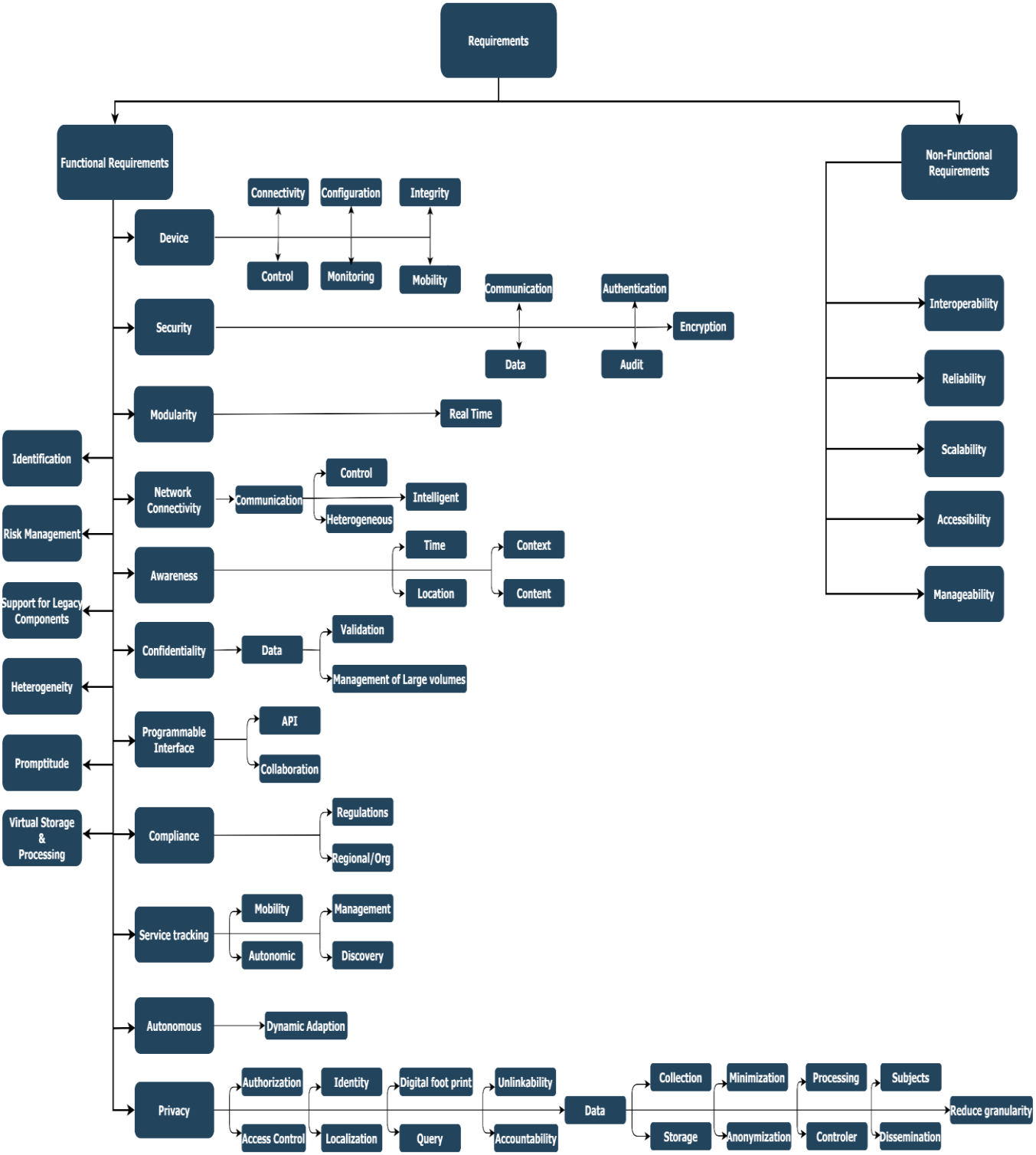

Figure 15 Quantifiable Metrics

3.3. ANALYSIS OF THE EXISTING IOT REFERENCE ARCHITECTURES

Table 2 Analysis of Existing IoT Reference Architectures

Requirements			Intel	Microsoft Azure	Mongo DB	IBM	Ser IoT	Cisco	IoT Arm	KSG	CCSA	WSO2	Google	Amazon
	Device	Connectivity	✓	✓	✓	✓	✓	✓	×	✓	✓	✓	✓	✓
		Control	✓	✓	✓	✓	✓	✓	×	✓	✓	✓	✓	✓
		Configuration	✓	✓	✓	×	✓	✓	✓	✓	×	✓	✓	✓
		Monitoring	✓	✓	✓	✓	✓	✓	×	✓	✓	✓	✓	✓
		Integrity	✓	✓	✓	✓	✓	✓	✓	✓	×	×	×	×
		Mobility	✓	✓	✓	✓	×	✓	×	✓	✓	✓	✓	✓
	Security	Communication	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
		Data	✓	✓	✓	✓	✓	✓	×	✓	×	×	✓	✓
		Authentication	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
		Audit	×	✓	✓	✓	×	×	×	×	×	×	✓	✓
		Encryption	✓	✓	✓	✓	✓	✓	✓	✓	×	✓	✓	✓
	Modularity	Real Time	✓	✓	✓	✓	✓	✓	×	✓	✓	✓	✓	✓
	Identification		×	✓	✓	✓	✓	✓	✓	✓	×	✓	✓	✓
	Network Connectivity-Communication	Control	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
		Heterogeneous	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
		Intelligent	✓	✓	✓	✓	×	✓	×	×	×	×	×	×
	Risk Management		×	✓	×	✓	×	✓	✓	✓	×	×	×	×
	Awareness	Time	✓	✓	×	×	×	×	×	×	×	×	×	×
		Location	✓	✓	✓	✓	✓	×	×	×	×	×	×	×
		Context	✓	✓	✓	✓	✓	×	×	✓	×	×	✓	×
		Content	✓	✓	×	✓	✓	✓	×	✓	×	×	×	×
Functional	Support for Legacy components		✓	✓	×	✓	×	✓	×	×	×	✓	✓	✓

	Confidentiality -Data	Validation	×	✓	✓	×	✓	✓	×	✓	×	×	×	✓
		Management of large volumes	✓	✓	✓	✓	×	✓	×	✓	×	✓	×	×
	Heterogeneity		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Programmable Interface	API	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
		Collaboration	✓	✓	×	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Promptitude		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Compliance	Regulations	✓	✓	✓	✓	×	✓	✓	✓	✓	✓	✓	✓
		Regional/org	×	✓	✓	✓	×	✓	✓	✓	✓	✓	×	✓
	Virtual Storage & Processing		✓	✓	×	×	✓	✓	×	✓	×	✓	✓	✓
	Service tracking	Mobility	✓	✓	✓	✓	×	×	✓	✓	✓	×	✓	✓
		Autonomic	✓	✓	✓	✓	✓	×	✓	✓	✓	×	✓	✓
		Management	✓	✓	✓	✓	×	×	✓	✓	✓	×	✓	✓
		Discovery	✓	✓	✓	✓	×	×	✓	✓	✓	✓	×	✓
	Autonomous	Dynamic adaption	✓	✓	✓	✓	✓	×	×	✓	✓	×	✓	✓
	Privacy	Authorization	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
		Access Control	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
		Identity	✓	✓	×	✓	✓	✓	✓	✓	×	✓	✓	✓
		Geo-Location	×	✓	×	✓	×	✓	✓	×	×	×	✓	×
		Digital footprint	×	✓	×	×	✓	×	×	×	×	×	×	×
		Query	×	×	×	×	×	×	×	×	×	×	✓	×
		Unlink ability	×	✓	×	×	×	×	✓	×	×	×	×	×
		Accountability	×	×	×	×	×	×	×	×	×	×	×	×
	Privacy-Data	Collection	×	✓	✓	✓	✓	×	✓	×	×	×	✓	×
		Storage	✓	✓	✓	✓	✓	✓	×	×	×	×	✓	✓
		Minimization	×	×	×	×	×	✓	✓	×	×	×	×	×
		Anonymization	×	✓	×	×	✓	×	✓	✓	×	×	✓	×
		Processing	×	×	×	×	✓	✓	✓	✓	×	×	×	×
		Controller	✓	✓	✓	×	✓	✓	×	✓	×	×	✓	✓

		Subjects	×	✓	×	×	×	×	✓	×	×	×	×	×
		Dissemination	×	×	×	×	×	×	✓	×	×	×	×	×
		Reduce Granularity	✓	✓	×	×	×	✓	✓	×	×	×	×	×
Non-Functional	Interoperability		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Reliability		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Scalability		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Accessibility		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Manageability		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

The tick check indicates the presence of a particular metric in the IoT reference architecture and the cross indicates the absence. The analysis gives us a detailed overview of which metric is lagging and is not included in most of the architectures.

3.3.1. Device Integrity:

The reference architecture of the IoT proposed by different organizations, research projects, and vendors should be analyzed on the integrity of the devices. It is a property through which the data cannot be altered or destroyed by unauthorized users. The integrity of the data is very important also for the reliability of the IoT systems. The software IoT applications data should not be altered through any sort of malicious activity and the reference architectures should support such requirement in the IoT system. This ensures the security of the system. IoT systems consisting of wireless sensor networks have intermediate nodes that can alter the data which can lead to the error in the functionality. The air conditioning system will not increase the cooling of the room due to the increase in the value of room temperature through the intermediate node.

3.3.2. Encryption:

To improve the security of the IoT systems the encryption algorithms and techniques could be applied and analyzed whether the mechanism is present in the architecture or not. The storage and communication of the data should be encrypted. There should be private data communication in the form of hidden data routing. It is a process of encoding a message from the sender to the intended recipient. No other user should be able to read it other than the intended user. It consists of a secret key or password that allows the user to decrypt the message.

3.3.3. Security Audit:

The analysis indicates the lack of auditing mechanisms adopted by IoT reference architectures. Most of the architectures are not auditing the security mechanisms implemented and conforms to the information security standards. The data access, processing, and storage should have a proper purpose defined under the laws and regulations. The Vulnerabilities got exposed in the form of cyber-attacks like DDOS and eavesdropping in the IoT systems. These things devices when hacked or exposed were used as a helping hand to disrupt the services of the server. The IoT architecture should incorporate the security audit.

3.3.4. Intelligent Network Communication:

The IoT devices should be intelligent in communication and architecture should inhibit this particular metric but analysis indicates that most of the reference architecture lacks this particular metric. Intelligent, autonomic, redundant networking is required to possess the capabilities of self-healing, self-rectification, self-path selection, or direction. Path selection redundancy and routing content-aware communication are required. The network flow analytics helps us to come up with better efficiency and results without any delay in communication. The congestion in the network traffic can be avoided through this intelligent network communication

3.3.5. Risk Management:

These IoT devices have vulnerabilities and can be exposed due to cyber-attacks for example a car that consists of sensors these days can be compromised and can lead to a lethal accident. The Risks can be calculated and avoided through risk management. This management can minimize the risk and countermeasure the vulnerabilities in the IoT system.

3.3.6. Support for Legacy Components:

The outdated components need support in the IoT systems along with the updated technologies. The analysis shows a deficiency in the support of such components in the IoT reference architecture. The integration of updated and legacy components is beneficial for the systems. It is good to come up with new components but should not limit the evolution from the start such as legacy systems. An example could be the transition of IPV4 compliance to IPV6 it’s a slow transition from legacy to future new technology. But the IPV4 is still not discontinued. The standards and applications still use IPV4. There is no clear answer to when to move to new technology leaving the legacy technology.

3.3.7. Data Validation:

Integrity is one of the major concerns in terms of security. Tampering the data should be avoided as it affects the reliability and functionality of the system. Validating the data is very important and should be incorporated in the IoT building block. Without validation, the corrupted or tampered data can be incorporated which can affect the efficiency of the system.

3.3.8. Virtual Storage and Processing:

Things that include devices, sensors, and actuators are in large number in the IoT systems which collect and process a large amount of data. The IoT devices are resource-constrained due to low power and processing capability. To overcome such constraints and the integration of big data we need cloud computing support in the form of virtual storage and processing. This can cover the deficiencies and constraints of the IoT system. Big data analytics should be incorporated in the reference architectures of IoT systems.

3.3.9. Service Tracking:

Services like mobility, autonomic, management, and discovery are important to be incorporated in the IoT systems. The static services can be easy as compared to mobile. The awareness of time, context, content, and location is essential for mobile services. The services should start automatically on the expiry of one and also warn the user before its expiry. The services should start without human intervention it should not be necessary to start service through human command and control.

3.3.10 Geo-Location Privacy:

Privacy of the users is very important and not incorporated in the IoT building blocks yet. The geolocation of the user can be identified through the identities of the user where the particular user device is at the moment. This device can trace an individual. Such type of information can be used for illegal purposes. Such types of data should be concealed. The analysis highlights that most of the reference architectures are not giving importance to the privacy of users in terms of their geolocation. The data should not be profiled based on geo-location.

3.3.11. Digital Footprint Privacy:

Digital footprint privacy addresses to use privacy settings and private data communication. IoT devices are connected to the internet all the time. Such a scenario can lead to vulnerabilities as the devices are continuously exposed to cyber-attacks. Data can be traced through the devices. The devices should be secured through effective security lightweight protocols to prevent the gathering of digital footprints of the devices and their owners. This should embed the checking of linking accounts and private data communication which includes encrypted data communication and hidden data routing.

3.3.12. Query Privacy:

The search query can reveal the identity of the person tracking the IP address of the user. In search query, it is suggested to answer high-level data instead of raw data. Giving raw data can lead to privacy violations of the users due to its secondary usage. This can overcome through open PDS/Safe Answers. This gives a high-level answer to the queries instead of raw data protecting the privacy of the end-users. Disseminating the information while answering the queries should be high-level answers instead of giving the raw data. Raw data can lead to privacy violations. Similarly, through analysis of the queries, we can block the repeated queries from the users which can lead to malicious activity that disclose the data of the users.

3.3.13. Privacy Accountability:

The data controller is responsible for the accountability of privacy in the IoT system. The data controller can control this through data collection defining the purpose of the collection of the data, limiting the required data, data dissemination, and only needed data should be collected. There should be no trade-off compromising the privacy of the users. Access controls should be defined in the form of ACLs and a digital certificate. Privacy impact assessment can be done through privacy SDS and Privacy Control Record (PCR). The analysis elaborates that none of the reference architectures of the IoT incorporating privacy accountability.

3.3.14. Data Privacy:

Data privacy should be embedded in the form of data subjects, collection, storage, minimization, anonymization, and processing. It should be incorporated in the design known as privacy by design. The data provider is the manager of privacy in the design of privacy.

3.4. SUMMARY:

This chapter consists of the requirements and metrics quantified through the standards that are essential for the development of the reference architecture of the IoT. We discuss in detail about these requirements and metrics divided into two categories functional and non-functional requirements. Based on these requirements, we analyze twelve reference architectures of the IoT and check through detail literature study whether a reference architecture from a particular vendor, organization address that particular metric. The privacy and security metrics are studied in detail and checked which metrics can be incorporated in the reference architecture of the IoT. The tick check indicates the presence of a particular metric and the cross indicates its absence. Through this analysis, we identify the shortcomings of the reference architectures and address those shortcomings in detail. In the next chapter, we will propose a reference architecture of the IoT addressing the shortcomings identified through a detailed analysis.

CHAPTER 4

4. PROPOSED PRIVACY FEDERATED IOT SECURITY REFERENCE ARCHITECTURE

4.1. PROPOSED REFERENCE ARCHITECTURE

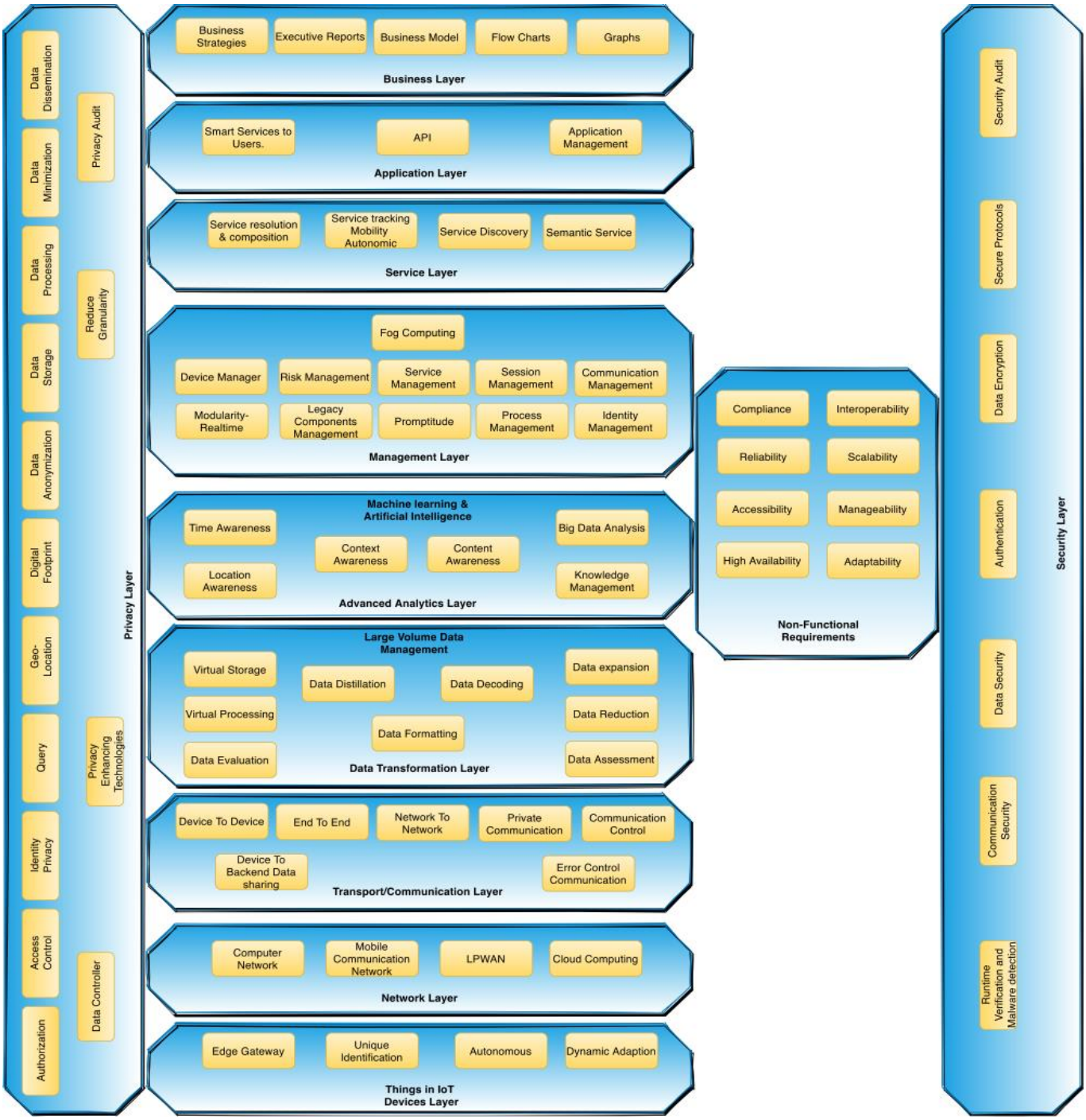

Figure 16 Proposed Reference Architecture

4.1.1. Things in IoT Devices Layer:

This layer consists of sensors, actuators and devices called things. The devices should have their identification and should be autonomous. They should be able to configure and rectify themselves without human intervention in case of any malfunctioning and errors. And this rectification should be done in the real-time environment supporting the dynamic adaption. The protocols used for unique identification are Electronic Product Code (EPC), Ubiquitous Code (uCode), Ipv6, Uniform Resource Identifier (URI). This layer should support the edge gateway it is present on the same layer where sensors and actuators are. These are high-powered devices that are capable of initial data collection, filtering, local aggregation, analysis, and offline data storage.

4.1.2. Network/Connectivity Layer:

This layer contains the computer network, mobile communication network, low power wide area network, and cloud computing. This layer helps the connectivity of the devices things with the network using connectivity protocols. Each metric has its own for example the computer network uses a wired or wireless medium for the connectivity which contains the following protocols

- WiMAX
- CAN (Controller Area Network) Bus
- Wi-Fi
- ZigBee
- ANT
- EnOcean
- Eddy stone
- NFC (Near Field Communication)
- Bluetooth
- Digi Mesh
- ISA 100.11a
- IEEE 802.15.4
- Wireless Hart.

The mobile communication network that uses mobile devices contains the following protocols.

- GPRS
- 2G
- 3G
- 4G
- 5G

The low power wide area network contains the following protocols

- Weightless
- Lora WAN
- LTE Machine Type Communication (LTE-MTC)
- Narrow Band IoT (NB-IoT)
- Random Phase Multiple Access (RPMA)
- Extended Coverage-GSM-IoT (EC-GSM-IoT)

This layer also incorporates cloud computing in the reference architecture. It is a server-less platform and can support many connected devices without needing any server sizing, provisioning, tuning, reconfiguration, or burdening any other IT tasks. The dispersed data of many IoT devices can be converted to the cloud IoT platform.

4.1.3. Transport/Communication Layer:

This layer incorporates backend data sharing from the devices. Private communication and error control communication. Automatic communication modes are required between users and the devices. The error control is important to handle the interference with the devices regarding the communication and to minimize the errors. Private communication helps to prevent cyber-attacks during the communication, bits of advice to incorporate security protocols so that the communication should be encrypted and cannot be breached. The protocols used for the communication are following.

- Ipv6
- Time Synchronized Mesh Protocol (TSMP)
- User Datagram Protocol (UDP)
- Content-Centric Networking (CCN)
- Ipv6 over Low Power Wireless Personal Area Network (6LoWPAN)
- Nano IP
- Aeron
- RPL/Roll
- Datagram Transport Layer (DTLS)
- uIP
- Quick UDP Internet Connection (QUIC)

4.1.4. Data Transformation Layer:

This layer transforms the data for the upper layers through data assessment, data reduction, data decoding, data formatting, data distillation, and data evaluation. The data is being evaluated and checked whether it is in a suitable format for the upper layer. It handles a large amount of data therefore virtual storage and processing are also embedded. Big data is managed through virtual storage and its processing.

4.1.5. Advanced Analytics Layer:

Machine learning and artificial intelligence algorithms are applied to the data collected from the below layers to get the best results from the upcoming data. This includes big data analysis, content awareness, context awareness, knowledge management, time awareness, and location awareness. Advanced analytics can be achieved through knowledge management which consists of gathering and intelligent learning. This incorporates deep business insights to predict before the failure of a component through analytics.

4.1.6. Management Layer:

The data after the advanced analytics layer then go to the management layer which manages data in terms of legacy component support, as to how to carry the old components and technologies with the new ones how to manage them both together. This layer provides management services which include risk management through asset categorization and risk value. Fog computing, Promptitude, Service management, Session management, Communication management, and Identity management.

4.1.7. Service Layer:

This layer is responsible for service resolution and composition, service tracking, mobility, autonomic, service discovery, and semantic service. Multicast domain name system (mDNS), universal plug and play (Upnp), physical web, and hyper cat are some protocols used in service discovery.

4.1.8. Application Layer:

This layer consists of the smart services in the form of applications for the users. API's are provided through this layer. The management of the applications is also done in this layer.

4.1.9. Business Layer:

This layer provides the business insights of the IoT system through graphs, flow charts, and executive reports for the top management. These reports play a vital role in the development. Strategies are also developed to capture the market. It carries the profit models for the system.

4.1.10. Privacy Layer:

The previous works focus on the issues like assisting the users with mobile application permissions, protecting the data regarding location, and privacy-aware video streaming. Our goal is to design an architecture that will allow the users to store manage the data according to the level of privacy they want for themselves trading the data for services rather than allowing the individuals to view, control, and disclose their data.
Privacy of the users can be ensured through the integration of privacy by design metrics to the reference architecture of the IoT. This includes the privacy validation chain that acts between the data owner, the data controller, and the data processor to define the purpose of the usage of the user's data. This acts between the data provider and the data controller who are the manager and accountable for the privacy protection respectively. Privacy-enhancing technologies (PETs) are used to enhance the privacy of the system. The main threats to privacy are identification, localization, and tracking. The privacy validation chain (PVC) answers the most important question that who is collecting the data under what defined purpose. The user authorization is required according to the predefined security policies to access the IoT followed by the access control list and digital certificates.

4.1.10.1. Data Anonymization:

This technique can prevent personally identifiable information before it is used by the IoT application. This leads to the data being anonymous. This reduces the risk of identification of personal information and privacy violations. This can include the secret key encryption mechanism and k-anonymity with a large value of k which exploits the quasi-identifier attributes to preserve the sensitive data. Strong identities with no unique identifiers in the database can lead to the prevention of the privacy of the users.

4.1.10.2. Data Storage:

The data storage should be minimized for example the raw data should be deleted after deriving the secondary contexts. Privacy can be enhanced by no long-term personal characteristics, distributed data storage, limiting the storage of the data, defining the legal needs to store the user's data, the purpose of storage, and encrypted data storage.

4.1.10.3. Data Processing:

The processing of the data should be distributed and encrypted so that the data may not get tampered by malicious attacks. Encryption is the encoding of data in which only authorized users can read the data. Those who are processing the data should not be always allowed to read the data as well.

4.1.10.4. Data Minimization:

Data minimization incorporates minimum knowledge discovery by discovering the data that is only needed to achieve the primary objectives by the IoT application and the rest of the detailed information should not be collected. Minimize the raw data intake. The raw data intake can lead to the secondary usage of the data which leads to privacy violation. Minimize data retention period, the retention of the data for a longer period of time should be avoided. A longer retention period can give more probability to do a malicious activity and breach the privacy of the user.

4.1.10.5. Reduced Data Granularity:

The IoT technologies should implement the lower level of granularity because if a higher level of granularity is implemented the fine-grained will be the data and information that will result in more privacy risk as compared to reduced data granularity.

4.1.10.6. Data Controller:

The data controller is accountable for the protection of privacy it includes privacy auditing through systematic checking of the logs and procedures. To control privacy through privacy SDS and PCR. The data subjects should be controlled through a mechanism.

4.1.11. Security Layer:

This layer consists of lightweight authentication mechanisms with communication and data security. Security audits should be done in the form of fairness, clearly informed, and transparent data access. The data access should be according to the rule of law and regulations. Runtime verification, malware detection, and data encryption are done in this layer. The following protocols can be used for lightweight data security protocols.

- ONS 2.0
- Reactive Streams
- Simple Sensor Interface (SSI)
- Message Queuing Telemetry Transport (MQTT)
- Constrained Application Protocol (CoAP)
- Simple Text Oriented Messaging Protocol (STOMP)
- Advanced Message Queuing Protocol (AMQP)
- Extensible Messaging and Presence Protocol (XMPP)
- Representational State Transfer (REST)
- Light Weight M2M (LWM2M)
- Light Weight Local Automation Protocol (LLAP)

- Data Distribution Service for Real-Time System (DDS)
- Java Message Service (JMS)
- Mihini/M3DA

4.1.12. Non-Functional Requirements:

This layer incorporated includes high availability, adaptability, accessibility, manageability, reliability, scalability, interoperability, and compliance as the NFR.

4.2. VALIDATION

The reference architecture proposed is privacy and security federated reference architecture. It consists of nine horizontal and two vertical layers with a layer describing nonfunctional requirements. It is a step towards standardization and building a concrete architecture on which the heterogeneous IoT system can rely. Each layer collaborates with the other through refined metrics. The design of the architecture is refined and optimized in terms of metrics and layers addressing the security and privacy concerns of the users in detail.

Validation is having a shred of documentary evidence demonstrating that the process or procedure carried out in testing and production conforms at all stages. It is the checking of validity or accuracy against the quality attributes. Including the acceptance and testing, it describes that the system designed or produced satisfies the user's needs or to check have we built the right system?

To validate the proposed privacy federated IoT security reference architecture we have followed the industry-recognized scenario-based approach. Researchers have termed this approach to be better in comparison to the questionnaire-driven approach and decision-based approach. We have adopted Architecture Tradeoff Analysis Method (ATAM) to validate the proposed architecture. This methodology provides us insight into how the quality goals interact with each other and how they can tradeoff with each other. ATAM is the leading methodology to evaluate and validate the architecture. This methodology consists of the following steps.

4.2.1. Present the ATAM.
4.2.2. Present the business drivers.
4.2.3. Present the architecture

The architecture has been proposed and presented in chapter 4.

4.2.4. Identify the architectural approaches
4.2.5. Generate a quality attribute utility tree.
4.2.6. Brainstorm and prioritize scenario.
4.2.7. Analyze architectural approaches.
4.2.8. Present the results.

4.2.1. Present the ATAM:

ATAM evaluation can identify and expose the risks that can inhibit the achievement of an organization's business goals. It is a scenario-based approach in which the proposed reference architecture is evaluated and validated through the quality attributes in brainstormed scenarios. This checks whether it meets the functional requirements addressed in the standards of NIST, ISO/IEC, and ITU-T regarding IoT. This results in the identification of the tradeoffs, sensitivity points, and risks associated with the architecture.

4.2.2. Business Drivers:

The business drivers are following for the Internet of things.

- Revenue and innovation, large investments on the internet of things.
- The low cost of sensors and the shift from traditional sensors to smart sensors have contributed to the business growth of the IoT.
- Better customer service, support, and improved customer experience.
- High mobile adaptation ratio.
- Product service improvement and innovation.
- Supply chain and logistics.
- New consumer demands.
- Diverse and expanded internet connectivity.
- Asset tracking, utilization, and inventory management.

The drive for the business also incorporates the problems relevant to implementation and security. To identify how to benefit from deploying IoT architecture to connected devices and services. To ensure privacy, data management, analytics, and rules automation. Lack of standardization is a major factor that can hinder business growth.

4.2.4. Identify Architectural Approaches:

Proposed privacy federated IoT security reference architecture is scalable, secure, and flexible architecture. It is a layered architecture that has no restrictions in terms of numbers and type of layers. It consists of nine horizontal and two vertical layers along with a layer with non-functional requirements. We have followed the ITU-T Y.2066 and ISO/IEC 30141 standards for the proposed IoT reference architecture. The other standards are ITU-T Y.2060 and NIST (Network of Things) but these both focus on the device and physical object communication. They do not completely address the end-to-end IoT systems reference architecture model. The ISO/IEC 30141 provides more elaboration on the system architecture of the IoT in terms of conceptual, system, domain, network, functional, and cross-sectional service view of the ecosystem. It is a modular and scalable architecture that provides an understanding of the key aspects of the architecture and how they will operate independently before embedding them into the IoT solution. Two vertical layers integrate the security and privacy concerns of the end-users. The metrics embedded are incorporated through the requirements defined in standards.

4.2.5. Quality Attribute Utility Tree:

The utility tree identifies the quality attributes to achieve the most important quality goals in architecture to validate the architecture based on the requirements. This follows the top-down approach. The quality factors which possess system utility are (Performance, Usability, Reliability, Install ability, Functionality, Security, Portability, and Privacy). In the next level, there are refinements of the quality attributes. Specified down to the scenarios which are also called the leaves of the trees. ASR’s provided by the business drivers for the quality attributes are mapped in the quality attribute tree.

Scenarios are generated through brainstorming considering the events in real-time scenarios. Day-to-day usage of the IoT applications and the utilization of the proposed reference architecture metrics, can generate the scenarios to validate. We have then validated through mapping whether the proposed architecture's significant metrics meet the defined quality attributes and their refinements or not. The figure below shows the utility tree along with the scenarios in accordance with the quality attributes.

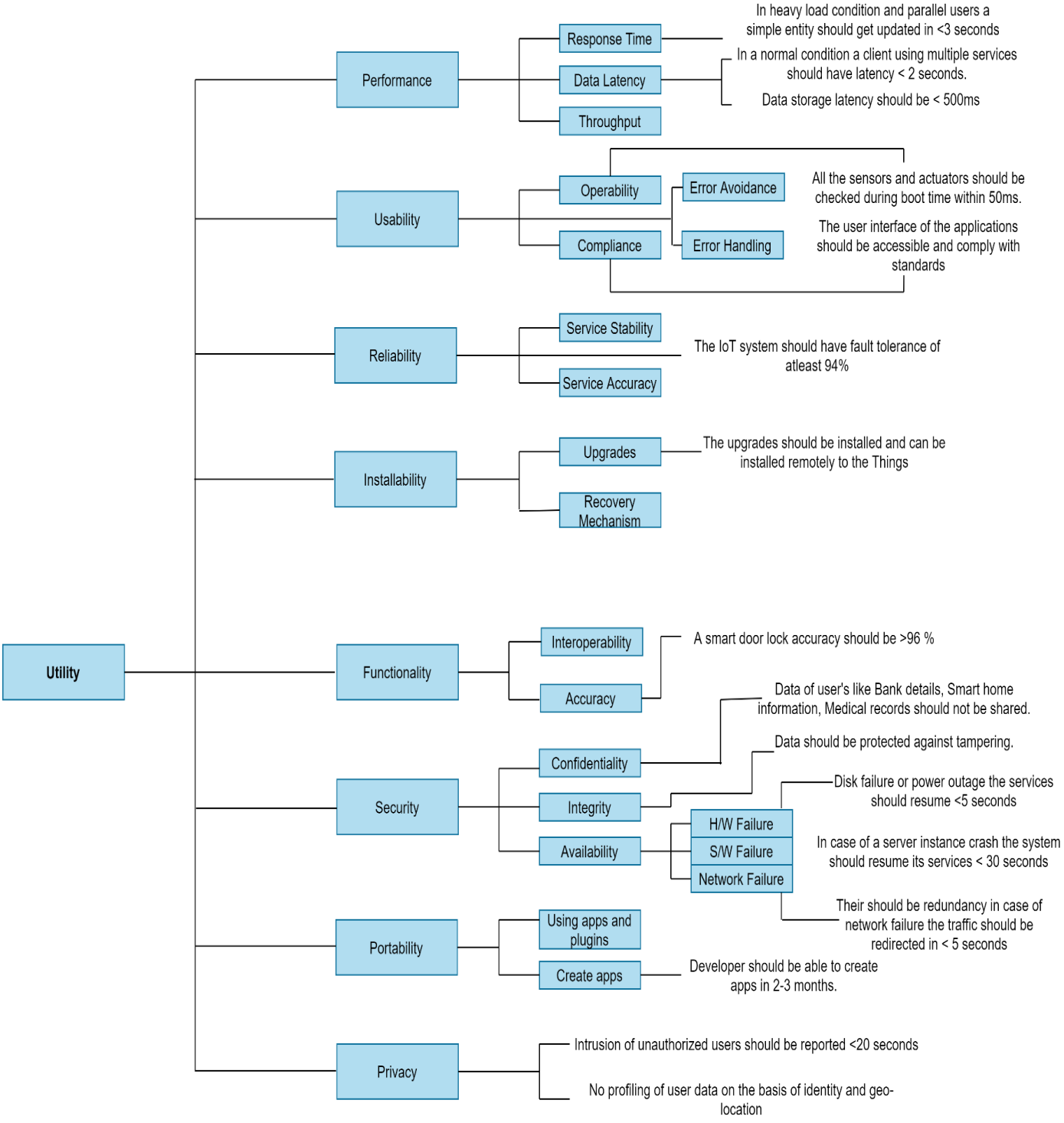

Figure 17 Quality Attribute Utility Tree

4.2.6. Brainstorm and Prioritize Scenarios:

Based on the scenarios generated in the utility tree a larger set of scenarios are elicited from the stakeholders. These will be prioritized by the stakeholders using the ASR's collected previously. These can be real-time scenarios. The characterization and prioritization of the quality attribute can be defined as the importance for the success of the systems and the difficulty to achieve it can be the architect's assessment. This will be prioritized as (High, Medium, and Low). The scenarios following their priorities are described in the table below.

Table 3 Brainstormed Scenarios following Priority.

Scenario Number	**Scenario text**	**Priority**
1	A smart home where all the appliances are connected to the internet. A user requests to unlock the door through a mobile application rather than just normal keys. (Functionality, A smart door lock accuracy should be > 96%).	(H, M)
2	A connected car self-driven can optimize its operation and maintenance driving on the road without a driver. (Reliability, The IoT system should have fault tolerance of 94%).	(H, M)
3	Industrial internet of things also known as industry 4.0 the revolution of industry. Production units highly rely on sensors, actuators, and controllers. Temperature, Voltage, frequency, Seismic sensors not giving correct readings to Programmable logic Controllers (PLCs). Giving false negatives. (Usability, All the sensors, and actuators should be checked during boot time within 50ms).	(M, M)
4	In Smart health care, Patients using a connected battery-powered pacemaker to control abnormal heart rhythms. (Security, Hardware disk failure or power outage the services should resume < 5 seconds)	(H, H)

5	In Smart retail large number of users requesting for transaction check out at a time using mobile POS. (Performance, In heavy load conditions and parallel users a simple entity, should get updated in < 3 seconds).	(M, L)
6	IoT medical devices collect healthcare data including blood pressure, sugar level, oxygen, and weight the data of users are stored online. (Privacy, No profiling of user data based on identity, geolocation)	(M, H)
7	The developer should be able to create new applications in the IoT ecosystem. (Portability, The developer should be able to create apps in 2-3 months).	(M, L)
8	The Patches should be installed on the software and Operating systems of Things. (Install ability, The upgrades should be remotely installed to the things).	(H, M)

4.2.7. Analyze Architectural Approaches:

In this step of architecture evaluation, the validation we will map the scenarios brainstormed and defined in the quality attribute tree to the proposed reference architecture of the IoT to see whether architecture has the response to the stimulus of the scenario. The source of the stimulus sends a stimulus which is any kind of condition that makes the system respond. End-user input is the source of the stimulus. The environment is the mode or state of the system while receiving a stimulus. It could be starting up the system, shutting down the system, recovering from failure, or normal operations. This will identify the risks, sensitivity points, and tradeoffs. The architecture decisions will be specified. Through mapping, we will check our proposed reference architecture that whether it has the response against a particular stimulus in the prioritized scenarios. It will also evaluate the architecture against the quality goals. The analysis is presented below in the form of tables by their scenario numbers.

The importance of this scenario for the success of the system is high and its difficulty to achieve is medium. To validate the architecture in terms of decision-taking capability in a particular scenario we will generate a stimulus in normal operating conditions to check the response and architecture decision. In this scenario, we will evaluate the functionality of our proposed architecture and the tradeoff that can be made.

Table 4 Scenario 1.

Scenario#1	**Scenario: A smart door lock accuracy should be > 96%.**
Attribute	Functionality
Environment	Normal Operations.
Stimulus	The mobile application fails to unlock the door using communication protocols Z-Wave, Wi-Fi, ZigBee.
Response	Will not affect overall system functionality and accuracy.
Architecture Decision	(Layer 3) Transport/ communication layer Error Control Communication.
Sensitivity	This layer should be able to control communication and errors from multiple IoT devices with the capability of intelligent networking.
Tradeoff	Performance, Reliability
Risk	The interoperability in the functionality could result in security vulnerabilities and the smart home could be compromised by unauthorized users.

The importance of this scenario for the success of the system is high and its difficulty to achieve is medium. The quality attribute that will be checked is reliability under normal operating conditions. We will generate a stimulus about the failure of the system to check what will be the decision of our proposed architecture to handle the failed state of the system.

Table 5 Scenario 2.

Scenario#2	**Scenario: The IoT system should have a fault tolerance of 94%).**
Attribute	Reliability
Environment	Normal Operations.
Stimulus	A self-driven car has failed to sense a hurdle on the road components failure.
Response	Without affecting the reliability of the self-driven car.
Architecture Decision	(Layer 5) Advanced Analytics layer. Machine learning and Artificial intelligence
Sensitivity	The IoT devices should be autonomous to detect any failure, change and adjust themselves according to the environment.
Tradeoff	No tradeoff.
Risk	If there is less fault tolerance percentage the system cannot be termed as reliable and can lead to a major hazard like an accident in this scenario.

The importance of this scenario for the success of the system is medium and its difficulty to achieve is medium. A scenario is generated to check the usability attribute while starting up the system. We will check the architecture decision in a particular stimulus i.e. failure of the system.

Table 6 Scenario 3.

Scenario#3	**Scenario: All the sensors and actuators should be checked during boot time within 50ms**
Attribute	Usability
Environment	Starting up the system.
Stimulus	Failure of boot-time check of sensors and actuators within 50ms
Response	Will not affect overall system operations.
Architecture Decision	(Layer 1) Devices layer. Things in IoT.
Sensitivity	The devices like sensors, actuators, wearables should be able to check, protect and configure themselves within specified boot time.
Tradeoff	Portability, Reliability, Functionality
Risk	Could result in false negatives, can halt the production units resulting in financial loss.

The importance of this scenario for the success of the system is high and its difficulty to achieve is high. The system is recovering from failure and the quality attribute that will be addressed are security and availability. The scenario will address whether the system has the response to the state of the system and which particular metric will address the response to the stimulus.

Table 7 Scenario 4

Scenario#4	**Scenario: Hardware disk failure or power outage the services should resume < 5 seconds**
Attribute	Security, Availability.
Environment	Recovering from Failure.
Stimulus	The hardware or battery of the pacemaker fails during operation.
Response	The recovery mechanism supported will not affect the security and availability of the system.
Architecture Decision	(Layer 6) Management layer Risk Management.
Sensitivity	There should be no common mode of failure to ensure different hardware and operating system
Tradeoff	Install ability, Reliability
Risk	This could result in fatal hazards, Management layer might be helpful in risk minimization might not address the hardware redundancy, What type of redundancy analytical or functional?

The importance of this scenario for the success of the system is medium and its difficulty to achieve is low. The quality attribute that will be checked in this scenario will be the performance of the system in extreme working conditions. The risk attached to this scenario is it can affect the goodwill of the consumers and halt the sales.

Table 8 Scenario 5

Scenario#5	**Scenario: In heavy load conditions and parallel users a simple entity should get updated in < 3 seconds.**
Attribute	Performance
Environment	Extreme operations.
Stimulus	Due to the increased number of processing at a time the POS system gets hanged.
Response	Heavy load and parallel processing will not affect the response time of the smart retail system.
Architecture Decision	(Layer 5) Advanced Analytics layer, (Layer 6) Management layer Big Data Analysis, Process Management
Sensitivity	Virtual storage and processing using cloud computing should be secure and reliable.
Tradeoff	Portability, Reliability, Security, Privacy
Risk	Could damage the goodwill of the consumer experience and halt sales.

The importance of this scenario for the success of the system is medium and its difficulty to achieve is high. The quality attribute checked in this scenario is privacy under normal operating conditions. The architecture proposed should have mitigation to the risk associated with the specific scenario.

Table 9 Scenario 6

Scenario#6	**Scenario: No profiling of user data based on identity, geolocation.**
Attribute	Privacy
Environment	Normal operations.
Stimulus	Medical records of the patients get profiled based on unique identifiers.
Response	Will not disclose and profile the data based on identities in the database.
Architecture Decision	(Vertical layer 1) Privacy layer. Identity privacy, Geolocation privacy, Privacy Audit.
Sensitivity	Health care IoT devices should have a Privacy validation chain and a defined purpose of collection and profiling of data.
Tradeoff	Security, Reliability
Risk	Unauthorized data collection and profiling of healthcare records could lead to exposure.

The importance of this scenario for the success of the system is medium and its difficulty to achieve is low. This scenario will check the portability quality attribute from the utility tree. The response of the system will identify the architecture decision.

Table 9 Scenario 7

Scenario#7	**Scenario: The developer should be able to create apps in 2-3 months**
Attribute	Portability
Environment	Normal Operations.
Stimulus	A new application or version of an operating system fails to configure with things sensors, actuators, and devices.
Response	It does not affect the software's ability to get transferred from one hardware to another.
Architecture Decision	(Layer 4) Data transformation layer (Layer 8) Application layer Data assessment, Data expansion API
Sensitivity	The things in IoT should be open source and be able to create and modify applications in case of any incompatibility or failure.
Tradeoff	Functionality, Usability, Security
Risk	If not tested the new applications could lead to major security vulnerabilities which could result in exposure.

The importance of this scenario for the success of the system is high and its difficulty to achieve is medium. Install ability attribute is checked whether it is achieved or not and which layer or particular metric will have the response to the system. The tradeoffs that can be made are identified in the table below.

Table 10 Scenario 8

Scenario# 8	**Scenario: The upgrades should be remotely installed to the things.**
Attribute	Install ability
Environment	Normal operations.
Stimulus	Failure to connect a mobile device to the target controller to install updates.
Response	Will not affect the communication network of low power resource-constrained IoT devices.
Architecture Decision	(Layer 2) Network Layer Mobile Communication network, LPWAN
Sensitivity	Things should be autonomous to carry on and manage Legacy components will become difficult.
Tradeoff	Security, Usability.
Risk	This could result in bugs and viruses and loss of data while upgrading the things firmware or OS. Potential downtime while upgrading.

4.2.8. Present the Results:

The process of the architecture tradeoff analysis method gives us the tradeoffs, sensitivity points, and risks associated with the proposed internet of things reference architecture. It gives us a clear sight of how the reference architecture should perform under the brainstormed real-time scenarios. We generate the stimulus in brainstormed scenarios of the failures of the system, check and map it with our proposed architecture whether our architecture addresses that particular stimulus in the given environment, and what would be the response of the system. The tradeoffs give us insight into which quality attribute could be given up to gain the other. The achievement of the quality goals and attributes refines, evaluates, and validates the proposed reference architecture.

4.3. SUMMARY:

In this chapter, we propose and validate privacy federated IoT security reference architecture. It is a modular and interoperable architecture. It addresses the shortcomings of the reference architectures analyzed. It consists of nine horizontal and two vertical layers along with a layer of non-functional requirements. Each layer addresses specific metrics and functionality. This architecture is a step towards standardization as it follows three standards combined that is ISO, IEC, and NIST. Each layer is discussed in detail. To validate the proposed architecture, we have adopted the industry-recognized technique known as Architecture Tradeoff Analysis Method. It is a scenario-based approach that consists of detailed eight steps that address the real-time scenarios need of the architecture and its decisions based on those scenarios.

CHAPTER 5

5. CONCLUSION AND FUTURE WORK DIRECTIONS

Internet of things has transformed our planet into a smarter and intelligent planet through communications between objects and humans. It is finding its path in our daily lives in the form of its applications smart devices and technologies. There are different standardization bodies for the IoT and it lacks standard architecture. In this thesis, we have federated privacy and security to the reference architecture of the IoT as privacy has not been embedded thoroughly in the reference architecture of IoT identifying the core requirements through the standards. The metrics have been identified and analyzed which privacy metrics could be embedded in the architecture. Based on these requirements and metrics twelve architectures have been analyzed which includes the recently published reference architectures. These architectures have not yet been analyzed based on privacy metrics in detail. Shortcomings were identified and based on these shortcomings we have proposed a privacy federated IoT security reference architecture which will help towards the making of a concrete and standard architecture addressing all the concerns for the domain system and functional point of view. We have validated our proposed reference architecture through an industry-recognized scenario-based technique known as Architecture Tradeoff Analysis Method (ATAM) which will help the proposed reference architecture from a business perspective.

For future work, we recommend proposing privacy-enhancing technologies (PET's) to be embedded with the smart IoT devices incorporating all the metrics in the privacy layer. To come up with lightweight protocols considering the resource-constrained environment of the IoT. To optimize the IoT network in the federation to privacy and security. Complex encryption and authentication algorithms consisting of less latency and computing resources on tiny IoT devices could be a great breakthrough in the future. We recommend coming up with a lightweight trust management system to address the hardware insecurities of the IoT devices.

BIBLIOGRAPHY

[1] W. Zhou, Y. Jia, A. Peng, Y. Zhang, and P. Liu, "The Effect of IoT New Features on Security and Privacy: New Threats, Existing Solutions, and Challenges Yet to Be Solved," in IEEE Internet of Things Journal, vol. 6, no. 2, pp. 1606-1616, April 2019.

[2] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari and M. Ayyash, "Internet of Things: A Survey on Enabling Technologies, Protocols, and Applications," in IEEE Communications Surveys & Tutorials, vol. 17, no. 4, pp. 2347-2376, Fourth quarter 2015, DOI: 10.1109/COMST.2015.2444095.

[3] M. Frustaci, P. Pace, G. Aloi, and G. Fortino, "Evaluating Critical Security Issues of the IoT World: Present and Future Challenges," in IEEE Internet of Things Journal, vol. 5, no. 4, pp. 2483-2495, Aug. 2018, DOI: 10.1109/JIOT.2017.2767291.

[4] I. Psychoula, L. Chen, X. Yao, and H. Ning, "A Privacy-Aware Architecture for IoT Enabled Systems," 2019 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI), Leicester, United Kingdom, 2019, pp. 178-183

[5] P.P. Ray, "Internet of Things: A survey related to various recent Architectures and Platforms available," Computer and Information Sciences, vol. 30, no. 3, pp. 291-319, 2016

[6] S. A. Al-Qaseemi, H. A. Almulhim, M. F. Almulhim, and S. R. Chaudhry, "IoT architecture challenges and issues: Lack of standardization," 2016 Future Technologies Conference (FTC), San Francisco, CA, USA, 2016, pp. 731-738, DOI: 10.1109/FTC.2016.7821686.

[7] S. Kraijak and P. Tuwanut, "A survey on IoT architectures, protocols, applications, security, privacy, real-world implementation and future trends," 11th International Conference on Wireless Communications, Networking and Mobile Computing (WiCOM 2015), Shanghai, 2015, pp. 1-6, DOI: 10.1049/cp.2015.0714.

[8] Https://threatpost.com/researchers-allege-systemic-privacy-security-flaws-in-popular-iot-devices/141244/

[9] Fadele Ayotunde Alaba, Mazliza Othman, Ibrahim Abaker Targio Hashem, Faiz Alotaibi, Internet of Things security: A survey, Journal of Network and Computer Applications, Volume 88,2017, Pages 10-28, ISSN 1084-8045,https://doi.org/10.1016/j.jnca.2017.04.002.

[10] Fatma Alshohoumi, Mohammed Sarrab, Abdulla AlHamadani, and Dawood Al-Abri, "Systematic Review of Existing IoT Architectures Security and Privacy Issues and Concerns" International Journal of Advanced Computer Science and Applications (IJACSA), 10(7), 2019. http://dx.doi.org/10.14569/IJACSA.2019.0100733

[11] Torkaman, A., & Seyyedi, M.A. (2016). Analyzing IoT Reference Architecture Models.

[12] I. Psychoula, L. Chen, X. Yao, and H. Ning, "A Privacy-Aware Architecture for IoT Enabled Systems," 2019 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI), Leicester, UK, 2019, pp. 178-183, DOI: 10.1109/SmartWorld-UIC-ATC-SCALCOM-IOP-SCI.2019.00073.

[13] Kate Grant AHG, SWG5, JTC1, "Study Report on IoT Reference Architectures/Frameworks," ISO/IEC, 2014.

[14] S. Chen, H. Xu, D. Liu, B. Hu, and H. Wang, "A Vision of IoT: Applications, Challenges, and Opportunities With China Perspective," in IEEE Internet of Things Journal, vol. 1, no. 4, pp. 349-359, Aug. 2014, DOI: 10.1109/JIOT.2014.2337336

[15] P. Pierleoni, R. Concetti, A. Belli and L. Palma, "Amazon, Google and Microsoft Solutions for IoT: Architectures and a Performance Comparison," in IEEE Access, vol. 8, pp. 5455-5470, 2020, DOI: 10.1109/ACCESS.2019.2961511

[16] Anna Gerber, "Simplify the development of your IoT solutions with IoT architectures," IBM 7 August 2017. [Online]. Available: https://www.ibm.com/developerworks/library/iot-lp201-iot-architectures/index.html. [Accessed 22 March 2021].

[17] Alessandro Bassi, Rob van Kranenburg, Martin Bauer, Sebastian Lange, Martin Fiedler, Stefan Meissner, Thorsten Kramp, Enabling Things to Talk Designing IoT solutions with the, Berlin Heidelberg: Springer-Verlag, 2013.

[18] C. Li and B. Palanisamy, "Privacy in Internet of Things: From Principles to Technologies," in IEEE Internet of Things Journal, vol. 6, no. 1, pp. 488-505, Feb. 2019, DOI: 10.1109/JIOT.2018.2864168.

[19] Intel IoT Platform Architecture Specification White paper internet of things, Online Available http://d885pvmm0z6oe.cloudfront.net/hubs/intel_80616/assets/downloads/general/Architecture_Specification_Of_An_IOT_Platform.pdf

[20] Microsoft Azure IoT Reference Architecture V 2.1 26/09/2018 Online Available https://download.microsoft.com/download/A/4/D/A4DAD253-BC21-41D3-B9D9-87D2AE6F0719/Microsoft_Azure_IoT_Reference_Architecture.pdf.

[21] A Mongo DB White Paper IoT Reference Architecture, July 2019 Online Available https://www.mongodb.com/collateral/iot-reference-architecture

[22] Marcela G.dos Santos, Darine Ameyed, Fabio Petrillo, Fehmi Jaffar, Mohamed Cheriet, Internet of things architectures: A comparative study. Arxiv:2004.12936v1 [cs.SE] 27 Apr 2020.

[23] Reference Architecture for Secure and Safe Internet of Things by SerIoT Project, 14-Jan-2019. https://seriot-project.eu/2019/01/14/reference-architecture-for-secure-and-safe-internet-of-things-by-the-seriot-project

[24] Internet of things Reference Model Cisco, 2014, Available Online https://www.cisco.com/c/dam/global/en_ph/assets/ciscoconnect/pdf/bigdata/jim_green_cisco_connect.pdf

[25] Alessandro Bassi, Rob van Kranenburg, Martin Bauer, Sebastian Lange, Martin Fiedler, Stefan Meissner, Thorsten Kramp, Enabling Things to Talk Designing IoT solutions with the, Berlin Heidelberg: Springer-Verlag, 2013.

[26] Shanzhi Chen, Hui Xu, Dake Liu, Bo Hu, Hucheng Wang, “A Vision of IoT: Applications, Challenges, and Opportunities with China Perspective,” IEEE Internet of Things Journal, vol. 1, no. 4, 2014.

[27] Paul Fremantle, "A Reference Architecture for the Internet of Things," WSO2, 2015.