Chapter 16: Secure and Safe IIoT systems via Machine and Deep Learning approaches

Aris S. Lalos, Athanasios Kalogeras, Christos Koulamas, Christos Tselios, Christos Alexakos, Dimitrios Serpanos

Abstract This chapter reviews security and engineering system safety challenges for Internet of Things (IoT) applications in industrial environments. On the one hand security concerns arise from the expanding attack surface of long-running technical systems due to the increasing connectivity on all levels of the industrial automation pyramid. On the other hand, safety concerns magnify the consequences of traditional security attacks. Based on the thorough analysis of potential security and safety issues of IoT systems, the chapter surveys machine and deep learning methods (ML/DL) that can be applied to counter the security and safety threats that emerge in this context. In particular, the chapter explores how ML/DL methods can be leveraged in the engineering phase for designing more secure and safe IoT-enabled long-running technical systems. However, the peculiarities of IoT environments (e.g., resource-constrained devices with limited memory, energy and computational capabilities) still represent a barrier to the adoption of these methods. Thus, this chapter also discusses the limitations of ML/DL methods for IoT security and how they might be overcome in future work by pursuing the suggested research directions.

Key words: Machine Learning; Deep Learning; Security Threats in IoT

16.1 Introduction

The Internet of Things is envisioned as a multitude of heterogeneous devices densely interconnected and communicating with the objective of accomplishing a diverse range of objectives, often collaboratively. The term “Internet of things” was used

Aris S. Lalos, Athanasios Kalogeras, Christos Koulamas, , Christos Alexakos, Dimitrios Serpanos
Industrial System Institute, Athena Research Center, Platani, Patras, Greece e-mail: *surname*@isi.gr

Christos Tselios
Citrix Systems, Patras, Greece e-mail: christos.tselios@citrix.com

for the first time in Mr. Kevin Ashton's presentation in 1999[1] while a significant milestone from the perspective of the IoT was the period between the years 2008 and 2009, when, according to the Cisco estimation, the number of devices (in general) connected to the Internet exceeds the number of the world's population (Evans, 2011). The advent of IoT is accompanied by a number of developments: miniaturization of devices and sensors, increasing mobility of devices, wearable devices, ubiquitous robotics and growing automation of all functions of IoT, presenting numerous benefits in a diverse number of applications ranging from smart homes, smart health and energy management to connected cars and smart farming.

As a term, Industrial IoT has been introduced to describe the application of IoT in the industry, namely the utilization of disruptive elements such as sensors, actuators, control systems, machine-to-machine communication interfaces and enhanced security mechanisms to improve industrial systems and shape the futuristic Smart Factory concept. The proliferation of IoT in industrial environments and value chains will allow companies, manufactures and workers to operate in a more efficient manner and will have a great impact in several fields, such as automation, industrial manufacturing, logistics, business processes, process management and transportation (Schmidt et al., 2015). Along with the overall expansion of the core manufacturing process, the digital transformation advancements and the constantly rising node interconnectivity allows new applications to emerge, mostly related to (i) process automation and optimization, (ii) optimized resource consumption and (iii) autonomous system generation and security intensification. It is already identified that IIoT radically changes the product lifecycle, thus providing a new way of doing business in general and highly affecting the overall competitiveness of any organization. As mentioned in (Schmidt et al., 2015) IIoT will integrate products and processes in such a way, that will eventually shift the productivity line effectiveness from mass production to mass customization. This simply translates to more modular and configurable products, tailor-made according to specific customer requirements (Jazdi, 2014). In a nutshell, IIoT will transform manufacturing as we know it through innovative and highly agile products and services, that can become partially independent, responsive and interactive, track their activity in real-time and optimize the whole value chain into providing relevant status information throughout their lifecycle.

The imminent adoption of the emerging IIoT paradigm will provide a significant boost also to the concept of Industry 4.0, a convoluted technological system that has been gaining significant traction over the last few years. Industry 4.0 can be seen as a superordinate term for describing a novel industrial paradigm which aims to combine among others Cyber-Physical Manufacturing Systems (CPMS), omnipresent and time-sensitive networks, Robotics, Big Data analytics and edge computing paradigms. The adoption of these technological pillars is crucial for the development of a highly intelligent manufacturing process, that will incorporate machines, sensors, production modules and incomplete products, all enhanced with the ability to independently exchange information, trigger actions and control each

[1] "I could be wrong, but I'm fairly sure the phrase 'Internet of Things' started life as the title of a presentation I made at Procter & Gamble (P&G) in 1999", Kevin Ashton, RFID Journal, 22 June 2009.

other, thus creating a fully automated, optimized and independent manufacturing environment (Weyer et al., 2015). The Industrial IoT is a key element of Industry 4.0, bringing together modern sensor technology, fog - cloud computing platforms, and AI to create intelligent, self-optimizing industrial equipment and facilities.

The aforementioned advancements can be definitely perceived as a big blessing, however, big challenges also arise related to the dynamic management and security mechanisms of Industrial IoT (IIoT) components across heterogeneous objects, transmission technologies, and networking architectures. Another major area of concern is privacy with regards to personal information that will potentially reside on networks, also a likely target for cyber criminals. Finally, it should be mentioned that IoT allows the virtual world to interact with the physical world and this brings big safety issues. Machine and deep learning (ML/DL) have advanced considerably over the last few years (Jordan and Mitchell, 2015; Goodfellow et al., 2016) and machine intelligence has transitioned from laboratory curiosity to practical machinery in several important applications. The ability to monitor IoT devices intelligently provides a significant solution to new or zero-day attacks. ML and DL are powerful methods of data exploration for learning about 'normal' and 'abnormal' behaviour according to how IoT components and devices perform within the IoT environment. Consequently, these methods are important in transforming the security of IoT systems from merely facilitating secure communication between devices to security-based intelligence systems.

The goal of this chapter is to provide a comprehensive survey of ML methods and recent advances in DL methods that can be used to develop enhanced security methods for modern IoT and IIoT systems that are used in smart manufacturing environments. IoT security threats, either inherent or newly introduced, are presented, and various potential IoT system areas of the attack surface and the possible threats and vulnerabilities are discussed. A thorough discussion of the opportunities and challenges involved in applying ML/DL to IoT security is offered. The presented solutions and challenges are expected to provide a novel insight at a key area with renewed research interest, where high potential for novel improvements is feasible in the near future.

The rest part of this chapter is organized as follows: Section 16.2 provides a review of different layered architecture for IoT and IIoT systems. A comprehensive discussion on the potential vulnerabilities and areas of the attack surface of IoT systems is provided in Section 16.3. Section 16.4 presents an in-depth review of the Machine Learning (ML) and recent advances in Deep Learning (DL) methods that have been applied for identifying IoT security threats and vulnerabilities. IoT physical world applications and Safety Challenges are presented in Section 16.5, and conclusions are drawn in Section 16.6.

16.2 IoT and IIoT layered Architecture Review

The cornerstone for the successful design and deployment of IoT infrastructure and relevant IoT applications, is the efficient combination of cutting-edge technological achievements in the areas of networks, hardware and informatics (Atzori et al., 2010). Only hierarchical, modular, loosely coupled, flexible and scalable system architectures can manage and coordinate this complex system of different components, networks, data, and software. From the architectural perspective, the first approaches of IoT ecosystems deploy the Service Oriented Architecture (SOA) as the inspiration for designing and implementing their IoT solutions (Xu et al., 2014). SOA key idea is the fact that each system exposes its independent functionalities in terms of web services, which can be invoked by other systems over computer networks. IoT consists of devices (systems) that are connected through networking. Thus, SOA is considered appropriate to support IoT at the early years (Atzori et al., 2010; Miorandi et al., 2012).

The evolution of IoT brought new challenges such as utilization of limited computational resources, low power consumption, networked devices distributed in a large geographical area, real-time and latency sensitivity, collection and processing of large amount of data, new business models and social requirements. Although the multi-layer SOA architecture provided a workable solution for IoT, these new challenges forced researchers to seek out alternatives to the SOA. After a decade of IoT existence, there is no widely accepted reference architecture that is established as a standardized design approach for IoT. Closer to SOA, most of researchers' opinions about conventional IoT architecture (Mashal et al., 2015; Mainetti et al., 2011; Wu et al., 2010) follow a three-layer approach which comprises:

i. The *perception (or sensing) layer* being the physical layer, consisting of smart objects/devices such as sensors and actuators that are able for sensing and gathering information about the environment as well as interacting with it and its elements.
ii. The *network layer* realizing the connection and communication of the smart objects, network devices, and servers. Furthermore, the network layer is responsible for the transmission and processing of sensor data.
iii. The *application layer* consisting of applications that deliver IoT-based services to the end users, including smart homes, smart energy, smart health and smart cities.

As the three-layer architecture, due to its simplicity, was a popular solution, researchers identify that the complexity of orchestrating the large number of smart devices, as well as the size of associated information, cannot be handled efficiently at the network or application layer. The solution was the introduction of a layer between them, usually named as *middleware layer*, thus defining a four-layer architecture. This layer is responsible for service management and storage of data, as well as for decision making based on the results of information processing. Such a paradigm is the IoT reference architecture proposed by ITU-T (International Telecommunications Union - Telecommunication Standardization Sector) (ITY-T, 2012), where the Service and

Application support layer (Middleware layer) provides generic services, such as data processing or data storage, and application-specific services, which cater for the requirements of diversified applications.

The four-layer model provides the flexibility in designing IoT applications, overcoming the most of technical challenges. But the IoT applications are more complex than the classic computer applications regarding to their target users. Due to their nature, IoT applications involve many collaborative devices satisfying the needs of various stakeholders as end-users (Evans, 2011), meaning that different user requirements must be met by a distributed network of heterogeneous nodes. Organizations from both private and public sector, or even individual citizens are some examples of potential end users of an IoT application, i.e a smart city paradigm. The diversity of business requirements and the social impact of the IoT applications, lead to the specification of another layer on top of the application layer, separating the data analysis and machine and deep learning from the business models that provide this data to the users. The commonly known Business Layer has to do with the conversion of the data received by application level to meaningful services to the different group of users (Wu et al., 2010; Sethi and Sarangi, 2017; Aazam et al., 2014; Khan et al., 2012). Furthermore, data analytics provide insights with practical and useful knowledge to the users. Furthermore, data access management and users' privacy are some of the most important features of this layer.

The evolution of the layered IoT architecture was unavoidable, and the addition of new layers permitted both the inclusion of all the factors that affect the operation of IoT applications and the development of technologies and tools to deal with the modern challenges. Figure 16.1 shows the evolution of IoT reference architectures, ending with the five-layered architecture. At this point, it should be mentioned that IoT

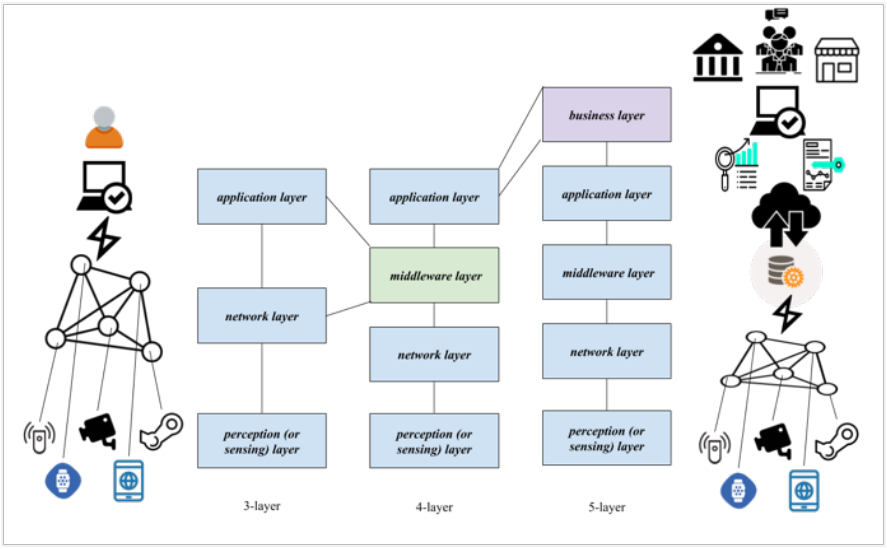

Fig. 16.1 Evolution of IoT Layered Architecture

is one of the technologies, that was rapidly integrated by the industry into its products. Today, following academia's paradigm and the concept of the Fourth Industrial Revolution (4IR), the Industry took large steps towards the well-known Industrial Internet of Things (IIoT) by establishing large and complex IIoT applications in various deployment areas (i.e. cities, energy grid, buildings, manufacturing, etc.). Although, the industry is favorable to work with standards, still there is a lack of standardization relevant to the architectural design of IIoT applications. Nevertheless, significant consortiums, consisting of the key industry players, were created worldwidely in order to define such standards. The Industrial Internet Consortium (IIC)[2] (US) and the Industrie 4.0 Platform[3] (Europe) are two of the mainstream initiatives towards standardization of IIoT systems, supplemented by further initiatives such as Japan's Society 5.0[4] and Made in China 2025[5]. As early results, each of the first two initiatives have proposed IIoT Architecture reference models providing a guidance by specifications for the development of system and application architectures.

The Industry 4.0 Platform introduced the Reference Architectural Model Industry 4.0 (RAMI 4.0) (Adolphs et al., 2016). RAMI 4.0 is recognized as a DIN standard (DIN SPEC 91345) and an international pre-standard (IEC PAS 63088). RAMI 4.0 is based on a three-dimensional model covering all the industrial aspects from the industrial hierarchy to the product life cycle. Its three dimensions are: a) the Hierarchy defining the functional areas of the IIoT applications selecting from Smart Product, Smart Factory and Connected World; b) Architecture, which provides the system architecture, and finally c) the Product Life Cycle, which covers development, production, and maintenance aspects. Focusing on the Architecture dimension, RAMI 4.0 defines six-layers:

- The *Asset Layer* representing the physical layer including devices and their hardware parts as well as the human factor.
- The *Integration Layer* defining the provision of informational data and asset control services.
- The *Communication Layer* applying standardized communication between the assets and the applications at the higher layer, always following the formality of the information at the Integration Layer.
- The *Information Layer* dealing with the pre-processing of the information and the generation of events. In the case of events, the asset control services may be invoked.
- The *Functional Layer* receiving pre-processed information from the Information Layer and implementing rules and decision-making logic. Furthermore, Functional Layer is the only remote access point to the data as in the layers below the data is protected for ensuring information integrity.
- The *Business Layer* being the layer where the functions of the Functional Layer are integrated to the business processes.

[2] https://www.iiconsortium.org/

[3] https://www.plattform-i40.de/

[4] https://www8.cao.go.jp/cstp/english/society5_0/index.html

[5] http://english.gov.cn/2016special/madeinchina2025/

In the USA, the Industrial Internet Consortium (IIC) proposed the Industrial Internet Reference Architecture (IIRA) (Lin et al., 2017a). Contrary to the RAMI 4.0, which is specialized in the manufacturing business processes, IIRA deals with a wider range of IIoT applications, from transportation to energy. IIRA also follows a three dimensional model, but with a different approach to RAMI 4.0. Its three dimensions are the a) *Product Life Cycle*; b) *the Industrial Sectors* that define the area of deployment and c) *a four-level layer consisting of viewpoints*, each one associated with particular stakeholders and their concerns. The Business viewpoint deals with business-oriented aspects, such as business value, expected return on investment, cost of maintenance, and product liability. The realization of the key capabilities defined by the Business viewpoint is the main concern of the Usage viewpoint. The next viewpoint, the Functional viewpoint, deals with system functional components, interfaces, and interactions. The last viewpoint, the Implementation viewpoint is concerned with the technologies and system components required, implementing the functional requirements defined at the Functional viewpoint. IIRA is a general reference model, which doesn't define a specific architecture but proposes some architectural patterns that can be used to deal with functional requirements of an IIoT application. These three patterns are: a) Three-tier architecture pattern, b) Gateway-Mediated Edge Connectivity and Management architecture pattern, and c) Layered Databus pattern.

The existence of two different reference architectures led to the collaboration of the two involved consortiums towards the publication of mapping and alignment guidelines between RAMI 4.0 and IIRA (Lin et al., 2017b). From the architectural perspective, this effort focuses on the mapping of the functional blocks that can be defined in the IIRA model with the layers of Architecture dimension of RAMI 4.0. In the context of this model alignment, Figure 16.2 presents a mapping of the functional blocks of the IIoT architecture with the aforementioned IoT five-layer architecture. The adaptation of a layered architecture for the design of an IIoT infrastructure assists

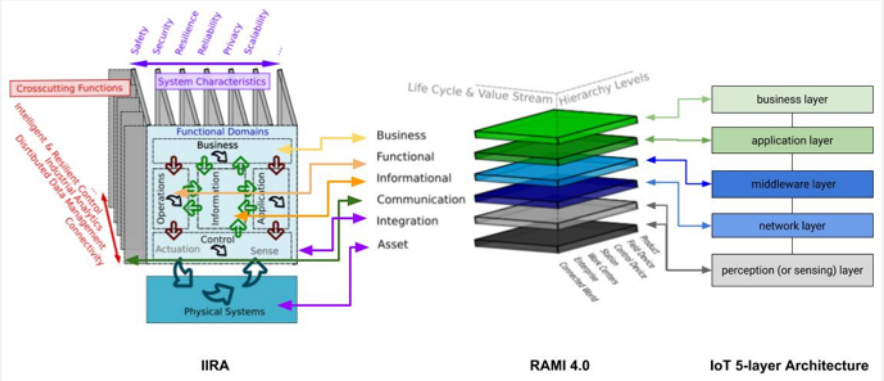

Fig. 16.2 Mapping between IIRA, RAMI 4.0 and IoT five layered architecture. Partially taken from (Lin et al., 2017b)

the engineers to clarify both the technologies that will be used at each layer and the implementation of the provided operations/services. Due to the difference of the functionalities and of the technologies/standards used at each layer, the abstraction of an IIoT infrastructure to independent layers allows the examination of security vulnerabilities and safety challenges separately for each layer. The following sections deal with the challenges of IoT security threats and vulnerabilities classifying them in the basic layers of an IoT Architecture.

16.3 IoT Security Threats and Vulnerabilities

As with any IT system, the principal information security requirements of availability, integrity, confidentiality, authentication/authorization and non-repudiation, constitute critical requirements of IoT based systems as well. However, the specific characteristics of IoT system components define a well-differentiated domain, requiring thus unique approaches in identifying threats and vulnerabilities (Xu et al., 2014), as well as in detecting and responding to relevant attacks, in order to guarantee the trustworthiness of modern IoT based systems. Important specificities of IoT systems are related to the typical involvement of i) a large number of resource constrained, wirelessly networked, miniaturized embedded devices, ii) distributed and/or centralized Big-Data processing infrastructures introducing significant security challenges. In such systems, these challenges become even harder to be addressed, due to the criticality of supported applications, considering also the Industrial IoT (IIoT) and its applications in critical infrastructures, as well as other systems in avionics, automotive and medical equipment domains, where safety, reliability and resilience are of highest priority. There are already various existing studies and proposals in the literature to identify the peculiarities of IoT security threats (Humayed et al., 2017; Mena et al., 2018; Chen et al., 2018). Among the most representative efforts to structure the typically extensive threat taxonomies, in a way tailored to the IoT specifics, are these of the ENISA agency (ENISA Report, 2017, 2018a,b,c) and the OWASP-IoT project (OWASP, 2018), which are also referenced by the IIC Security Framework Architecture document (IIC, 2016). According to (ENISA Report, 2017, 2018a,b,c) there are 8-9 high-level threat groups, and a large number of identified threats, depending on the case, while in (OWASP, 2018) there are 18 identified areas of the attack surface, and a multitude of possible vulnerabilities.

For a smart manufacturing application context case, the different threats identified by ENISA, are grouped under the following high-level threat categories (ENISA Report, 2018a):

- **Nefarious Activity / Abuse:** It classifies the most widely known threats, such as the *Denial of Service (DoS), malware, manipulation of hardware and software, manipulation of information, personal data abuse, brute force*, and other *targeted* attacks.
- **Eavesdropping / Interception / Hijacking:** This group contains main network related threats, including the *Man-in-the-Middle* attacks or *session hijacking*,

which involve eavesdropping and actively relaying of messages accompanied possibly by modifications or deletion of the transmitted data. It also contains *protocol hijacking* and *network reconnaissance*, which mainly lead to information leakage, including information related to passwords or network structure.
- **Physical Attacks:** It includes threats related to *device modifications*, such as tampering physically unsecured ports, and *device destruction* or theft (i.e. the attacker's goal is typically sabotage) attacks.
- **Unintentional Damage:** Unintentional changes of data or configuration or erroneous use and administration of devices and systems, as well as damages caused by a third party, such as a maintenance subcontractor or a manufacturer software update, are all considered as threats of this category.
- **Failures or malfunctions:** This category describes the threats of a general device failure, either at the sensor/actuator, or at the control system level. It also contains malfunctions due to various uncategorized *software vulnerabilities*, e.g. due to *weak or default passwords, software bugs* and *configuration errors*, as well as failures due to *services* which the system depends on.
- **Outages:** This group includes the loss of availability of communication links, or power supply, as well as of higher level needed support services.
- **Disaster:** *Natural disasters* (floods, landslides etc.), as well as other *environmental disasters* related to the immediate IoT equipment environment, fall under this threat group.
- **Legal:** It refers to threats related to violation of *rules and regulations*, or *breach of legislation* and *abuse of personal data*, as well as to threats related to *failures* to meet *contractual requirements*, all leading to possible financial losses either direct (fines) or indirect (reputation).

On a different perspective, the OWASP-IoT approach starts from the definition of the set of areas of the attack surface, for which then the various vulnerabilities are enumerated. The attack surface list is rather elaborate and includes (OWASP, 2018):

- **Ecosystem (general):** interoperability standards, security enrollment, system decommissioning, lost access procedures, and other system wide vulnerabilities
- **Device Memory:** Leakage of sensitive data (various types of credentials)
- **Device Physical Interfaces:** Firmware extraction, command interfaces, privilege escalations, tamper resistance, removable storage media, debug ports, and device ID exposure
- **Device Web Interface:** Code injection, broken authentication, sensitive data exposure, broken access control, security misconfigurations, cross-site scripting, insecure deserialization, vulnerable components, insufficient logging and monitoring, credential management
- **Device Firmware:** Sensitive data exposure, backdoor accounts, hardcoded credentials, encryption implementation, vulnerable services due to old software versions, security API exposure, firmware downgrades
- **Device Network Services:** Information disclosure, command interfaces, injection, DoS, unencrypted channels, poor encryption implementations, existence of

development/test services, OTA update blocks, replay, no payload verification, no integrity checks, credential management
- **Administrative Interface:** Common web interface vulnerabilities, credential management, security/encryption options, logging options, two-factor authentication, insecure direct object references, inability to wipe device
- **Local Data Storage:** Unencrypted or weakly encrypted data, discovered keys, no integrity checks, static keys
- **Cloud Web Interface:** Common web interface vulnerabilities, credential management, transport encryption, two-factor authentication
- **Third-party Backend APIs:** Device information leakage, location leakage
- **Update Mechanism:** Unencrypted updates, not signed, verified or authenticated updates, malicious updates, missing update mechanisms, no manual update mechanisms
- **Mobile Application:** Implicit trusts, username enumeration, account lockout, default credentials, weak passwords, insecure data storage, transport encryption, insecure password recovery, two-factor authentication
- **Vendor Backend APIs:** Inherent trusts, weak authentication and access controls, Injection attacks, hidden services
- **Ecosystem Communication:** Heath checks, heartbeats, de-provisioning, updates
- **Network Traffic:** Protocol fuzzing, wireless medium, range
- **Authentication/Authorization:** Data disclosure or reuse, multiple schemes, weak authentication
- **Privacy:** Data disclosure
- **Hardware (Sensors):** Sensing environment manipulation, physical tampering and damage

Other taxonomies may follow a threat classification based on a purpose / target threat model, as in (Humayed et al., 2017), where the five threat classes are Criminal, Financial, Political, Privacy and Physical threats, followed by a detailed enumeration of application domain specific, physical, cyber and cyber-physical vulnerabilities. Alternatively, they follow a layered approach, as in (Chen et al., 2018), where the attack threats are categorized on a four-layer basis:

- **Application Layer:** Code Injection, Buffer overflow, Sensitive data Permission / Manipulation
- **Middleware Layer:** Flooding attack, cloud malware injection, signature wrapping attack, web browser attack, SQL injection attack
- **Network Layer:** Traffic Analysis, Sniffing attack, DoS, Sybil, Sinkhole, Replay, Man-in-the-Middle attacks
- **Perception Layer:** Unauthorized Tag Access, Tag cloning, Eavesdropping, RF Jamming, Spoofing attack

Finally, as the overall IoT architecture contains also typical web components and interfaces, detailed classifications, that apply to the wider web environment, may also get into the picture (WASC, 2012).

Attempting to organize the broad set of threats and areas of the attack surface under the structural view presented in the previous section, Table 16.1 can be constructed.

Table 16.1 Classification of Vulnerabilities and Threats in modern IoT and IIoT systems

Vulnerabilities, Threats	Physical	Cyber	
Attack surface		Passive	Active
Physical Device	Modifications Destruction Tampering Theft Failure Malfunction Power Outage Link Outage Environmental Disasters Natural Disasters	HW/SW Failure Personal Data Leakage Unauthorized Tag Access	DoS Malware False Data Injection HW/SW Manipulation Info. manipulation Personal Data Abuse Brute Force Attacks Tag Clonning
Network Service	Failure Malfunction Environmental Disasters Natural Disasters Power Outage Link Outage	Network Reconnaissance Traffic Analysis Eavesdropping Sniffing	DoS Man in the Middle Session Hijacking Protocol Hijacking False Data Injection Sybil Sinkhole Replay Spoofing RF Jamming
Cloud, Web and Application Service	Failure Malfunction Environmental Disasters Natural Disasters Power Outage Link Outage	HW/SW Failure Personal Data Leakage	DoS Malware HW/SW Manipulation Info. manipulation Personal Data Abuse Brute Force & Targeted attacks Code Injection Buffer overflow Signature wrapping Web Browser attack SQL injection attack

16.4 Detailed Review of ML and DL methods for securing IoT Systems

Pervasive sensors continuously collecting massive amounts of information have rendered data-driven learning increasingly important. Learning algorithms focus on the construction of schemes that progress automatically through experience (Jordan

and Mitchell, 2015). Machine and deep learning approaches have been widely applied in a surprising number of applications including medical, financial and automotive industry, and recently they are finding their way into the manufacturing industry, providing from increased production capacity to more efficient plant operation and everything in between (Sharp et al., 2018).

Machine-learning algorithms are usually classified into the following learning categories: supervised, unsupervised, semi-supervised, active, and reinforcement, as it is shown also in Figure 16.3. Supervised algorithms are used for learning a function that maps an input to an output, based on several input-output pairs known as training data. Supervised learning approaches are applied for solving classification and regression problems, where the output variable is either a category (e.g., "threat" or "no threat") or a real value. Unsupervised learning algorithms model the underlying structure or distribution of data to learn more about the data without using any corresponding output variables. The unsupervised learning problems are further grouped into clustering and association problems. In the clustering case the goal is to discover inherent groupings in the data, while in the association case the focus is on finding rules that describe large portions of data, like for example learning temporal state-based specifications for electric power systems to accurately differentiate between disturbances, normal control operations, and cyber-attacks (Pan et al., 2015). Active learning emphasizes on learning from limited amount of training samples, based on the experience of users that play the role of "omniscient" to label the selected data (Yang et al., 2018). It is naturally suited for the design of Intrusion Detection Systems, provided that the labeling process for intrusion detection is either a very time-consuming process or even impossible for cases that intrusion never happened before. Active learning boosts the power of machine learning by exploiting the experience of a domain expert, significantly decreasing the labeling efforts and increasing at the same time the reliability of a supervised learning model for intrusion detection. Finally, in reinforcement learning uses a software agent that learns an optimal policy of actions over the set of states in an environment. Depending on the performed action, the environment sends a reward to the agent, while each agent tries to maximize its rewards over time by choosing action that results in higher rewards. This approach has been widely adopted to obtain optimal or near-optimal, integrated maintenance and production control policies for deteriorating, stochastic production/inventory systems (Xanthopoulos et al., 2018).

Deep learning is a subcategory of machine learning, that focuses on learning data representations. Most deep learning approaches are based on artificial neural networks and more specifically they use a cascade of multiple layers of non-linear processing units for extracting informative features. Successive layers use the output from the previous layer as input. DL approaches can be also classified into supervised and unsupervised schemes and their main characteristic is their ability to learn multiple machine and deeplevels that correspond to different levels of abstraction. In the following part of this section we discuss both ML and DL approaches, to provide readers with in-depth review of both of them, and we focus on applications in securing Industrial IoT systems.

16.4.1 Machine learning (ML) methods for IIoT security

This subsection focuses on the presentation of the most common ML approaches including decision trees, support vector machines, Bayesian algorithms, k-nearest neighbors and random forests. More specifically, we will briefly describe their strengths and weakness and the threats that are usually detected in IIoT security challenges.

Decision Trees (DTs) are used for solving classification problems by sorting samples according to some indicative feature values. Each vertex (node) in a tree represents a feature, and each edge (branch) denotes a value that is assigned to the vertex corresponding to the sample that needs to be classified. The samples are then classified starting from the origin vertex and with respect to their feature values. The identification of the optimal feature is based on different metrics including information gain (Quinlan, 1986) and Gini index (Du and Zhan, 2002). Despite their wide adoption in different security applications, including intrusion detection (Kim et al., 2014) and detection of suspicious traffic sources (Alharbi et al., 2017), they usually involve a massive construction of trees with several decision nodes, increasing significantly the computation and storage requirements.

Support Vector Machines (SVMs) classify samples by assigning them to points in space and creating a splitting hyperplane between two or more classes, such that the distance between the hyperplane and the most adjacent points of each class is maximized and thus the separate classes are divided by a clear gap that is as wide as possible. Although they are fairly robust against overfitting, especially in high dimensional space, it is trickier to be tuned due to the importance of selecting the right parameters and do not scale well to larger datasets. SVMs have been employed to improve the effectiveness of prediction and diagnosis of induction motor faults particularly during the maintenance judgment process (Gangsar and Tiwari, 2017), to detect attacks in a smart grid (Ozay et al., 2016) or as a tool to exploit IoT device security (Lerman et al., 2015).

Bayesian Methods: Bayes' theorem describes the probability of an event based on previous information related to the event. **Naive Bayes (NB)** is a well-known ML technique for constructing classifiers that calculate the posterior probability of an event and use the Bayes theorem to evaluate the probability that a particular feature set of unlabeled samples fits a specific label, assuming independence among features. For example, NB can be used for classifying network traffic as normal or abnormal, using as features the connection duration, the connection protocol (e.g., TCP, UDP), the connection status flag. These features are considered independent although in practice there are dependencies. The aforementioned schemes can be easily implemented and have been applied both in binary and multiclass problems, though they completely ignore interactions among features, which in many complex tasks contribute in increasing the discrimination power of a classification model (Ng and Jordan, 2001). They have been successfully applied for identifying nefarious Activity / Abuse in smart manufacturing systems, including malware attacks (Ye et al., 2017).

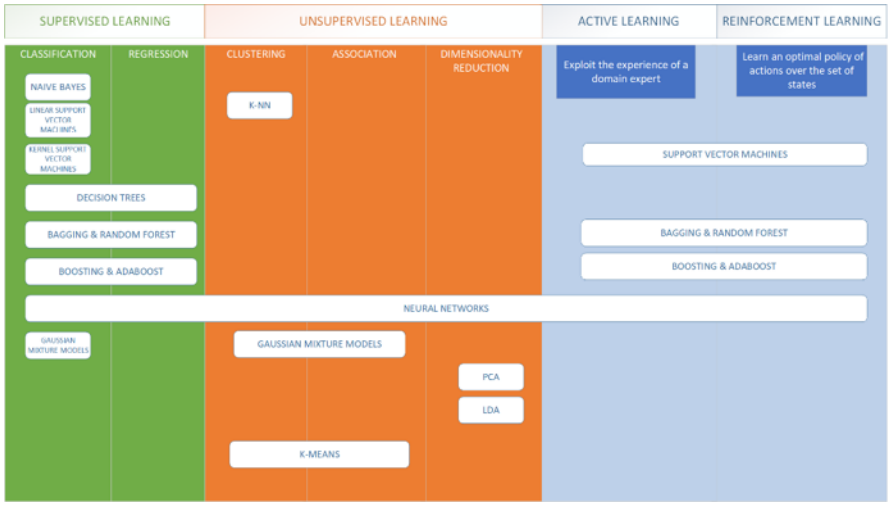

Fig. 16.3 Categorization of different ML applications based on their working principle

k-Nearest Neighbor Classifiers categorize data patterns (e.g., node behavior's) in normal and malicious, on the basis of the votes provided by a selected number (i.e., k) of its nearest neighbors. Euclidean distance between features representing different patterns is usually adopted for identifying neighborhoods. Determining the optimal value of k may become a challenging and time consuming process, though this method has been successfully applied in several network intrusion and anomaly detection schemes (Syarif and Gata, 2017; Su, 2011).

Random Forests are supervised learning algorithms that use several Decision Trees DTs to combine decisions and thus acquire precise and more reliable classification results. They are composed of several trees which are constructed randomly and they are trained to vote for an output class. The class with the most votes is selected as the final classification output. The number of required trees depends on the size of the training dataset and the construction of several trees may be impractical in several real time applications. Several works suggest their application for DDoS detection (Doshi et al., 2018) and for detection of unauthorized IoT devices (Meidan et al., 2017).

Other ML approaches, such as **Association rule, Ensemble learning, Bugging and Boosting** and **k-means clustering** algorithms, have been also applied for the detection of intrusion, anomalies and malware, as well for improving the efficiency of protection of data, without reducing the quality of anonymization (Xie et al., 2017).

16.4.2 Deep learning (DL) methods for IIoT security

Recently, several researchers, system engineers and software developers have shown increasing interest in the application of DL approaches for addressing security threats and vulnerabilities in modern IoT systems. This phenomenon is mainly attributed to their superior performance over traditional ML schemes, especially when both methods utilize large datasets. DL approaches are capable of learning data representations with several levels of abstraction by using computational architectures with several non-linear processing layers. This is also the reason why they are known as hierarchical learning methods. Most modern deep learning methods are based on Neural Networks (NNs) (please refer to Figure 4), while learning can be supervised, alternatively known as discriminative (e.g. Convolutional and recurrent NN), unsupervised (generative learning, e.g. generative adversarial networks) or semi-supervised (e.g. auto-encoders, deep belief networks, restricted Boltzmann machines). In the remaining part of this section we briefly review the working principles of the aforementioned DL schemes and their potential application for identifying different IoT security threats.

Convolutional Neural Networks (CNNs) focus on reducing the connection between layers by exploiting sparse interactions, parameter sharing and translation invariant characteristics. They consist of two different type of layers: i) the convolutional layer where data parameters are convolved with multiple filters of equal size, ii) the pooling layer, where different approaches for subsampling the output and decreasing the size of subsequent layers are applied. Their benefits compared to traditional NN are increased scalability and reduced training complexity, while their wide adoption is attributed to their ability to automatically learn features from raw data. Still their complexity is quite high, making their integration to resource constrained devices a very challenging task. CNNs have been successfully utilized for Malware detection (McLaughlin et al., 2017)and for also breaking cryptographic implementations (Maghrebi et al., 2016).

Recurrent Neural Networks (RNNs) have been utilized in applications where the data is available sequentially (e.g. speech, video, sensor measurements). RNNs are created by applying the same set of weights recursively over a differentiable graph-like structure by traversing the structure topologically. They are very efficient in processing data in an adaptive manner, though their main limitation is the issue of vanishing or exploding gradients (Pascanu et al., 2013). RNNs have been previously used for detecting anomalies in time-series based threats, e.g. monitoring network traffic flow to detect potential malicious behaviors (Torres et al., 2016).

An Auto-encoder (AE) is a NN composed of two parts, the encoder and the decoder, which obtains the input and provides an abstraction (code) as an output and vice versa. The encoding and decoding weights are selected by minimizing the error between the encoder's input and the decoder's output. AEs are important for feature extraction and dimensionality reduction without any data prior knowledge, though in order to operate satisfactory the training dataset should be representative of the testing dataset, while they also consume considerable computation time. Previous studies have used AEs to extract features, which were proven informative

for detecting impersonation attacks in Wi-Fi environments (Aminanto et al., 2018) and cyberattacks in Fog Computing systems (Abeshu and Chilamkurti, 2018).

Restricted Boltzmann Machines (RBMs) are deep generative models utilized for learning a probability distribution over the input data. They are undirected models, while there is no link between any nodes in the same layer. They consist of visible and hidden layers and they hierarchically understand features from data. Again, their complexity is increased making their integration to resource constrained devices a challenging task. The most common applications that use RBMs are related to network anomaly detection (Fiore et al., 2013).

Generative adversarial networks (GANs) have recently emerged as a promising DL approach. GANs are based on the training and use of two different models called generative and discriminative models. The generative model goal is to learn a distribution over the input dataset and generate a data sample and the discriminative model prediction whether the input is from the dataset or from the generative model. GANs generate samples very fast, though its training is hard and usually unstable. Despite this drawback, GANs have been used to build an architecture for securing an IoT system cyberspace (Hiromoto et al., 2017). GANs have a potential application in IoT security, since they are capable of learning different attack scenarios and generate samples similar to a zero-day attack (e.g. variations of existing attacks), providing security approaches that are robust against unknown attacks (Zenati et al., 2018). All the aforementioned ML and DL approaches provide solutions for detecting threats on how IoT devices interact with each other and with the environment, using the data collected by different heterogeneous devices that can be integrated in dynamic environments. Table 16.2 summarizes the various security/vulnerability threats that are detected using aforementioned ML and DL approaches.

In Figure 16.4 we present different DL NN-based architectures for detecting threats in IIoT systems. Red nodes indicate the classification output (e.g. Normal, malicious behavior e.t.c.), (light) green nodes correspond to (probabilistic) hidden layer, while blue nodes denote recurrent cells, and purple nodes correspond to convolutional cells.

16.5 Achieving Safety using ML and DL approaches

Learning from large volumes of data using powerful algorithms, as those presented above, brings significant benefits in securing IIoT systems, though questions about safety still need to be carefully examined. Although workhorse machine and deep learning tools are expected to have intelligence that in many cases surpasses human abilities or something in between, they are still technological components that have to be engineered with safety in mind (Conn, 2015). The term “safety” is widely used in a large number of diverse engineering disciplines, indicating the absence of system failures or the absence of dangerous conditions. Authors in (Maller and Hansson, 2008) introduce a decision-theoretic definition of safety, making a link to the minimization or reduction of risk and uncertainty to undesirable states which can

Table 16.2 Summary of studies on ML and DL for securing IoT and IIoT

Reference	Method	Threats Detected or Security Application	Areas of the attack surface			
			Physical device	Network service	Cloud service	Web service
(Kim et al., 2014)	DT	Intrusion Detection	✓	✓	-	-
(Alharbi et al., 2017)	DT	Denial of Service	✓	✓	✓	-
(Gangsar and Tiwari, 2017)	SVM	Fault Prediction	✓	-	-	-
(Ozay et al., 2016)	SVM	False Data Injection	-	-	✓	✓
(Lerman et al., 2015)	SVM	Attacks to Masked Advanced Encryption Schemes (AES)	-	-	-	✓
(Ye et al., 2017)	NB	Malware Attack	✓	-	-	✓
(Syarif and Gata, 2017)	kNN	Intrusion Detection	✓	✓	-	-
(Su, 2011)	kNN	Denial of Service	✓	✓	✓	-
(Doshi et al., 2018)	RF	Denial of Service	✓	✓	✓	-
(Meidan et al., 2017)	RF	Unauthorized Access	-	-	-	✓
(Maghrebi et al., 2016)	CNN	Masked AES Attacks	-	-	✓	✓
(McLaughlin et al., 2017)	CNN	Malware Attacks	✓	-	-	✓
(Torres et al., 2016)	RNN	Malicious Behaviour	-	-	✓	✓
(Aminanto et al., 2018)	AE	Anomaly-based IDS	-	✓	-	-
(Abeshu and Chilamkurti, 2018)	AE	Fog Cyberattacks	-	✓	✓	-
(Fiore et al., 2013)	RBM	Network Anomaly Detection	-	✓	-	-
(Hiromoto et al., 2017)	GAN	Vulnerabilities to malicious supply chain risk	✓	-	-	✓

be considered as harmful. This generic definition applies to many different domains and systems and indicates that the cost of undesirable states is expected to be quite high in a human sense for events that are harmful, and that safety is achieved by minimizing the probability of both expected and unexpected harms.

In IIoT, safety is related to i) the ability of reasoning about the behavior of the IIoT devices and more specifically that of the actuators, ii) the ability of identifying and preventing unintended and unexpected failures or harmful events. Those are very hard challenges, since they require the system(s) to be able to identify "normal" behaviors and at the same time develop device interaction approaches, mechanisms that enforce safety properties. More importantly, they usually become even harder to be addressed, due to their heterogeneity, the lack of standardization and the ineffectiveness of traditional defense mechanisms, including firewalls and antivirus software.

The strategies that could be applied for ensuring safety are strongly related to the specific application, though the authors in (Maller and Hansson, 2008) have analyzed different strategies across different domains suggesting four main categories of safety approaches. The first approach known as safe design, suggests the exclusion instead of the control of a hazard (e.g. excluding hydrogen from the buoyant material of

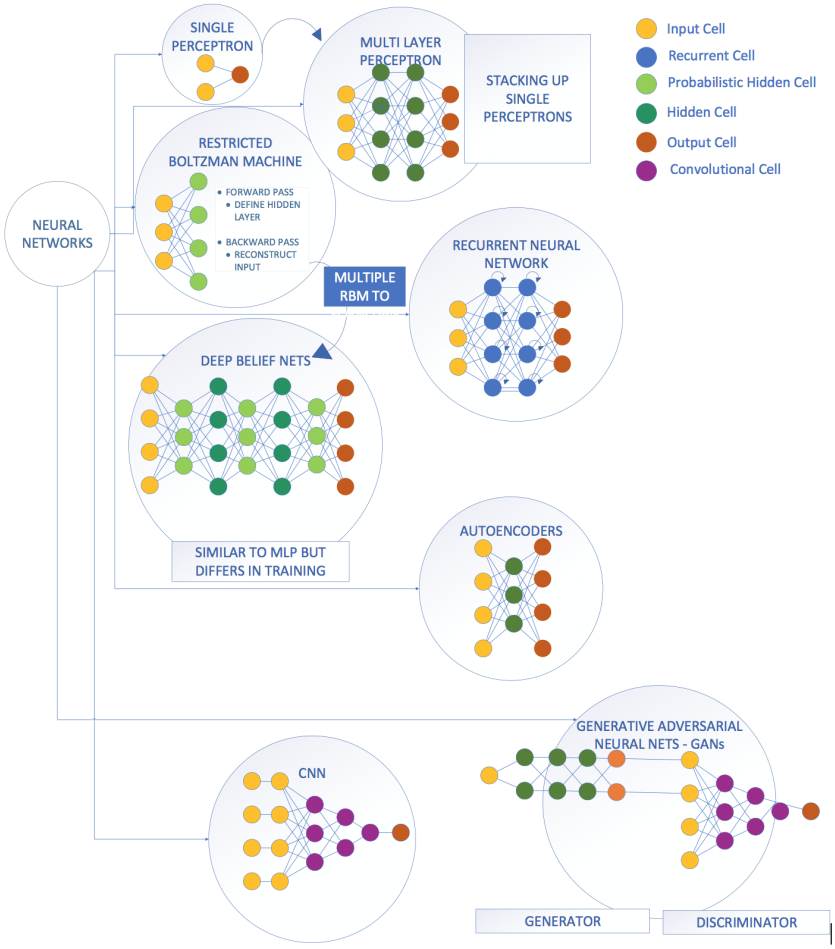

Fig. 16.4 Different DL NN-based working principles for Detecting threats in IIoT systems. Red nodes indicate the classification output (e.g., Normal, malicious behavior e.t.c.)

a dirigible airship ensures safety). The second one suggests using multiplicative or additive reserves, known also as safety factors/margins. A safety factor in mechanical engineering is a ratio between the maximum load that does not lead to failure and the optimal load that the system was designed to support, while the corresponding safety margin is determined as the difference between the two. The third category is known as the "safe fail strategy" according to which a system remains safe, even when it fails in its intended operation (e.g. dead man's switches on trains, safety valves on boilers e.t.c.). The fourth final category, suggests including measures, known as procedural safeguards, that are beyond the ones designed in the core functionality of the system, such us audit diagrams, posted warnings, e.t.c. In the rest part of this section we

provide details about deploying these strategies using machine and deep learning methods.

Inherently Safe Design: One of the major goals in the ML context is to provide robust approaches that address the uncertainties when the training set has not sampled from the test distribution. Training dataset may have biases and patterns, which are unknown to the users, will not be present at the test and might lead to unsafe or undesirable operations. Recent approaches including gradient boosting and deep neural networks are capable of exploiting the biases, achieving higher accuracies, however, making safe predictions of unknown shifts in data, incorrect patterns or harmful rules seems to remain a safety challenge (Caruana et al., 2015).

These models are usually complex introducing difficulties in understanding their behavior in such shifts or whether their outcome will be unsafe. Therefore, the most widely adopted best practices to introduce inherently safe design is by deploying models that can be interpreted by humans and by excluding features which are not casually related to the outcome (Freitas, 2014; Rudin, 2014; Athey and Imbens, 2015; Welling, 2015). The use of interpretable models, features or processing approaches that are capable of identifying and excluding irregular patterns are expected to enhance safety. Moreover, the successful identification of variables that are causally linked to the outcome, could lead to exclusion of behaviors which are not part of the true "physics" of the system, ensuring that any undesirable operation can be avoided. At this point it should be noted that post hoc interpretation and repair of complex uninterpretable models is not the decision rule of a decision making process and therefore it does not assure safety via inherently safe design.

Safe Fail Technique: It is used in ML for rejecting options, which are not confident (Varshney et al., 2013). More specifically, the model reports whether it cannot provide a reliable output, thus avoiding any unsafe or undesirable output. In cases that the output of a model is the reject option, then the user intervenes checks and test sample and provides a manual prediction. This actually means that there is an assumption that a distance from the decision boundary is inversely related to confidence. This assumption is valid in parts of feature space with high probability density and large number of training samples, since the decision boundary is located where there is a large overlap in likelihood functions, though parts of the feature space with low density may not contain any training samples at all, introducing uncertainties in the decision boundaries. In this case, the distance from the decision boundary is fairly meaningless and the typical rule for triggering the reject option should be avoided (Attenberg et al., 2015). For a rare combination of features in a test sample, a safe fail strategy is to manually examine the test sample. At this point, it should be noted that manual intervention options are suitable for applications with long time scales, while when working in ms scale, only options similar to dead man's switches that stop operations in a reasonable manner are applicable.

Procedural Safeguards: Two relevant directions in ML and DL that can be deployed for increasing safety are user experience and openness. Despite the fact that many decision making systems in several IIoT applications, are based on ML and DL systems, the operators and the designers of these systems are usually non specialists in the ML and DL domains. However, the definition of the training data

and the set-up of the evaluation procedures have certain constraints that could lead to undesirable outcomes if they are not done correctly. User experience design can certainly guide and warn non specialists to address the aforementioned issues properly, increasing significantly safety. In addition, it's worth mentioning that most ML and DL approaches are open source, allowing their wide deployment and for the possibility also of the public audit, facilitating the identification of safety hazards and potential harms via the examination of the source code. Of course, one should also take into account that the source software is not sufficient, since these approaches are driven by data. Opening therefore data, making them available to be freely used, reused and redistributed by anyone, is a widely adopted procedural safeguard for increasing safety (Shaw, 2015; Kapoor et al., 2015).

16.6 Future challenges, Discussion and Conclusion

The latest advancements in learning approaches facilitated the development of machine and deep learning methods for addressing different security threats and vulnerabilities. However, there are still challenges that need to be addressed for satisfying complex requirements related to the physical devices, the collected data wireless transmission technologies, mobile and cloud architectures, which are described in detail in the following part of this section.

Availability of security related datasets: One of the major challenges that should be addressed in IIoT systems using ML and DL approaches is the extraction and generation of realistic and high quality training data that contain various possible attacks. A vital future research approach towards this direction is the use of crowd sourcing methods for generating datasets related to IoT treats and attacks. This approach could lead to the inclusion of all the potential attacks in rich training datasets that could be used for benchmarking the accuracy of new algorithms. At this point however, it should be also noted that generating collaborative IoT threat dataset that will be continuously updated with new attacks is a challenging task mainly due to the large diversity in the technical characteristics of the various IoT devices. More importantly many privacy concerns also arise, since sensitive and critical information may be shared publicly especially when we focus on industrial and medical IoT devices.

Learning to secure IoT with low quality data: IoT and IIoT systems deploy a large number of heterogeneous connected devices with memory, power and computational constraints that usually affect also the data quality (e.g., data with missing entries, outliers, noise). Therefore, learning to secure IoT systems require effective algorithms capable of handling and learning from noisy and low quality data. Towards this direction, there is clear need for multimodal and effective ML and DL models that are capable of handling heterogeneous data and with contaminated/noisy data segments.

Lifelong Learning for learning IoT threats: IoT and IIoT systems represent dynamic systems were several new devices either join or leave the system for satisfying

the need of various application with evolving needs. Due to their dynamic nature distinguishing between normal and abnormal behaviors cannot be predefined, thus becoming a challenging task. To address this issue, frequent updates of the security models are required in order to track and understand the system modifications. Therefore, lifelong learning is a significant attribute that should be supported in long term real-world applications and it is directed towards the construction of a model that can perform the retraining process repeatedly for the learning of new emerging patterns related to each behavior. The model should be able to continuously adapt to and learn from new environments.

Implementation of ML and DL at the edge: Edge computing is an essential solution that immigrates IoT service solutions to the network edge, minimizing delays, realizing real time processing performance, improving energy efficiency and enhancing the scalability of lightweight IoT devices. Thus, the implementation of DL and ML approaches at the edge for IoT security are expected to offer an effective framework for data processing with reduced network traffic load. Though, there are still significant challenges that need to be addressed for exploiting the benefits of edge computing. The design of ML and DL approaches that can process scalable data representations compatible with adaptive data transmission protocols is an interesting and important direction for improving the performance of transparent computing. In addition, the programming language of the framework should take into account the heterogeneity of hardware and the capacity of the resources in the workflow. Thus an appropriate ML and DL end-to-end framework that will take into account hardware and software reconfigurations is still a challenging problem. Finally, distributed security solutions is still an open direction of research, meaning that future security solutions should not only exploit the capability of edge servers for building more secure IoT devices, but also be able to guarantee the security of the distributed and sometimes resource-constrained edge servers.

Data Security and Privacy Concerns: Because of the everyday and pervasive nature of IoT scenarios, security and privacy concerns take a broader dimension, demanding for cross and multidisciplinary approaches through efforts from different areas in order to bring citizens into the loop. Nowadays, when talking about the strong development of the IoT, most estimates provide very impressive data on the number of interconnected devices in coming years. Consequently, many security and privacy approaches in IoT are proposed from a device perspective with the aim of addressing these concerns in a broader environment. However, the IoT ecosystem is not only composed of communication-enabled devices, but of a huge amount of heterogeneous smart objects, middleware and services, where security and privacy requirements from different actors (citizens, companies, or regulatory bodies) need to be reconciled. Given the degree of heterogeneity, one of the most significant challenges is to build a secure, privacy-aware, but still interoperable IoT framework. Therefore, there is a strong need to move towards a holistic security and privacy approach by addressing the IoT ecosystem as a whole, beyond such device-centric vision.

Yet, industry has already realized that the true value of IoT is not on the physical interconnected devices per se, but on the massive datasets and crude, unrefined

information they contain, and consequently how this hidden commodity can be efficiently processed in a fast and meaningful manner. Through IoT, individuals will produce an unprecedented amount of raw information about their daily routine which can be exploited manifold by operators and malicious eavesdroppers alike. This clearly violates user privacy, especially when considering that the user has willingly purchased the device which now may handle his personal data over to third-party data silos to be further processed. Consequently, it is essential for users to demand and legislative authorities to enforce a certain move towards data-centric security schemes that will penalize paradox and improper data usage. Users also need to be empowered with mechanisms to control how data from their devices are shared, to whom, and under what circumstances. Machine and deep learning can be used for properly identifying the once again fuzzy lines between using data analysis for benevolent service optimization or arbitrary behavior mapping which can be later sold to the highest bidder.

This chapter presented the working principles together with the strength and weakness of several machine and deep learning approaches, focusing on the identification and mitigation of modern IoT security threats and vulnerabilities. Therefore, it is expected to serve as a useful manual encouraging researchers to advance the security of IoT systems either by addressing device or end-to-end security challenges.

Acknowledgements We acknowledge support of this work by the project "I3T - Innovative Application of Industrial Internet of Things (IIoT) in Smart Environments" (MIS 5002434) which is implemented under the "Action for the Strategic Development on the Research and Technological Sector", funded by the Operational Programme "Competitiveness, Entrepreneurship and Innovation" (NSRF 2014-2020) and co-financed by Greece and the European Union (European Regional Development Fund).

The views and opinions expressed are those of the authors and do not necessary reflect the official position of Citrix Systems Inc.

References

M. Aazam, I. Khan, A. A. Alsaffar, and E. Huh. Cloud of things: Integrating internet of things and cloud computing and the issues involved. In *Proceedings of 2014 11th International Bhurban Conference on Applied Sciences Technology (IBCAST) Islamabad, Pakistan, 14th - 18th January, 2014*, pages 414–419, Jan 2014. doi: 10.1109/IBCAST.2014.6778179.

A. Abeshu and N. Chilamkurti. Deep learning: The frontier for distributed attack detection in fog-to-things computing. *IEEE Communications Magazine*, 56(2): 169–175, Feb 2018. ISSN 0163-6804. doi: 10.1109/MCOM.2018.1700332.

P. Adolphs, J. Cabot, and M. Wimmer. *Structure of the Administration Shell: Continuation of the Development of the Reference Model for the Industrie 4.0 Component*. Platform Industrie 4.0, 2016. URL https://www.plattform-i40.de/I40/Redaktion/EN/Downloads/Publikation/structure-of-the-administration-shell.pdf.

S. Alharbi, P. Rodriguez, R. Maharaja, P. Iyer, N. Subaschandrabose, and Z. Ye. Secure the internet of things with challenge response authentication in fog computing. In *2017 IEEE 36th International Performance Computing and Communications Conference (IPCCC)*, pages 1–2, Dec 2017. doi: 10.1109/PCCC.2017.8280489.

M. E. Aminanto, R. Choi, H. C. Tanuwidjaja, P. D. Yoo, and K. Kim. Deep abstraction and weighted feature selection for wi-fi impersonation detection. *IEEE Transactions on Information Forensics and Security*, 13(3):621–636, March 2018. ISSN 1556-6013. doi: 10.1109/TIFS.2017.2762828.

S. Athey and G. Imbens. Machine learning methods for estimating heterogeneous causal effects. 2015.

J. Attenberg, P. Ipeirotis, and F. Provost. Beat the machine: Challenging humans to find a predictive model's "unknown unknowns". *J. Data and Information Quality*, 6(1):1:1–1:17, Mar. 2015. ISSN 1936-1955. doi: 10.1145/2700832. URL http://doi.acm.org/10.1145/2700832.

L. Atzori, A. Iera, and G. Morabito. The internet of things: A survey. *Comput. Netw.*, 54(15):2787–2805, Oct. 2010. ISSN 1389-1286. doi: 10.1016/j.comnet.2010.05.010. URL http://dx.doi.org/10.1016/j.comnet.2010.05.010.

R. Caruana, Y. Lou, J. Gehrke, P. Koch, M. Sturm, and N. Elhadad. Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '15, pages 1721–1730, New York, NY, USA, 2015. ACM. ISBN 978-1-4503-3664-2. doi: 10.1145/2783258.2788613. URL http://doi.acm.org/10.1145/2783258.2788613.

K. Chen, S. Zhang, Z. Li, Y. Zhang, Q. Deng, S. Ray, and Y. Jin. Internet-of-things security and vulnerabilities: Taxonomy, challenges, and practice. *Journal of Hardware and Systems Security*, 2(2):97–110, 6 2018. ISSN 2509-3428. doi: 10.1007/s41635-017-0029-7. URL https://doi.org/10.1007/s41635-017-0029-7.

A. Conn. *The AI wars: The battle of the human minds to keep artificial intelligence safe*. Industrial Internet Consortium, 2015. URL http://futureoflife.org/2015/12/17/the-ai-wars-the-battle-of-the-human-minds-to-keep-artificial-intelligence-safe.

R. Doshi, N. Apthorpe, and N. Feamster. Machine learning ddos detection for consumer internet of things devices. In *2018 IEEE Security and Privacy Workshops (SPW)*, pages 29–35, May 2018. doi: 10.1109/SPW.2018.00013.

W. Du and Z. Zhan. Building decision tree classifier on private data. In *Proceedings of the IEEE International Conference on Privacy, Security and Data Mining - Volume 14*, CRPIT '14, pages 1–8, Darlinghurst, Australia, Australia, 2002. Australian Computer Society, Inc. ISBN 0-909-92592-5. URL http://dl.acm.org/citation.cfm?id=850782.850784.

ENISA Report. Baseline Security Recommendations for IoT, 2017. URL https://www.enisa.europa.eu/publications/baseline-security-recommendations-for-iot.

ENISA Report. Good Practices for Security of Internet of Things, 2018a. URL https://www.enisa.europa.eu/publications/good-practices-for-security-of-iot.

ENISA Report. Hardware Threat Landscape and Good Practice Guide, 2018b. URL https://www.enisa.europa.eu/publications/hardware-threat-landscape.

ENISA Report. Ad-hoc and sensor networking for M2M Communications, 2018c. URL https://www.enisa.europa.eu/publications/m2m-communications-threat-landscape.

D. Evans. *The Internet of Things—How the Next Evolution of the Internet Is Changing Everything*. White Paper. CISCO, 2011.

U. Fiore, F. Palmieri, A. Castiglione, and A. De Santis. Network anomaly detection with the restricted boltzmann machine. *Neurocomput.*, 122:13–23, Dec. 2013. ISSN 0925-2312. doi: 10.1016/j.neucom.2012.11.050. URL http://dx.doi.org/10.1016/j.neucom.2012.11.050.

A. A. Freitas. Comprehensible classification models: A position paper. *SIGKDD Explor. Newsl.*, 15(1):1–10, Mar. 2014. ISSN 1931-0145. doi: 10.1145/2594473.2594475. URL http://doi.acm.org/10.1145/2594473.2594475.

P. Gangsar and R. Tiwari. Comparative investigation of vibration and current monitoring for prediction of mechanical and electrical faults in induction motor based on multiclass-support vector machine algorithms. *Mechanical Systems and Signal Processing*, 94:464 – 481, 2017. ISSN 0888-3270. doi: https://doi.org/10.1016/j.ymssp.2017.03.016. URL http://www.sciencedirect.com/science/article/pii/S088832701730136X.

I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. The MIT Press, 2016. ISBN 0262035618, 9780262035613.

R. E. Hiromoto, M. Haney, and A. Vakanski. A secure architecture for iot with supply chain risk management. In *2017 9th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS)*, volume 1, pages 431–435, Sep. 2017. doi: 10.1109/IDAACS.2017.8095118.

A. Humayed, J. Lin, F. Li, and B. Luo. Cyber-physical systems security—a survey. *IEEE Internet of Things Journal*, 4(6):1802–1831, Dec 2017. ISSN 2327-4662. doi: 10.1109/JIOT.2017.2703172.

IIC. Industrial Internet of Things Volume G4: Security Framework, 2016. URL https://www.iiconsortium.org/IISF.htm.

ITY-T. Overview of Internet of Things, 2012.

N. Jazdi. Cyber physical systems in the context of industry 4.0. In *2014 IEEE International Conference on Automation, Quality and Testing, Robotics*, pages 1–4, May 2014. doi: 10.1109/AQTR.2014.6857843.

M. I. Jordan and T. M. Mitchell. Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245):255–260, 2015. ISSN 0036-8075. doi: 10.1126/science.aaa8415. URL http://science.sciencemag.org/content/349/6245/255.

S. Kapoor, A. Mojsilovic, J. N. Strattner, and K. R. Varshney. From open data ecosystems to systems of innovation : A journey to realize the promise of open data. 2015.

R. Khan, S. U. Khan, R. Zaheer, and S. Khan. Future internet: The internet of things architecture, possible applications and key challenges. In *2012 10th International*

Conference on Frontiers of Information Technology, pages 257–260, Dec 2012. doi: 10.1109/FIT.2012.53.

G. Kim, S. Lee, and S. Kim. A novel hybrid intrusion detection method integrating anomaly detection with misuse detection. *Expert Syst. Appl.*, 41(4):1690–1700, Mar. 2014. ISSN 0957-4174. doi: 10.1016/j.eswa.2013.08.066. URL http://dx.doi.org/10.1016/j.eswa.2013.08.066.

L. Lerman, G. Bontempi, and O. Markowitch. A machine learning approach against a masked aes. *Journal of Cryptographic Engineering*, 5(2):123–139, Jun 2015. ISSN 2190-8516. doi: 10.1007/s13389-014-0089-3.

S.-W. Lin, M. Crawford, B. Miller, J. Durand, and G. Bleakley. *The Industrial Internet of Things Volume G1: Reference Architecture*. Industrial Internet Consortium, 2017a. URL https://www.iiconsortium.org/IIC_PUB_G1_V1.80_2017-01-31.pdf.

S.-W. Lin, B. Murphy, E. Clauer, U. Loewen, and G. Bleakley. *Architecture Alignment and Interoperability*. Industrial Internet Consortium and Plattform Industrie 4.0 Joint Whitepaper, 2017b. URL http://www.iiconsortium.org/pdf/JTG2_Whitepaper_final_20171205.pdf.

H. Maghrebi, T. Portigliatti, and E. Prouff. Breaking cryptographic implementations using deep learning techniques. In *IACR Cryptology ePrint Archive*, 2016.

L. Mainetti, L. Patrono, and A. Vilei. Evolution of wireless sensor networks towards the internet of things: A survey. In *SoftCOM 2011, 19th International Conference on Software, Telecommunications and Computer Networks*, pages 1–6, Sep. 2011.

N. Maller and S. O. Hansson. Principles of engineering safety: Risk and uncertainty reduction. *Reliability Engineering & System Safety*, 93(6):798 – 805, 2008. ISSN 0951-8320. doi: https://doi.org/10.1016/j.ress.2007.03.031. URL http://www.sciencedirect.com/science/article/pii/S0951832007001251.

I. Mashal, O. Alsaryrah, T.-Y. Chung, C.-Z. Yang, W.-H. Kuo, and D. P. Agrawal. Choices for interaction with things on internet and underlying issues. *Ad Hoc Networks*, 28:68 – 90, 2015. ISSN 1570-8705. doi: https://doi.org/10.1016/j.adhoc.2014.12.006. URL http://www.sciencedirect.com/science/article/pii/S1570870514003138.

N. McLaughlin, J. Martinez del Rincon, B. Kang, S. Yerima, P. Miller, S. Sezer, Y. Safaei, E. Trickel, Z. Zhao, A. Doupé, and G. Joon Ahn. Deep android malware detection. In *Proceedings of the Seventh ACM on Conference on Data and Application Security and Privacy*, CODASPY '17, pages 301–308, New York, NY, USA, 2017. ACM. ISBN 978-1-4503-4523-1. doi: 10.1145/3029806.3029823. URL http://doi.acm.org/10.1145/3029806.3029823.

Y. Meidan, M. Bohadana, A. Shabtai, M. Ochoa, N. O. Tippenhauer, J. D. Guarnizo, and Y. Elovici. Detection of unauthorized iot devices using machine learning techniques. *CoRR*, abs/1709.04647, 2017. URL http://arxiv.org/abs/1709.04647.

D. M. Mena, I. Papapanagiotou, and B. Yang. Internet of things: Survey on security. *Information Security Journal: A Global Perspective*, 27(3):162–182, 2018. doi: 10.1080/19393555.2018.1458258. URL https://doi.org/10.1080/19393555.2018.1458258.

D. Miorandi, S. Sicari, F. D. Pellegrini, and I. Chlamtac. Internet of things: Vision, applications and research challenges. *Ad Hoc Networks*, 10(7):1497 – 1516, 2012. ISSN 1570-8705. doi: https://doi.org/10.1016/j.adhoc.2012.02.016. URL http://www.sciencedirect.com/science/article/pii/S1570870512000674.

A. Y. Ng and M. I. Jordan. On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes. In *Proceedings of the 14th International Conference on Neural Information Processing Systems: Natural and Synthetic*, NIPS'01, pages 841–848, Cambridge, MA, USA, 2001. MIT Press. URL http://dl.acm.org/citation.cfm?id=2980539.2980648.

OWASP. The free and open software security community, 2018. URL http://www.owasp.org/index.php/OWASP_Internet_of_Things_Project.

M. Ozay, I. Esnaola, F. T. Yarman Vural, S. R. Kulkarni, and H. V. Poor. Machine learning methods for attack detection in the smart grid. *IEEE Transactions on Neural Networks and Learning Systems*, 27(8):1773–1786, Aug 2016. ISSN 2162-237X. doi: 10.1109/TNNLS.2015.2404803.

S. Pan, T. Morris, and U. Adhikari. Developing a hybrid intrusion detection system using data mining for power systems. *IEEE Transactions on Smart Grid*, 6(6): 3104–3113, Nov 2015. ISSN 1949-3053. doi: 10.1109/TSG.2015.2409775.

R. Pascanu, T. Mikolov, and Y. Bengio. On the difficulty of training recurrent neural networks. In *Proceedings of the 30th International Conference on International Conference on Machine Learning - Volume 28*, ICML'13, pages 1310–1318. JMLR.org, 2013. URL http://dl.acm.org/citation.cfm?id=3042817.3043083.

J. R. Quinlan. Induction of decision trees. *Mach. Learn.*, 1(1):81–106, Mar. 1986. ISSN 0885-6125. doi: 10.1023/A:1022643204877. URL http://dx.doi.org/10.1023/A:1022643204877.

C. Rudin. Algorithms for interpretable machine learning. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '14, pages 1519–1519, New York, NY, USA, 2014. ACM. ISBN 978-1-4503-2956-9. doi: 10.1145/2623330.2630823. URL http://doi.acm.org/10.1145/2623330.2630823.

R. Schmidt, M. Möhring, R.-C. Härting, C. Reichstein, P. Neumaier, and P. Jozinović. Industry 4.0 - potentials for creating smart products: Empirical research results. In W. Abramowicz, editor, *Business Information Systems*, pages 16–27, Cham, 2015. Springer International Publishing. ISBN 978-3-319-19027-3.

P. Sethi and S. R. Sarangi. Internet of things: Architectures, protocols, and applications. *J. Electrical and Computer Engineering*, 2017:9324035:1–9324035:25, 2017.

M. Sharp, R. Ak, and T. Hedberg. A survey of the advancing use and development of machine learning in smart manufacturing. *Journal of Manufacturing Systems*, 48:170 – 179, 2018. ISSN 0278-6125. doi: https://doi.org/10.1016/j.jmsy.2018.02.004. URL http://www.sciencedirect.com/science/article/pii/S0278612518300153. Special Issue on Smart Manufacturing.

E. Shaw. *Improving Service and Communication with Open Data*. Data Smart City solutions, 2015. URL https://datasmart.ash.harvard.edu/news/article/improving-service-and-communication-with-open-data-702.

M.-Y. Su. Real-time anomaly detection systems for denial-of-service attacks by weighted k-nearest-neighbor classifiers. *Expert Syst. Appl.*, 38(4):3492–3498, Apr. 2011. ISSN 0957-4174. doi: 10.1016/j.eswa.2010.08.137. URL http://dx.doi.org/10.1016/j.eswa.2010.08.137.

A. R. Syarif and W. Gata. Intrusion detection system using hybrid binary pso and k-nearest neighborhood algorithm. In *2017 11th International Conference on Information Communication Technology and System (ICTS)*, pages 181–186, Oct 2017. doi: 10.1109/ICTS.2017.8265667.

P. Torres, C. Catania, S. Garcia, and C. G. Garino. An analysis of recurrent neural networks for botnet detection behavior. In *2016 IEEE Biennial Congress of Argentina (ARGENCON)*, pages 1–6, June 2016. doi: 10.1109/ARGENCON.2016.7585247.

K. R. Varshney, R. J. Prenger, T. L. Marlatt, B. Y. Chen, and W. G. Hanley. Practical ensemble classification error bounds for different operating points. *IEEE Transactions on Knowledge and Data Engineering*, 25(11):2590–2601, Nov 2013. ISSN 1041-4347. doi: 10.1109/TKDE.2012.219.

WASC. Threat Classification v2.0, 2012. URL http://projects.webappsec.org/w/page/13246978/Threat%20Classification.

M. Welling. Are ml and statistics complementary. IMS-ISBA Meeting on Data Science in the Next 50 Years, 12 2015.

S. Weyer, M. Schmitt, M. Ohmer, and D. Gorecky. Towards industry 4.0 - standardization as the crucial challenge for highly modular, multi-vendor production systems. *IFAC-PapersOnLine*, 48(3):579 – 584, 2015. ISSN 2405-8963. doi: https://doi.org/10.1016/j.ifacol.2015.06.143. URL http://www.sciencedirect.com/science/article/pii/S2405896315003821. 15th IFAC Symposium onInformation Control Problems inManufacturing.

M. Wu, T.-J. Lu, F.-Y. Ling, J. Sun, and H.-Y. Du. Research on the architecture of internet of things. In *2010 3rd International Conference on Advanced Computer Theory and Engineering(ICACTE)*, volume 5, pages V5–484–V5–487, Aug 2010. doi: 10.1109/ICACTE.2010.5579493.

A. S. Xanthopoulos, A. Kiatipis, D. E. Koulouriotis, and S. Stieger. Reinforcement learning-based and parametric production-maintenance control policies for a deteriorating manufacturing system. *IEEE Access*, 6:576–588, 2018. ISSN 2169-3536. doi: 10.1109/ACCESS.2017.2771827.

M. Xie, M. Huang, Y. Bai, and Z. Hu. The anonymization protection algorithm based on fuzzy clustering for the ego of data in the internet of things. *Journal of Electrical and Computer Engineering, Hindawi*, 1(1):1–10, 2 2017. Article ID 2970673.

L. D. Xu, W. He, and S. Li. Internet of things in industries: A survey. *IEEE Transactions on Industrial Informatics*, 10(4):2233–2243, Nov 2014. ISSN 1551-3203. doi: 10.1109/TII.2014.2300753.

K. Yang, J. Ren, Y. Zhu, and W. Zhang. Active learning for wireless iot intrusion detection. *IEEE Wireless Communications*, 25(6):19–25, December 2018. ISSN 1536-1284. doi: 10.1109/MWC.2017.1800079.

Y. Ye, T. Li, D. Adjeroh, and S. S. Iyengar. A survey on malware detection using data mining techniques. *ACM Comput. Surv.*, 50(3):41:1–41:40, June 2017. ISSN 0360-0300. doi: 10.1145/3073559. URL http://doi.acm.org/10.1145/3073559.

H. Zenati, C. S. Foo, B. Lecouat, G. Manek, and V. R. Chandrasekhar. Efficient GAN-Based Anomaly Detection. *CoRR*, abs/1802.06222, 2018. URL http://arxiv.org/abs/1802.06222.