DOI: 10.24427/978-83-67185-29-5_9

9. BIG DATA AND IOT SECURITY ISSUES

9.1. Big Data Technology

Big Data or big data is structured or unstructured data arrays of large volume. They are processed using special automated tools to be used for statistics, analysis, forecasts, and decision making. The term "big data" was coined by Nature editor Clifford Lynch in a 2008 special issue. He talked about the explosive growth of information in the world. Lynch referred to big data any arrays of heterogeneous data more than 150 GB per day, but there is still no single criterion (Gantz & Reinsel, 2020).

Until 2011, big data analysis was done only within the framework of scientific and statistical research. But by the beginning of 2012, data volumes had grown to enormous proportions, and there was a need for their systematization and practical application.

Since 2014, the world's leading universities have paid attention to Big Data, where they teach applied engineering and IT specialties. Then IT corporations such as Microsoft, IBM, Oracle, EMC, and then Google, Apple, Facebook, and Amazon joined the collection and analysis. Today, big data is used by large companies in all industries, as well as government agencies.

9.1.1. Big Data Basics

To designate an array of information with the prefix "big", it must have the following features:

- **Volume** – data is measured by physical quantity and the space occupied on a digital medium. "Big" refers to arrays over 150 GB per day.
- **Velocity** – information is regularly updated and real-time processing requires intelligent big data technologies.
- **Variety** – information in arrays can have heterogeneous formats, be structured partially, completely, and accumulate unsystematically.

For example, social networks use big data in the form of texts, videos, audio, financial transactions, pictures, and more. In modern systems, two additional factors are considered:

- **Variability** – data streams can have peaks and valleys, seasonality, periodicity. Bursts of unstructured information are difficult to manage and require powerful processing technologies.
- **Value** – information can have different complexity for perception and processing, which makes it difficult for intelligent systems to work. For example, an array of messages from social networks is one level of data, and transactional operations are another.

The task of machines is to determine the degree of importance of incoming information to quickly structure. The principle of operation of big data technology is based on the maximum informing the user about any object or phenomenon. The purpose of this familiarization with the data is to help you weigh the pros and cons to make the right decision. In intelligent machines, a model of the future is built based on an array of information, and then various options are simulated, and the results are monitored.

Modern analytics agencies run millions of these simulations when they test an idea, a hypothesis, or solve a problem. The process is automated.

Big data sources include:

- internet of things and devices connected to it;
- social networks, blogs, and media;
- company data: transactions, orders for goods and services, taxi and car sharing trips, customer profiles;
- device readings: meteorological stations, air and water composition meters, satellite data;
- statistics of cities and states: data on movements, births, and deaths; medical data: tests, diseases, diagnostic images.

The principles of working with data arrays include three main factors (Perry, 2018):

- **System extensibility**. It is usually understood as the horizontal scalability of storage media. That is, the volumes of incoming data have grown – the capacity and number of servers for their storage have increased.
- **Resilience to failure**. It is possible to increase the number of digital media, intelligent machines in proportion to the amount of data indefinitely. But this does not mean that some of the machines will not fail or become obsolete. Therefore, one of the factors for stable work with big data is the fault tolerance of servers.
- **Localization**. Separate arrays of information are stored and processed within one dedicated server to save time, resources, and data transfer costs.

Modern computing systems provide instant access to big data arrays. For their storage, special data centers with the most powerful servers are used.

In addition to traditional, physical servers, they use cloud storage, data lakes (data lakes – storage of a large amount of unstructured data from a single source) and Hadoop, a framework consisting of a set of utilities for developing and executing

distributed computing programs. To work with Big Data, advanced methods of integration and management are used, as well as data preparation for analytics.

9.1.2. Big Data and Other Technologies

Blockchain is a decentralized transaction system, where every transaction is verified by every element of the network. Such a system guarantees the immutability and impossibility of data manipulation. Cryptocurrencies and other blockchain technologies are becoming more and more popular. In Japan alone, nearly 50 banks have partnered with Ripple, the open-source blockchain network and third-largest crypto market capitalization in the world. For banks, cooperation will provide instant risk-free transactions at a low cost. Interest in such operations is shown by financial structures in other countries, which means the further development of new technologies in the banking sector.

The popularity of the technology portends an exponential growth in the volume of transactional data recorded in registers. By 2030, the information contained in the blockchain ledger will account for up to 20% of the global Big Data market and generate up to $100 billion in annual revenue (Rayes & Salam, 2019). Storing these "data lakes" with traditional cloud storage providers (AWS or Azure) will cost a fortune. Decentralized data storage providers have entered the market in a timely manner, offering cost savings of up to 90%. Their work facilitates the implementation of the blockchain around the world and guarantees the development of the sphere.

The use of blockchain opens a new level of Big Data analytics. Such information is structured, complete, and secure, as it cannot be faked due to the network architecture. By analyzing it, the algorithms will be able to check every transaction in real time, which will practically destroy fraud in the digital sphere. Instead of analyzing fraud records that have already taken place, banks can instantly identify risky or fraudulent activities and prevent them. Blockchain technology is applicable not only to the financial sector. Immutable records, audit trails and confidence in the origin of data – all this applies to any business area. Already, companies are implementing blockchain in the food trade, and on the other hand, they are studying the prospects of technology in space exploration. Future Big Data and blockchain solutions are expected to radically change the way business is done.

Today, many industries are implementing machine learning to automate business processes and modernize the economic sphere. The concept provides for the training and management of artificial intelligence (AI) using special algorithms. They teach the system based on open data or experience.

Over time, such an application can predict the development of events without explicit human programming and hours spent writing code. For example, using machine learning, you can create an algorithm for the technical analysis of stocks and their estimated prices.

Using regression and predictive analysis, statistical modeling and action analysis, experts create programs that calculate the time of profitable purchases in the stock market. They analyze open data from exchanges and offer the most likely course of events.

When working with Big Data, machine learning performs a similar function: special programs analyze impressive amounts of information without human intervention. All that is required of the operator is to "teach" the algorithm to select useful data that the company needs to optimize processes. This allows analysts to generate reports in a few mouse clicks, freeing up their time and resources for more productive tasks: processing results and finding the most effective strategies.

In a fast-paced world where customer expectations are ever higher and human resources ever more valuable, machine learning and data science play a critical role in a company's growth. Digital technologization of the workflow is vital to maintaining a leading position in a competitive environment.

9.1.3. Big Data Analytics

Thanks to high-performance technologies such as grid computing or in-memory analytics, companies can use any amount of big data for analysis. Sometimes Big Data is first structured, selecting only those that are needed for analysis. Increasingly, big data is used for tasks within advanced analytics, including artificial intelligence.

There are four main methods of Big Data analysis (Perry, 2018):

1. **Descriptive analytics** is the most common. It answers the question "What happened?", analyzes real-time data and historical data. The main goal is to find out the causes and patterns of success or failure in a particular area to use this data for the most effective models. For descriptive analytics, basic mathematical functions are used. A typical example is sociological research or web statistics data that a company receives through Google Analytics.
2. **Predictive analytics** – helps to predict the most likely development of events based on the available data. To do this, use ready-made templates based on any objects or phenomena with a similar set of characteristics. With the help of predictive (or predictive, predictive) analytics, you can, for example, calculate a collapse or price change in the stock market. Or assess the potential borrower's ability to repay a loan.
3. **Prescriptive analytics** is the next level up from predictive. With the help of Big Data and modern technologies, it is possible to identify problem points in a business or any other activity and calculate under what scenario they can be avoided in the future.
4. **Diagnostic analytics** – uses data to analyze the causes of what happened. This helps to detect anomalies and random connections between events and activities.

Data is processed and analyzed using various tools and technologies:

- **Special software**: NoSQL, MapReduce, Hadoop, R;
- **Data mining** – extracting previously unknown data from arrays using a large set of techniques;
- **AI and neural networks** – for building models based on Big Data, including text and image recognition. For example, the lottery operator Stoloto has made big data the basis of its strategy within the Data-driven Organization. Using Big Data and artificial intelligence, the company analyzes customer experience and offers personalized products and services;
- **Analytical data visualization** – Animated models or graphs based on big data.

Developers adhere to two criteria for collecting information: anonymization of data which makes personal information of users inaccessible to some extent and aggregation of data which allows us to operate only with average indicators. To process large amounts of data online, supercomputers are used: their power and computing capabilities are many times greater than conventional ones.

Big data techniques are widely used in many areas like, for example:

- **Public administration**. The study and analysis of big data helps governments make decisions in areas such as health, employment, economic regulation, crime and security, emergency response; Industry. The introduction of Big Data tools helps to increase the transparency of industrial processes and introduce "predictive production", which makes it possible to predict the demand more accurately for products and, accordingly, plan the expenditure of resources.
- **Medicine**. The huge amount of data collected by medical institutions and various electronic devices (fitness bracelets, etc.) opens up fundamentally new opportunities for the healthcare industry. Big data helps to find new medicines, make more accurate diagnoses, select effective treatment, fight pandemics.
- **Retail**. The development of network and electronic commerce is impossible to imagine without solutions based on Big Data – this is how stores personalize assortment and delivery.
- **Internet of things**. Big Data and the Internet of Things are inextricably linked. Industrial and household appliances connected to the Internet of things collect a huge amount of data, based on the analysis of which the operation of these devices is subsequently regulated.
- **Real estate market**. Developers use Big Data technologies to collect and analyze the entire array of information, and then give the user the most interesting options for him. Already now, a future buyer can see the house he likes without a seller;
- **Sport**. With the help of big data, football clubs select the most promising players and develop an effective strategy for each opponent.
- **Agriculture**. An IoT solution from the field of so-called precision farming is when special weather stations that stand in the fields collect data (temperature, humidity) using sensors and send them to the IoT platform using transmitting radio-GSM modules. On it, using big data algorithms, the information collected from

the sensors is processed and a high-precision hourly weather forecast is built. The client sees it in the interface on a computer, tablet or smartphone and can quickly make decisions.

The world leaders in the collection and analysis of big data are the United States and China. So, in the United States, even under Barack Obama, the government launched six federal programs for the development of big data for a total of $200 million. Large corporations are considered the main consumers of Big Data, but their data collection activities are limited in some states – for example, in California.

China has more than 200 laws and regulations regarding the protection of personal information (Rayes & Salam, 2019). Since 2019, all popular smartphone apps have been checked and blocked if they collect user data in violation of the law. As a result, the state collects data through local services, and many of them are inaccessible from the outside.

Since 2018, the European Union has adopted the GDPR – the General Data Protection Regulation. It regulates everything related to the collection, storage, and use of online user data. When the law went into effect a year ago, it was considered the world's toughest system to protect people's online privacy.

There main problems of big data are:

- Big data is heterogeneous and therefore difficult to process for statistical inference. The more parameters required for forecasting; the more errors accumulate in the analysis;
- Working with large amounts of data online requires huge computing power. Such resources are very expensive, and so far, only available to large corporations;
- The storage and processing of Big Data is associated with increased vulnerability to cyber-attacks and all kinds of leaks. A prime example is the Facebook profile scandals;
- The collection of big data is often associated with a privacy issue: not everyone wants their every action to be tracked and transferred to third parties;
- Big data is used not only by corporations, but also by politicians: for example, to influence elections.

9.1.4. Big Data Ethical Considerations

Any technology has both positive and negative features to some extent. On the one hand, research Massachusetts Institute of Technology (MIT) showed that companies using methods of analysis of large data, were able to improve profitability by an average of 5–6% (Betts, 2016).

On the other hand, this young field of computer science gives rise to several ethical problems. These problems need to be identified and addressed, as they may ultimately override the benefits of intelligent systems. Collection, processing, and analysis of client data has now become the way making money, and in some cases a factor that can

change an entire industry. Most of this data, directly or indirectly, can be considered personal data and therefore subject to protection. However, this is not enough since data protection laws usually do not cover ethical and moral aspects.

Illustrative is the situation with American company Target. Its specialists have developed an algorithm for offering personalized services by analyzing customer data, including their behavior on the Internet. In 2012 The father of a high school student complained that Target sent his daughter coupons for goods for pregnant women. The company admitted the mistake, but it soon became clear that the girl was indeed pregnant, and the big data algorithm only considered the changes in the behavior of the girl, which was like behavior of pregnant women (Betts, 2016). In this situation, no law was violated, but the public was outraged by such an invasion of privacy.

Also, an example of the use of personal data that is contrary to moral standards can be the work of special services, in particular, FBI. In 2016 it was found that FBI database includes 412 million photos people, among whom many foreign citizens and persons who have never broken the law (Perry, 2018). At the same time, the FBI intentionally hides information about how and to what extent new technologies are used (contrary to the requirements of protecting confidential data). It should be noted that the use of the facial recognition system was effective and helped in the capture of a number of criminals. However, there were some false positives: due to the error of methods analysis of big data, the system called law-abiding citizens criminals.

Based on the mentioned above, some ethical problems and possible ways to solve them can be identified (McEwen & Hakim, 2014):

- **Data collection and use**. People's concern is what will happen to the data after how they were obtained, and what will be the limits of their application. To earn the trust of customers, companies need to indicate all ways of using data in the user agreement, avoiding ambiguous wording.
- **Transfer of data to third parties**. Confidential data must be provided to other companies or individuals without any indication of the identity of its owner. It is also necessary to notify the client about the transfer of his data.
- **Observation**. The use of video analytics or behavioral analytics makes users feel that they are under surveillance. Technology that they feel is intrusive and therefore disadvantages them independence. To solve this problem, the person should be warned about the application of the data analysis algorithm. For example, in places where video analytics cameras are installed. Developers also need to thoroughly test their system to avoid situations like the incident with Target.
- **Prejudice**. Big data should not perpetuate stereotypes such as racism or sexism. For example, the FBI's facial recognition system was the most frequently mistaken when processing photos of African Americans. To avoid such situations, developers need to improve the quality of forecasts, or think about the appropriateness of using the algorithm.

By addressing ethical issues at an earlier stage, steps must be taken to either alleviate the problems or eliminate them. Organizations must implement moral and ethical

codes that cover the full life data cycle, including acquisition, preparation, processing, aggregation, sharing, storage, archiving and destruction.

The ethical aspect of big data is dual in nature: on the one hand, technologies (including big data) help in many areas: medicine, finance, education, etc. After a few years, the patient will be invited to see a doctor as soon as symptoms appear. But people will have to live in a world where there will be no place for privacy and where their privacy will be limited.

In a global sense, the problem is how people will use the data they receive. Therefore, the critical importance of modern life starts playing mental and moral level of people. If the society and each its member lacks something in ethical background the negative consequences could spread in all over the world.

9.2. Big Data in Modern Buildings and Smart Cities

According to the United Nations, the world population will grow to 9.7 billion by 2050. Of these, 6.3 billion people will live in cities. For such a high pace of urbanization has two reasons. Firstly, people move from villages or other small settlements to large cities for better living conditions and living standards in general, for better paid jobs, etc. Secondly, people do not move within one country, but on a global scale: migrants from remote poor areas of developing or underdeveloped countries move to large countries, where again the standard of living is higher, better conditions, easier to find work or favorable conditions for visiting residents.

As cities grow, they become more and more difficult to manage.

Many workers are needed to serve each inhabitant, at the same time qualitatively and quickly, so that the standard of living does not fall, but only increases over the years. This need has become one of the reasons for the start of automation of various processes. Starting from fare payment and self-service checkouts in stores, and ending with road map management, management decision support, electronic housing, and communal services – these are examples of the implementation of smart city and smart building projects.

9.2.1. Introduction to Smart Cities

A smart city is a safe, sustainable (green) and efficient urban center of the future with advanced infrastructure of sensors, electronics and networks that drives sustainable economic growth and high quality of life. Many scientists who consider this concept from a scientific point of view emphasize the importance of human capital, modern infrastructure, and information technology.

The British Standard Institution (BSI) describes a smart city as: a combination of different systems (human, physical, information, and others) in the most efficient

way to get as a result sustainable, highly intelligent, convenient, and comfortable future for the citizens of the city (McEwen & Hakim, 2014).

Information technology allows city government to interact directly with communities and city infrastructure, and monitor what is happening in the city, how the city is developing, and what ways can improve the quality of life. Using sensors integrated in real time, the accumulated data from urban residents and devices is processed and analyzed. The collected information is the key to solving problems of inefficiency. ICT is used to improve the quality, productivity, and interactivity of city services, reduce costs and resource consumption, improve communication between city dwellers and the state.

From an information technology point of view, a "smart city" is defined as a way to create more intelligent and efficient infrastructure elements: city administration, education and health systems, public order, transport infrastructure, and so on. From this point of view, at the heart of any "smart city" should be information. Information should be accumulated from various sensors installed on buildings and other smart city facilities. Data exchange should keep all the internal processes of the city connected, creating a single ecosystem. The data obtained from sensors must be used in such a way that the living conditions of citizens are stable and comfortable, as well as more economical. For the sustainable development of the city, a model of smart operational management is used.

Because "smart cities" is an innovative and popular topic in the modern world, there are many interpretations of what parts a "smart city" should consist of for its high-quality functioning, as well as how to describe it.

Some researchers studying the infrastructure of smart cities describe it as having the following distinguishing features or features (Perry, 2018):

- increasing attractiveness for investment and just life; creation of new jobs (this paragraph contributes to the implementation of the first paragraph, which indicates that the "smart city" is a complex and constantly improving system);
- an effective social and cultural environment (so that new residents do not feel the same, but retain their individuality); careful attitude to resources (both renewable, such as electricity, and non-renewable, such as water);
- optimization of traffic flows (since, due to the growing population, the pace of life of residents will also increase, therefore, the traffic flow will increase);
- so-called smart buildings, automated commercial services and "engineering infrastructure" (this term refers to the automation of daily events and actions of city residents: ordering a taxi, paying bills, buying groceries, paying transport fares, and so on).

This description of the infrastructure of smart cities clearly explains the essence: make the most of information technology (and constantly update technology, as in the IT field there is a very fast change of devices and methods to more modern ones) in order to provide new people who are increasingly arriving in large cities with a decent standard of living, a workplace and comfortable conditions in material

terms for existence (in particular, the ability to automatically perform daily activities). What is included in the concept of "smart city" and what are its components is also an open question. Many scientists divide in their own way, and sometimes each scientist calls the same concept in his own way. After analyzing several such divisions, the following list of smart city components was formulated (Rayes & Salam, 2019):

1. **Smart economy**. The following concepts are used to describe this term: productivity, a clear definition of how private and public spheres are interconnected (and how divided), entrepreneurship development as a way improving the economic condition of the city, observing, and maintaining economic trends in society and the close connection of life with the economy.
2. **Smart people**, which are distinguished by: flexibility and mobility as the basis of a lifestyle, they are highly qualified professionals in in their chosen field, they are ready to learn all their lives, craving for new knowledge and skills, as well as the desire for self-improvement both in work and in personal life. The supremacy of reason and logic over emotions when it comes to making some important and responsible decisions.
3. **Smart living**. High level of social living conditions, health care, high level of medical services, continuous training.
4. **Smart Governance**. Systems that are essential in case of an emergency (fire, ambulance, police, etc.) should be available 24/7 without restrictions to all members of society, support for adoption solutions using digital infrastructure and the latest technologies, the availability of social public services for the provision of services, the control of the level of urbanization (to prevent overpopulation of cities with insufficient resources).
5. **Smart Mobility**. For residents, there should be absolutely no barriers to movement both within their city, country, and between countries. There should be enough resources so that any resident can maintain a high level of automation in their lives. It is important to monitor the safety of residents in various areas.
6. **Smart Environment**. Resource management: special attention to non-renewable resources, but also the control of renewable resources, such as control and reduction of electricity consumption, environmental protection and assistance in its restoration, regulation of air pollution and minimizing them.

The main goal of developing and implementing a smart city system is the need for better city management. To make the implementation of the system of "smart cities" more efficient, scientists use international standards to improve the quality of management.

9.2.2. Life Cycle of Smart Cities

To be able to comprehensively automate an entire city, you need to understand how the system develops, what stages of the life cycle it goes through. To The analyzed

stages of the life cycle of a "smart city" (which are like the stages of the life cycle of a simple city) were assigned the following stages:

1. development,
2. expansion,
3. stagnation,
4. decline.

1. **Stage of development of the city.** The first stage of the life cycle of a city is characterized by a situation where the area of housing, the amount of goods and resources is growing rapidly. At the same time, the number of residents and jobs "does not keep pace" with the development of resources and there is an excess of the latter. Usually at this stage, the emergence of city-forming enterprises occurs, the influx of investors into the city increases, the city becomes attractive for life. If we talk about such indicators as unemployment, it is almost absent, and sometimes completely absent at this stage of the city's development. It can be noted that this stage is more characterized by a shortage of jobs than unemployment. The urban environment is improving at this stage at a very rapid pace. The very rapid and explosive growth of a smart city in the modern sense means that there will be an increase in the number of housings, public goods, jobs, and so on, together with a sharp increase in urban residents (the population will increase due to the very high rate of urbanization).

   In this context, it would be appropriate to mention one of the components of the "smart city" – "smart economy", which will primarily adapt to the increase in urban population. "Smart economy" at this stage of the city's development manifests itself as intensive construction, rapid growth of the city's economy, making the city attractive for investment. As a result, there is no unemployment, and the comfort of living conditions, transport accessibility and infrastructure, on the contrary, are at the stage of active growth.

   The use of intelligent methods of city management allows you to control the ratio of production volumes, employment of the population and the flow of personnel to the city. A "smart city" is characterized by the existence of a single information system that manages the information flow in the city and controls the main indicators. However, no matter how attractive this phase of the city's life cycle is, it physically cannot last long and passes into a phase of slowing growth (expansion).
2. **Stage of city expansion.** The next stage, which follows smoothly from the first, can be characterized as the state of the city, when the current number of jobs that can be provided to residents is not enough for everyone who wants to work. Because of this, unemployment arises, resources and vital goods begin to form a deficit. Infrastructure and businesses are no longer improving at the same pace as they were during the first stage of city development – much more slowly. At this phase of the development of the city, the city-forming enterprise usually stops its development (or greatly slows down the pace of development). Investors start

to lose interest in the city, resulting in a lack of funding to maintain public goods. The city is becoming less livable but still attractive to move to due to the availability of jobs, including highly paid. If this phase is maintained without improvement for a long period of time, the city gradually passes into the next stage of the life cycle – stagnation.

3. **Stage of city stagnation**. Stagnation is a stagnation in the economy, production, public life, etc. From the name of this phase, it is clear what happens to the city in this phase. The gap between the number of jobs and those willing to work is widening. The amount of goods, resources and money in the city stops growing, and is also very insufficient for the current number of residents. Since unemployment is increasing and the attractiveness of jobs in unprofitable enterprises does not cause a desire to stay in this city, people begin to plan to move to other places. The state of the urban environment sharply is deteriorating, the attractiveness of the city for moving to it is extremely low. Such a state of the city in a long period without improvement leads to the transition of the city to the final stage of the life cycle.
4. **Stage of decline of the city**. This stage is considered final. It is characterized by the discomfort of the urban environment: housing conditions are poor, the number of resources is insufficient, there are no jobs, the ecological situation adversely affects the lives of people in this place. The unemployment rate reaches such limits that people stop seeing prospects in this city and start migrating, looking for more profitable places to live. Businesses close, go bankrupt, investors don't pay any attention to the city. The state of the infrastructure is at the lowest level. Cities at this stage of development are called "depressed". These are cities that are not able to cope with the situation in which they find themselves and are no longer able to get out of such a strong decline without the help of the state. However, it is unprofitable for the state to invest resources in "depressed" cities because they become centers of social tension not only within themselves, but also seize nearby territories. The state prefers to permanently liquidate such cities.

9.2.3. Biggest Smart Cities

As of 2019, 278 smart city implementation projects have already been implemented in the world. To rank the „smartest cities" in the world, ratings were used from four independent companies from different countries: Forbes, PwC, Juniper Research (international market research agency) and EasyPark (Swedish IT- company).

Singapore is present in all the ratings considered, and in the rating from Juniper Research it ranks first. The core element of a smart city in Singapore is smart traffic. In this city-state intellectual solutions are implemented in both personal and public transport. Examples include smart traffic lights, the main task of which is to minimize traffic jams and congestion, as well as road sensors, which constantly measure traffic density and adjust the entire transport infrastructure to this. Another example

of smart traffic can be called „smart parking", which also uses a variety of sensors to register the number of free spaces in a particular parking lot, sending this information in a convenient form to the application. The user on the way can assess the state of the parking lot and change the route, if necessary.

Also, Singapore has already launched unmanned vehicles on its roads, and by 2020 all motorists are required to install a navigation system that tracks the position of the car. The city has developed the concept of „Virtual Singapore": a 3D simulation on which tests can be carried out. For example, plan the evacuation of the city in case of an emergency.

Another important aspect of a smart city is the wise use of resources. The Singapore government is trying to optimize water costs and reduce dependence on Malaysia, from where the city imports fresh water. To do this, blocks of Singapore are equipped with sensors that can track the consumption of electricity, water, and other indicators in real time. One of the blocks, for example, is already equipped with a vacuum waste management system and solar panels to generate electricity. All this not only allows you to save money, but also teaches you to take care of resources. As for healthcare, Singapore is introducing smart technologies in this area. Since 2014, the city has been testing a voluntary monitoring system for the elderly. Special sensors were installed in their apartments and on the doors to track the movement of elderly people. When the system determined that a person was motionless for a long time, she warned relatives and medical specialists about this.

London is in second place in the overall ranking. It has become one of the smart cities thanks to its large data center and high-tech solutions to traffic problems. Since London was among the first cities in Europe to face huge uncontrolled traffic on the roads, the authorities of this city have long begun to fight traffic jams and rebuild the transport infrastructure. A smart parking system (like the one in Singapore) has been in place in London since 2014. In addition, since 2002 there has been a system of payments for road congestion, which has now become completely digital (the driver pays for the right to use a car in a traffic-laden zone on weekdays). Since the metro schemes in London are very confusing, and the distances around the city often must be covered with transfers on public transport, so that residents can more comfortably plan their trip, various applications have been developed that build routes around the city. In particular, the SmartLondon system. Statistical analytical system allows you to identify the most fire hazardous houses. Modeling of each area of the city is made up of 60 criteria, including demographic, geological and historical data. In addition, for 1 sq. km of London accounts for over 300 outdoor video surveillance cameras.

New York. This city was added to the rating because of its advanced security systems: the city is covered by a network of video cameras; sensors are installed on the streets that record sound vibrations from shots and send a signal to the police. There is also a modern fire prevention system. BigBelly smart trash bins are installed in the city center – they are equipped with sensors that tell you when it's time to send a garbage truck for them. Also, smart technologies used for street lighting. The system

collects data on the congestion of streets and highways and selects the optimal mode of operation of the lamps.

Barcelona, like the previous cities on the list, uses intelligent parking and traffic systems to monitor congestion. However, this city stands out as "smart" for another very important reason. Barcelona actively uses solar energy. In 2000, the Solar Thermal Decree required all large buildings to produce their own hot water, and in 2006, the city mandated the use of solar water heaters.

In addition, Barcelona stands out for its public transport network. It's one of the cleanest in the world, thanks to its fleet of hybrid buses, as well as the Bicing smart bike initiative, which gives you access to over 400 bike stations through an annual subscription or phone payments.

The city authorities even take out the garbage from the streets "in a smart way." The following system has been introduced in the city: the container is equipped with ultrasonic sensors that give a signal when it is full, this allows significantly save the fuel of garbage trucks and the working hours of city services.

The main "smart" system in Barcelona is Sentilo. 550 sensors – devices for monitoring water supply, light, energy, traffic conditions, noise levels and so on – they collect information about the situation in the city. All data is public and this means that they not only help the authorities to plan development, but also provide a good basis for the development of independent commercial companies.

Copenhagen. The capital of Denmark has taken the top spot in the EasyPark ranking. Copenhagen has the unofficial title of "the most cycling city in Europe", as the cycling infrastructure is very developed here. In Copenhagen, "smart" technologies are used in the field of street lighting and house management. In 2017, a project was launched to equip bicycles with sensors, the main task of which will be to collect and transmit information about the level of road pollution and traffic.

In the same year, the Copenhagen authorities and Hitachi created an "urban data exchange base". Now any individual and legal entity, from an ordinary citizen to the administration of the capital, can place their data here. It looks like the cooperation of social institutions: society, police, administration, and emergency services.

Oslo strives for a progressive and cleaner life. The city authorities monitor the consumption of resources: thus, the city currently uses 65,000 LED lamps, they not only reduce the amount of energy consumed, but also independently regulate the degree of lighting. When it is foggy in the city, such bulbs shine more brightly, when it is light – on the contrary.

In the Norwegian capital, authorities are planning to build an additional 37 miles of bike lanes and ban cars from the city center to get rid of traffic and allow residents to commute comfortably to work. In Oslo, waste is one of the main fuels, and both industrial and standard waste are used. Because the city uses so much waste for fuel, their supply was exhausted in 2013 and the authorities had to import garbage due to frontier. In the future, Oslo aims to reduce fuel emissions by 50%.

9.3. Understanding Security in IoT

Today, there are billions of devices connecting the world to the Internet, and in a short time this number has increased by tens of percent, making the IoT the largest object of attack on the planet. Already, exploits and malware are being developed, deployed, and distributed globally, making life difficult for countless businesses, networks, and individuals.

IoT instances are a separate problem, the security of which was thought of last. Often, systems as simple as sensors are so limited that implementing the industrial-grade protection mechanisms traditional for PCs is too difficult, and sometimes even impossible.

The most important thing about security is to apply it at all levels: from sensors to communication systems, routers, and cloud platforms.

9.3.1. The Main Definitions of Cybersecurity

Here is the list of most popular definitions from cybersecurity area, adopted from (Perry, 2018):

- **Botnets** – hacked and infected devices connected to the internet are controlled to perform joint tasks – usually to send requests in unison to generate huge traffic from different clients. It is also possible to send spam and spyware;
- **brute force** – gaining access to the system or breaking the encryption by trial and error;
- **buffer overflow** exploits a bug or defect in the running software by sending data to a buffer or block of memory that exceeds the allocated space. This allows other information stored in adjacent addresses to be overwritten. An attacker can place malicious code there and execute it by redirecting the current instruction pointer to it. Compiled languages such as C and C++ are especially prone to buffer overflows because they have no internal protections. Most of these errors are due to poorly written code that does not check the bounds of input values;
- **power correlation analysis** is a four-stage attack that allows you to detect secret cryptographic keys stored on the device. First, the dynamic power consumption is analyzed; metrics are recorded for each phase of the normal encryption process. Then the target device is sent for encrypt a few simple text snippets and record energy consumption. Next, small segments of the key (subkeys) are cracked by searching through all possible combinations and calculating the Pearson correlation coefficient between the simulated and real current. At the end, the most successful subkey is collected, which is used is used to obtain the whole key;
- **dictionary brute-force** – hacking the network by systematically entering usernames and passwords from a ready-made dictionary;
- **DDoS attack** – an attempt to disrupt the operation of an Internet service or make it inaccessible by sending it many requests from different (distributed) sources;

- **Fault injection** – This attack consists of sending defective or non-standard data to the device and observing its reaction. For example, if a device starts to perform poorly or shows signs of failure, an attack could reveal a weak point;
- **A man-in-the-middle** attack is a common type of attack that involves placing a device in the middle of a communication flow between two unsuspecting parties. The device monitors, filters and isolates the transmitted information by sending receiving party its modified copy. An attacker can act from the inside, acting as a relay, or simply listen to the data from the outside, leaving it unchanged;
- **NOP-shift** is a sequence of NOP assembler instructions that allows you to „shift" the pointer of the current processor instruction to the area of malicious code. Usually part of a buffer overflow;
- **replay attack** – a network attack with malicious repetition or cyclic reproduction of data by the original sender or an attacker who intercepts, stores, and transmits this data at his own discretion;
- **remote code execution exploit** – allows an attacker to execute arbitrary code. Usually occurs as a buffer overflow over HTTP or other network protocol with the introduction of malicious code;
- **a rootkit** is usually malicious software (although often used to unlock smartphones) that hides the presence of other programs. Rootkits use several highly specialized techniques, such as buffer overflows in kernel components, the hypervisor, and user applications;
- **third-party attack** – extracting information from the victim's system by observing indirect signs of its physical activity; does not imply the search for fresh vulnerabilities or the selection of exploits. Side channel attack includes correlation analysis of power consumption, acoustic analysis, and residual reading. Any data after their removal from memory;
- **spoofing** – an attacker pretends to be another user or replaces a device on the network;
- **Zero-day vulnerabilities** are security holes or bugs in commercial/industrial software that are unknown to developers or manufacturers;
- **Address Space Layout Randomization** is a mechanism for protecting memory and preventing buffer overflow attacks. It makes unpredictable the choice of memory location for loading an executable file, malware, injecting its code after a buffer overflow does not know where it will be loaded, so managing the pointer to the current instruction becomes extremely difficult. ASLR also protects against a library return attack;
- **black hole (or funnel)** – when a DDoS attack is detected, routes are created that direct malicious data from a DNS server or IP address to nowhere. Funnels perform additional analysis to filter out useful data;
- **Data Execution Prevention** – marks areas of memory as executable and non-executable. This prevents an attacker from executing code injected into such areas because of a buffer overflow. The result will be a system error or exception;

- **Deep Packet Inspection** – analyzes each packet (its body and, possibly, header) in the data stream to isolate instructions, viruses, spam and other information filtered by certain criteria;
- **a firewall** is a network security mechanism that allows or blocks the flow of packets between trusted and untrusted zones. Access control lists, or ACLs, can be used to control and manage traffic on specific routes. The firewall can perform stateful filtering and enforce rules based on target ports and traffic state;
- **safe address spaces and non-executable memory** – protects non-executable memory areas that are writable. Protects against NOP shifts;
- **honeypot** – a tool for detecting, redirecting, or reverse-engineering malicious attacks. The honeypot looks like a normal website or network node, but it is isolated and subject to close monitoring. Data and requests entering the device are logged; instruction-based memory access control – a technique for separating data about the returned address inside the stack. Helps protect against ROP attacks and is especially useful in constrained IoT systems intrusion detection system – IDS (English Intrusion Detection System) (network mechanism for detecting network threats through out-of-band analysis of the packet flow) is not tied to the source or destination, therefore it allows you to respond in real time;
- **intrusion prevention system** – blocks network threats using full-fledged linear analysis, statistical methods of detection and filtering by signature;
- **a fake botnet** is a tool that emulates an infected device. Connects to the control server and receives malicious commands that it sends to its botnet;
- **port scanning** – a technique that allows you to find open and available ports on a local network;
- **public key infrastructure** – defines hierarchical mechanisms that guarantee the authenticity of the origin of the public key. The certificate is signed by a certification authority;
- **public key** – generated using the private key and available to external clients. The public key can be used to decrypt hashes;
- **private key** – generated using the public key and never exposed to the public. Stored in a secure location and used to encrypt hashes;
- **the root of trust** – is launched at the very beginning of the device boot and is in an immutable, trusted memory area (such as ROM). If the BIOS or bootloader is available for uncontrolled modification, the root of trust is lost every meaning. RoT is usually the first step in a multi-stage secure boot;
- **secure boot** – a sequence of steps for booting a device. Starts at the root of trust and goes through OS and application startup; each component must be authentically signed. Signatures are verified using the public keys loaded in the previous steps;
- **stack indicators** – protect the stack from overflow and prevent the execution of code placed on the stack;
- **a secure runtime environment** is a secure area of the processor where code and data are protected. Typically, this runtime resides in the core of the main

processor and ensures that secure boot, money transactions, and private key operations are performed at a higher level of security than normal code.

9.3.2. Examples of Cyberattacks

Since the Internet of things consists of hardware networks, protocols, signals, cloud components, frameworks, operating systems, and everything that connects them, all exploits, and attacks on IoT devices can be divided into three main types (Rayes & Salam, 2019):

- Mirai is the most destructive DDoS attack in history, triggered by poor security of IoT devices in remote areas;
- Stuxnet is a government cyberweapon that targets IoT devices in SCADA systems. Caused significant and irreparable damage to Iran's nuclear program;
- Chain Reaction is a research technique for exploiting personal networks. Uses devices like smart light bulbs and does not require an internet connection.

Mirai is the name of a malware that infected Linux-based IoT devices in August 2016. The attack was carried out from a botnet that generated a huge DDoS load. The most notorious victims included Krebs on Security (a popular security blog), Dyn (a very popular and in-demand DNS provider) and Lonestar (a major cellular operator in Liberia). Less significant targets included Italian political websites, Minecraft servers in Brazil, and Russian online auctions. Indirect victims of the DDoS attack on Dyn were its customers, among which were such huge services as PlayStation Network, Amazon, GitHub, Netflix, PayPal, Reddit and Twitter. A total of 600,000 devices were infected and became part of the botnet. The Mirai source code was published on the **hackforums.net** hacker forum. By analyzing it, as well as using tracing and studying log files, the researchers revealed the principle of operation and the chronology of the attack:

- **victim search** – fast asynchronous scanning using TCP SYN packets and checking arbitrary IPV4 addresses. The malware paid special attention to TCP port 23, which belongs to SSH/Telnet, and port 2323. If a suitable port was found, the second stage began. Mirai has a built-in blacklist of addresses to avoid. It contained 3.4 million records belonging to the US Postal Service, Hewlett Packard, GE, and the US Department of Defense. The scanning speed reached 250 bps. For a botnet, this is slow. Attacks like SQL Slammer generated scan queries at 1.5 Mbps. The fact is that IoT devices have much more modest computing capabilities compared to desktop and mobile computers;
- **simple Telnet brute-force** – at this stage, the malware tried to establish a working connection with the victim via Telnet by sending 10 randomly selected login-password pairs from a list of 62 pairs. If successful, the hacked computer connected to the C2 server. Later variants of the Mirai learned how to perform RCE exploits;

- **infection** – the server passed to the potential victim a loader program that was responsible for determining the version of the operating system and installing malware tailored for a specific device. The loader then looked for and terminated any competing processes that were using ports 22 or 23 (including other malware that might be on the device). After that, the bootloader was removed, and the process name was masked to hide its presence. The malware was not stored on persistent storage and disappeared after a reboot. At the end, the bot would go to sleep waiting for further commands.

Mirai's victims included IoT devices such as IP cameras, DVRs, consumer routers, IP telephony, printers, and digital set-top boxes. The malicious binaries supported 32-bit versions of ARM, MIPS, and x86.

Stuxnet is the first documented cyberweapon designed to damage the assets of another country. It was a worm that targeted Siemens Programmable Logic Controllers, or PLCs, based on SCADA. He used a rootkit to change the rotation speed of the motors directly controlled by the PLC. The creators of this virus have done everything to ensure that it attacks only devices with driven frequency-controlled drives, which are connected to Siemens S7-300 PLC modules and rotate at 807 Hz or 1210 Hz, as they are commonly used in pumps and gas centrifuges for uranium enrichment.

The attack began in April or March 2010. The infection process consisted of the following steps:

- **initial infection** – the worm first infected a Windows computer using vulnerabilities found in previous virus attacks. It is believed that this was done via a USB disk connection. At the same time, several zero-day exploits were used at once (unprecedented sophistication). The exploits ran the rootkit in user and kernel mode and then installed a device driver with the correct certificate stolen from Realtek. A kernel-mode driver was needed to hide Stuxnet from various antivirus packages;
- **attack on Windows and distribution** – after installation through a rootkit, the virus began to search the Windows system for files related to the Siemens SCADA controller version WinCC / PCS 7, also known as Step-7. If successful, the worm tried to connect to the C2 server over the Internet using fake URLs to update to the latest version. He then searched the disk for a file called s7otbdx.dll, which was an important communication library for communication between Windows and the PLC. The Step-7 controller included a built-in password database, which was hacked using another zero-sum exploit day. Stuxnet intruded between the WinCC system and s7otbdx.dll, performing a man-in-the-middle attack. First, the virus recorded information about the normal operation of centrifuges;
- **destruction** – when it was decided to coordinate the attack, the virus played back the pre-recorded data and sent it to the SCADA controllers, who were unaware that the system was compromised or was behaving unusually. The damage

was done in two different coordinated attacks based on PLC manipulation, resulting in damage to the entire uranium enrichment complex. Every 27 days, the centrifuge rotors were run for 15 or 50 minutes, causing them to gradually wear out and crack in their shafts. In addition, the enrichment process was disrupted.

The attack on Iran's main uranium enrichment facility at Netanz is believed to have disabled more than a thousand centrifuges. Since then, the Stuxnet code has become public domain and has become a kind of playground for creating derivative exploits (github.com/ microphone/stuxnet).

Chain Reaction is an academic study of a new type of cyberattack that targets personal mesh networks without an internet connection. It also shows how vulnerable remote IoT sensors and control systems can be. The target of the original attack was Philips Hue light bulbs, which can be found in so-called smart homes with support for control via the Internet or mobile applications.

The exploit can scale up to attack entire smart cities – enough just screw in one infected light bulb. Philips Hue light bulbs raise a mesh network based on the Zigbee protocol, which is part of the Zigbee Light Link (ZLL) initiative to ensure interoperability between different lighting methods. ZLL messages are not encrypted or signed, although cryptography is used to protect the keys exchanged when a light bulb is added to a mesh network. As a result, the master key, known to all members of the ZLL consortium, was leaked.

In addition, according to the ZLL standard, the light bulb to be connected must be near the initiator, which prevents it from taking control of the light bulbs of its neighbors. The Zigbee protocol also provides a contactless method for flashing devices, although firmware packages are encrypted and digitally signed. The attack plan that the researchers conducted consisted of four stages:

- crack the encryption and digital signature of the firmware package;
- write and deliver an infected firmware update to a single bulb using cracked encryption and keys;
- the compromised light bulb will join the network using the master key stolen earlier and bypass the proximity-based protection using a zero-day defect found in the widespread Atmel AtMega component;
- after successfully connecting to the Zigbee mesh network, the malicious code will be sent to neighboring light bulbs, which will lead to their rapid infection. The spread of the virus will occur according to the theory of percolation and will cover all lighting fixtures in the city.

To encrypt contactless firmware updates, Zigbee uses AES-CCM. To crack it, the attacker uses correlation and differential power analysis (Correlation Power Analysis [CPA] and Differential Power Analysis [DPA]). This is a sophisticated form of attack that involves placing a light bulb on a special stand and measuring the energy it consumes. Given advanced controls, it is possible to measure the dynamic power consumption of a processor that is executing an instruction or moving data (for example,

while an encryption algorithm is running). This is a simple power consumption analysis that leaves little chance of breaking the key. The CPA and DPA methods are more advanced and use statistical correlation. Instead of trying to recognize individual bits, CPA can operate on whole bytes. Power indicators are taken using an oscilloscope and are divided into two sets depending on the intermediate value that is hacked: in the first it is 1, and in the second it is 0. The real value is calculated by subtracting the average from these sets.

Using DPA and CPA, the researchers were able to hack the Philips Hue lighting system:

- the CPA method was used to crack AES-CBC. The attacker had no key, no random number, no initialization vector. Thanks to this approach, a key was obtained, which was then used using the same method to crack a random number;
- the DPA method was used to crack the AES-CTR counter mode and subsequently the encryption algorithm that was used when packaging the firmware. The researchers found 10 sites in which the AES-CTR mode was presumably performed, leaving 10 potential solutions;
- the researchers then focused on breaking Zigbee's proximity-based security to connect to the network. As a result of studying the source code of the bootloader that was used in the Atmel chip, a zero-day vulnerability was found. At the time of sending a scan request to Zigbee, the check for remoteness passes successfully. To get around it, it was enough to start the session with any other message.

This allowed the researchers to connect to the network. A real attack could cause a hacked light bulb to infect its neighbors within a 100-meter radius by giving them malicious code to disable updates so they can't be restored. In effect, the light bulbs would be under the control of an intruder and would have to be destroyed. The researchers were able to create a fully automated attack system and attach it to an unmanned aerial vehicle that systematically circled within range of Philips Hue bulbs on campus and infected each one.

9.3.3. Device and Physical Security

The first level of hardware security is to **establish a root of trust** (RoT). RoT is a hardware-authenticated boot process that ensures that the source of the first executable instruction cannot be changed. This is a key step in the boot process and is involved in the further startup of the system – from the BIOS to the OS and applications. RoT is the basic protection against rootkits. Each stage of the download process authenticates the next stage, thus forming a chain of trust.

The root of trust can be used different launch methods (Perry, 2018):

- loading image and root key from firmware or immutable memory;
- storing the root key in a one-time programmable memory using fuse bits;
- loading code from a protected memory area into a protected storage.

The root of trust must authenticate each subsequent download step. To do this, at each stage, a set of keys with a digital signature is used.

Public and private keys are the key to a secure system. A proper management mechanism is required to protect them. One of the most popular hardware key protection standards is TPM (Trusted Platform Module). Its specification was created by the Trusted Computing Group and is part of ISO and IEC. The current version of TMP 2.0 was released in September 2016. Equipment supplied to the US Department of Defense must support TPM 1.2.

The TPM is a separate hardware component with an RSA key embedded at the manufacturing stage. Typically, TPM is used to store, protect, and manage keys in scenarios such as disk encryption, trust root boot, hardware and software authentication, and password management. The TPM can generate a hash of a verified hardware or software configuration that can help detect third-party tampering at runtime. This technology is also used in creating SHA-1 and SHA-256 hashes, AES block encryption, asymmetric encryption, and random number generation.

The two **main protection mechanisms in the CPU and OS**, which noteworthy are non-executable memory and address space layout randomization. Both are designed to make it harder or to prevent the process of injecting malicious code based on a buffer or stack overflow (Perry, 2018):

- **non-executable memory** is a hardware mechanism by which the operating system makes memory regions non-executable. The goal is to ensure that only areas of memory that contain verified and genuine code can be executed. When malware attempts to inject via a stack overflow, the system will mark the corresponding section as non-executable, because of which the current instruction pointer shift to this section will result in a hardware exception. The marking of non-executable memory is done using the NX bit (via the associative translation buffer). On Intel and ARM platforms, this bit is called XD (English eXecute Disable – turning off the completion) and, accordingly, XN (English eXecute Never – never execute). This technology is supported on most operating systems such as Linux and Windows, as well as on some real-time systems;
- **address space allocation randomization** – ASLR is more of a feature of virtual address space handling in the operating system than a hardware feature, but it is also important to consider. This technology protects against buffer overflows and library return attacks. Such hacking techniques require the attacker to understand the structure of memory and involve the deliberate execution of certain benign code or libraries. This is not an easy task, especially if the address space changes randomly on every boot.

Many IoT devices use persistent storage at the edge or router/gateway. Smart fog nodes also need somewhere to store their data. **Data security** is key to preventing the installation of malware and protecting sensitive information if your device is stolen. Many storage devices, such as flash drives and hard drives, support encryption and security technologies. In addition to encryption, you should also

take care of the security of drives being decommissioned. Retrieving content from older storage systems is a relatively simple task. There are additional standards that describe the safe process of deleting data from a drive (whether it be a magnetic platter disk or phase change memory). In addition, the NIST Lab publishes documents on the secure destruction of content, such as the NIST Special Publication 800-88 for Secure Erase".

Penetration resistance and physical security play an important role in the Internet of Things. Many IoT devices are hosted remotely, without any protection. An attacker with direct access to an IoT device can use any tool to compromise the system, as we saw with the Chain Reaction exploit.

An example of a side-channel attack using power analysis has already been presented; hacking can also be done based on time, cache, electromagnetic field radiation and scan chain. The main feature of attacks by third-party channels is that a hacked device, in fact, turns into a test site. This means that it will be monitored in a controlled environment and its activity will be measured in every possible way. In addition, techniques such as DPA use statistical analysis to infer patterns between random input and output. This approach is applicable only if the system exhibits identical behavior with the same input.

Techniques for preventing these attacks are well known and can be licensed and used on a variety of hardware. Countermeasures include the following:

- changing the encryption function to minimize key usage. Using keys valid only for the duration of the current session and based on the hash of the original key;
- for timing attacks: random insertion of functions that do not break the algorithm;
- using random machine instructions to create a large working function that is hard to crack;
- removal of conditional branches that depend on the key;
- for power-based attacks: minimizing leaks and limiting key operations. This will reduce the attacker's working set of metrics;
- interference in power lines. Variation in the execution time of operations or shifting of timers;
- changing the order of independent operations. This reduces the degree of correlation around the calculations in the S-box.

9.3.4. Encryption

Encryption and secrecy are mandatory for IoT devices. They help secure the interaction by protecting the firmware and the authentication process. Encryption can be divided into three main categories (Perry, 2018):

- **symmetric encryption** – the same key is used for encryption and decryption. Symmetric algorithms are RC5, DES, 3DES and AES;
- **public key encryption** – the key used to encrypt data is publicly available. But only the receiving party has the private key to decrypt the message. This type

of encryption is also called asymmetric. Asymmetric cryptography is used to provide data privacy, authentication, and non-repudiation. Public keys are used in well-known Internet protocols for encryption and messaging, such as Elliptic Curve, PGP, RSA, TLS, and S/MIME;

- **cryptographic hashing** – binds data of arbitrary size to a bit string (called a digest). A hash function is created „one-way" from the start. In fact, the only way to recreate the final hash is to try all possible input combinations (the hash function cannot be performed in reverse). Examples of one-way hashes are MD5, SHA1, SHA2, and SHA3. They are commonly used to encrypt digital signatures in firmware images, imitations, and authentication. When encrypting small strings, such as a password, the input may be too short to produce a valid hash; in this case, a salt or a public string is added to the password to increase entropy. Salt is a kind of key derivation function (KDF).

In cryptography, terms such as plain text and ciphertext are used, which denote unencrypted input and, accordingly, encrypted output. The current encryption standard is AES (Advanced Encryption Standard); it replaced the old DES algorithm developed in the 1970s. AES is part of the FIPS specification and the ISO/IEC standard 18033-3 which are used all over the world. AES algorithms are based on fixed length blocks of 128, 192, or 256 bits. Messages that exceed the block length are split into several parts. AES consists of four basic encryption steps.

Block ciphers are algorithms that are based on a symmetric key and process data in sequential blocks. Modern ciphers are based on an article on industrial encryption written by Claude Shannon in 1949. An encryption mode is an algorithm that describes the repeated use of a block cipher to transform large amounts of data consisting of many blocks.

Majority Modern ciphers use an Initialization Vector, or IV, which turns the same input into a different ciphertext each time. The AES algorithm has several modes of operation:

- **Electronic Codebook**, or ECB is the simplest type of AES encryption; it is used in combination with other modes to improve security. The data is divided into blocks, each of which is encrypted separately. Identical blocks produce the same result, making this approach relatively unreliable;
- **Cipher Block Chaining**, or CBC – before encryption, the plain text is XORed with the previous encrypted block;
- **ciphertext feedback**, or CFB- similar to CBC, but forms a stream of ciphers (the output of the previous cipher serves as input for the next one). CFB uses the previous encrypted block to generate input for the current cipher. Because of this dependency, CFB cannot be executed in parallel. Stream ciphers allow block loss in transmission; in this case he will be restored based on subsequent blocks;
- **output feedback chaining**, or OFB mode – this mode is like CFB but allows you to apply error correction codes even before encryption;

- **Counter**, or CTR – turns a block cipher into a stream cipher using an incremental counter that parallelizes the input to each block cipher, which speeds up execution. The input is a combination of a counter and a randomly generated number;
- **Message Authentication Code**, or MAC is used to authenticate a message and confirm that it came from the claimed sender. The recipient then adds a mock insert to the message for later authentication.

AES-CCM uses a double cipher: CBC and CTR. AES-CTR (or counter mode) is used for general decryption of an incoming ciphertext stream that contains an encrypted spoof. AES-CTR decrypts both the simulation inserts and the data itself. At this stage of the algorithm, the so-called expected imitation insertion is formed; the original frame header and the decrypted blocks output from the AES-CTR are marked as input. The data is decrypted, but authentication requires a spoof that is computed in AES-CBC; if it is different from what is expected during the AES-CTR stage, this means that the data may have been changed during transmission.

Asymmetric cryptography is also called public key encryption. Asymmetric keys are generated in pairs (for encryption and decryption); they can be interchangeable – that is, one key can encrypt and decrypt, although this is not a requirement. But usually, a pair of keys is generated – one public and the other private.

The first public-key asymmetric encryption method was described in the **Rivest-Shamir-Adleman (RSA) algorithm**, developed in 1978. It implies that the user must find and publish the product of two large prime numbers and an auxiliary value (public key). The public key allows messages to be encrypted and is available to anyone, but the prime multipliers remain private.

Perhaps the most well-known type of asymmetric key exchange is the **Diffie-Hellman protocol** (named after Whitfield Diffie and Martin Hellman). Typical for asymmetric cryptography is the concept of a one-way function with a Trapdoor Function, which takes a given value A and returns an output B; but in this case, A cannot be obtained from B. The Diffie-Hellman method allows both parties to exchange keys without knowing in advance about the shared key s.

The strength of this key exchange is the generation of a truly random number for each private key. The slightest predictability in the operation of the pseudo-random number generator can lead to breaking the cipher. However, the fundamental disadvantage here is the lack of authentication, which opens up the possibility of a MITM attack.

The third type of encryption technology is hashing functions. They are commonly used to create digital signatures and are considered "one-way" – with no way to reverse. To recreate the original data that passed through the hashing function, one would have to iterate over all possible combinations of input. Key features of the hash function:

- always generates the same hash from the same input; fast in computation, but not instantaneous;
- irreversible;
- cannot generate the original message from the hash;

- the slightest change in input causes significant entropy and completely changes the output;
- two different messages will never have the same hash.

Algorithms of the SHA family are actively used in:

- Git repositories;
- digital signatures of TLS certificates for web browsers (HTTPS);
- authenticate the contents of a file or disk image.

Most hash functions are based on the Merkle-Damgard structure. The example below splits the input into blocks of the same size, each of which goes through a compression function using the compression results of the previous block. An initialization vector is used to select the initial value. Thanks to the compression function, the hash is resistant to collisions.

The SHA-1 algorithm is based on the Merkle-Damgard structure. In general, messages that are fed into the SHA algorithm should be less than 264 bits. They are sequentially processed in 512-bit blocks. The SHA-1 standard has been superseded by more robust versions such as SHA-256 and SHA-3. The possibility of "collisions" was found in SHA-1 hashes; and although for This requires approximately 2^{51} to 2^{57} operations, cracking the hash on a rented graphics adapter will cost only a few thousand dollars. In this regard, it is recommended to switch to other varieties of SHA.

Asymmetric cryptography (with a public key) is the basis of trading and interaction on the Internet. It is ubiquitous in SSL and TLS connections. A typical example is when the transmitted data can be encrypted with the public key (that is, by anyone), but it can be decrypted by the person who owns the private key.

Another use is with digital signatures, where the sender signs binary data with a private key, and the recipient can verify their authenticity if they have the public key. To establish a reliable issuance of public keys, a process called public key infrastructure (English Public Key Infrastructure, or PKI) is used. Authenticity is guaranteed by certifying centers (English Certificate Authorities, or CA), which manage roles and rules, creating distributed digital certificates.

The largest public publishers of TLS certificates are Symantec, Comodo and GoDaddy. Public key certificate formats are described by the X.509 standard. This is the basis for secure communication in the TLS/SSL and HTTPS protocols. X.509 defines attributes such as the encryption algorithm used, expiration dates, and the issuer of the certificate.

Transport Layer Security, TLS, has incorporated all the cryptographic protocols and technologies that we have already discussed. Initially, the level of secure sockets (eng. Secure Sockets Layer, or SSL) was introduced in 1990, but after 9 years it was replaced by TLS technology. TLS 1.2 includes the SHA-256 hash generator, which was added in place of SHA-1 to improve security.

The encryption process in TLS is as follows:

1. the client opens a connection to a server that supports TLS (port 443 for HTTPS);

2. the client provides a list of ciphers it supports;
3. the service selects the cipher and encryption function and notifies the client;
4. the server transmits to the client a digital certificate issued by a certification authority and containing a public key;
5. the client confirms the authenticity of the certificate;
6. One of two methods is used to generate the session key:
   - a random number is sent to the server, previously encrypted using its public key. The server and client then create a session key based on it, which is used throughout the interaction;
   - The session key for encryption and decryption is generated using the Diffie-Hellman protocol. The resulting key is used until the connection is closed.
7. the interaction goes into an encrypted channel.

9.3.5. Blockchain and Cryptocurrency in IoT

Blockchain is a public, digital, decentralized registry (ledger) or a chain of cryptocurrency transactions. The first cryptocurrency blockchain is Bitcoin, but in addition to it, there are more than 700 new currencies on the market. The strength of this technology lies in the absence of a single entity that controls the state of transactions. It also provides system redundancy, forcing each member to keep a copy of the registry.

Assuming that the participants in the system are not inclined to trust each other, their interaction should be based on consensus. This raises the question: why transfer data or currency on a blockchain if we have already solved the problems of authentication and security with asymmetric cryptography and key exchange? The fact is that the transfer of funds and valuable information requires something more.

Imagine that we have two devices (let's call them Bob and Alice). According to information theory, when Bob sends a message or piece of data to Alice, he keeps a copy of the transmitted information. When exchanging money or contracts, the data must leave the source and appear at the destination. They must exist in a single copy.

Authentication and encryption provide interoperability, but we a new tool is needed to transfer ownership. Secure blockchain-based cryptocurrencies are of great importance for the Internet of Things (Perry, 2018):

- **direct cash payments between devices** – the Internet of things should be ready to support devices that exchange services for currency;
- **supply chain management** – the immutability and security of the blockchain can come in handy in logistics, inventory, and movement of goods, making paper accounting unnecessary. All containers, movements, locations, and states can be tracked, verified, and certified. Attempts to change, delete, or change accounting information become impossible;
- **solar energy** – think of solar energy as a service. In this case, solar panels are installed on the roof of a residential building, which can not only generate electricity

for residents, but also supply it to the public grid (for example, in exchange for so-called carbon credits).

The part of Bitcoin that belongs to the cryptocurrency is not a blockchain as such. Bitcoin is an artificial currency, which has no value and is not backed by anything (unlike gold). It cannot be felt; it exists only within the network. And finally, the number of „coins" Bitcoin is not controlled by a central bank or government. It is a completely decentralized technology.

Like other blockchains, it is based on public key cryptography, a large and distributed peer-to-peer network, and a protocol that defines the structure of Bitcoin. In 2008, Satoshi Nakamoto (pseudonym) published a white paper in a cryptography mailing list entitled "Bitcoin: A Digital Peer-to-Peer Cash System." In 2009, the first Bitcoin network was launched, in which Satoshi generated the first block (Betts, 2016).

The concept of blockchain implies the existence of a block, which represents the current fragment of the ledger. A computer connected to a blockchain network is called a node. Each node participates in the authentication and transmission of transactions; to do this, he gets a copy of the registry and, in essence, becomes its administrator.

Distributed networks based on peer-to-peer topologies are ideal for Bitcoin. Metcalfe's law applies to the Bitcoin network since its size determines the value of the currency. The network keeps a chain of records (registry). The question arises: who would want to volunteer their computing resources to monitor a journal?

The answer is a mining-based reward system. First, a request is made, which is broadcast over a peer-to-peer (P2P) network of computers (nodes). This network is responsible for authenticating its users, during which the transaction itself is also verified. Then transactions are merged into a new block of data in a distributed ledger. Once filled, the block is added to the existing blockchain and is no longer subject to change.

Here is a qualitative analysis of the blockchain in general and the operation of Bitcoin in particular (Perry, 2018). It is important to understand these fundamental principles, which are based on all the security features that we covered earlier in this chapter:

- **digitally signed transaction** – Alice wants to give Bob 1 Bitcoin. The first thing to do is to announce it publicly. To do this, Alice writes the message „Alice will give Bob 1 Bitcoin" and confirms it with a digital signature based on her private key. Anyone with the public key can verify the authenticity of the message. However, Alice can repeat her message and thus counterfeit the money;
- **unique identification** – to solve the counterfeiting problem, Bitcoin creates a unique serial number, just like the US Treasury does on its banknotes. To do this, instead of a number that is assigned in a centralized way, a hash is used. This hash is automatically generated during the transaction and allows it to be identified. Another big problem is double spending. Even if the transaction is signed and has a unique hash, Alice can try to transfer the same bitcoin to other participants. Bob will validate the transaction initiated by Alice and everything will

fit. But if Alice performs the same transaction, only in relation to Charlie, she, in essence, will wrap the system around her finger. The Bitcoin network is very large, but the possibility of theft of funds in it, although insignificant, is still present. To protect against double spending, Bitcoin users accepting payments via the blockchain, awaiting confirmation. Over time, more and more confirmations appear, which increases the chance of successfully passing the test;

- **authentication by other nodes** – To avoid double spending on the blockchain, the recipients of the transaction (Bob and Charlie) transmit the payment information to the network and ask other participants to verify its authenticity. Such verification is not free of charge; proof of work – the problem of double spending is still not fully resolved. Alice can take control of the network with her own servers and claim that all her transactions are genuine.
- To eliminate this possibility once and for all, the **concept of proof-of-work** was added to Bitcoin. It has two aspects. First, verifying the authenticity of a transaction must be computationally intensive. It should be something more complex than matching keys, usernames, transaction IDs, and other trivial authentication attributes. Second, users should be rewarded for helping to confirm other participants' money transactions (see next step); to force users verifying transactions to do some work, a randomly generated number is added to the transaction headers. Bitcoin hashes this number, along with the header message, using the secure SHA-256 algorithm. This hash is called the target hash; it is less than 256 bits in size, and its content is constantly changing. The smaller this value, the more resources are spent searching for the original message. Because each hash essentially generates a completely random number, users must compute multiple SHA-256 values. On average, each value takes approximately 10 minutes.
- **Bitcoin Mining Incentives** – To encourage building a peer-to-peer network to verify other people's transactions, users are rewarded for their work. There are two ways. The first is through Bitcoin mining, which benefits participants who verify blocks with transactions. The second way is to pay a commission for the transaction. The miner gets the commission which helps to verify the authenticity of the block. Initially, the commission was equal to zero, but with the growth of the popularity of Bitcoin, it also began to grow. On average, a successful transaction is rewarded with $35 (in the form of BTC). If the block is processed in an accelerated mode, the commission can be raised. Thus, even after calculating all the hashes in the current generation of Bitcoin, users will have an incentive to support transactions;
- **Block chaining security**, the order in which transactions are executed, is also of great importance for the integrity of Bitcoin. If funds are transferred from Alice to Bob and then from Bob to Charlie, these events must be recorded in the ledger in that order. To do this, transactions in the blockchain are linked. Each new block that enters the network contains a pointer to the last block in the chain that was verified. In Bitcoin, a transaction is not valid until it is added to the longest chain, and until at least five other blocks are confirmed after it. This solves the asynchrony problem if Alice tries to pass the same funds to Bob and Charlie.

Recently, a new **cryptocurrency IOTA** has appeared, designed specifically for the Internet of things. Its architecture is based on a Directed Acyclic Graph (DAG), and the chain of trust is formed by the IoT devices themselves (Betts, 2016). Bitcoin provides a commission for each transaction.

There are no fees in IOTA. This makes it possible to conduct microtransactions, which is very important in the context of the Internet. of things. For example, multiple clients can subscribe to sensor readings via MQTT. Overall, this service has some value, but each individual transaction is so insignificant that the fees charged on the Bitcoin network would be higher than the cost of the data itself. The IOTA architecture has the following features:

- control over funds is not centralized;
- on the blockchain, users can form large groups to increase the number of blocks they can generate and the corresponding reward. This can lead to a concentration of influence and harm the network;
- no need for expensive equipment. Cryptocurrency mining on the Bitcoin network requires powerful processors that can handle complex calculations;
- micro- and nanotransactions at the level of individual IoT devices; reliable protection against hacking by simple enumeration, even if quantum computers are used; through IOTA, you can transfer not only currency, but also data. At the same time, there is full support for authentication and protection against substitution;
- in the IOTA network, the content of a transaction can be anything, so on its basis it is possible to build a national voting system that is protected from third-party interference;
- the role of a service can be performed by any device with a compact chip. IOTA allows you to rent anything: a drill, a personal router, a microwave oven, or a bicycle – all it takes is a small chip or microcontroller.

The DAG graph in IOTA is called a tangle and is used to store transactions in the form of a distributed ledger. Transactions are used by nodes (IoT devices) and make up a DAG tangle. If transactions A and B are not directly connected by a directed edge, but it is possible to pass from A to B a route no less than the distance between these two points, A is considered to indirectly confirm B. There is also the concept of a primary transaction. Since a tangle cannot be started by mining graph edges and there are no incentives or fees, each node must hold all the tokens; the primary event sends them to the founders' addresses.

This is the static set of all tokens new, which will never be replenished again. Each new transaction must confirm (or reject) the previous two; this process forms a straight edge in the graph and is called direct confirmation. To make a transaction, it is necessary to perform "work" on behalf of the tangle. The job is to find a random number (nonce) that matches the hash of the fragment of the confirmed transaction.

Thus, using IOTA, the network becomes more distributed and secure. Transactions can be confirmed many times. With the increase in their number, the confidence in their eligibility grows. When attempting to approve an unauthorized

transaction, a node runs the risk of rejecting its own transaction and being kicked out of the tangle.

9.3.6. Improving IoT Security

On the Internet of Things, security must be considered from the very beginning, and not after the fact, after design or commissioning is completed – at these stages it will be too late. In addition, the approach to security should be comprehensive and cover all aspects: from hardware provisioning to the cloud. This section looks at a simple IoT project with security that permeates all its layers, from the sensor to the cloud infrastructure. We'll try to deploy it with different precautions to make it harder for potential attackers.

If you focus on any one aspect of the Internet of Things, the resulting security chain will have weak links. Security must permeate all levels of the system: from the sensor to the cloud. This is an integrated approach. Each component in the control and data chain must have checklist of security settings and potential threats.

Below is a list of time-tested recommendations and ideas related to security (Perry, 2018):

- Use the latest versions of operating systems and libraries with all necessary patches.
- Use hardware that supports security features such as secure runtimes, TPMs, and non-executable address spaces.
- Obfuscation of code in the hope that an attacker cannot unravel it is a relatively hopeless undertaking.
- Sign, encrypt and secure your firmware and software images – especially those available on the company's website.
- Choose the initial password randomly.
- Use root of trust and secure boot to ensure that your customers' devices are running genuine software.
- Remove passwords from the firmware code.
- All IP ports should be closed by default; use address space allocation randomization, stack indicators and safe memory segments that are supported in modern operating systems.
- Use automatic updates. Give manufacturers a mechanism to fix bugs and vulnerabilities in production systems. To do this, the software architecture must be modular.
- Plan decommissioning ahead of time. IoT devices can work for a long time and productively, but someday they will have to be disposed of. This includes safely removing and destroying all read-only memory (flash) modules in the device.
- Offer your customers and users rewards for found bugs – especially those that can lead to zero-day vulnerabilities.
- Subscribe to active threat alerts sent by US-CERT to stay up to date on the latest exploits and cyberattacks; building a project around simple (and insecure)

protocols like MQTT or HTTP might be tempting, but you should ship your devices with TLS or DTLS-based authentication enabled.
- Encrypt data from the sensor to the cloud.
- Use anti-debug fuse bits. Detonate them during production for safe debugging before release.

With well-known viruses such as Mirai and Stuxnet specifically targeting IoT devices, IoT system specialists must be concerned about the security of their architectures from the outset. The Internet of Things is an ideal environment for executing all sorts of attacks. Usually, systems of this type have less mature protection compared to PCs. IoT devices represent the largest attack surface on the planet, and the remoteness of some of them allows attackers to gain physical access to equipment unthinkable in a secure office environment. These threats require serious attention, as their consequences can affect individual devices, cities, or even entire countries.

References

[1] Betts, R. (2016). Architecting for the Internet of Things. O'Reilly: Beijing, Boston, Tokyo.
[2] Gantz J., Reinsel D. The Digital Universe in 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East. –URL: https://www.emc.com/collateral/analyst-reports/idc-digital-universe-united-states.pdf
[3] McEwen, A., Cassimally, H. (2014). Designing the Internet of Things. John Wiley and sons: Chichester, West Sussex.
[4] Perry, L. (2018). Architecting Internet of Things. Packt Publishing: Birmingham, Mumbai.
[5] Rayes, A., Salam, S. (2019) Internet of Things from Hype to Reality. The Road to Digitization. Second Edition. Springer: Cham, Switzerland.