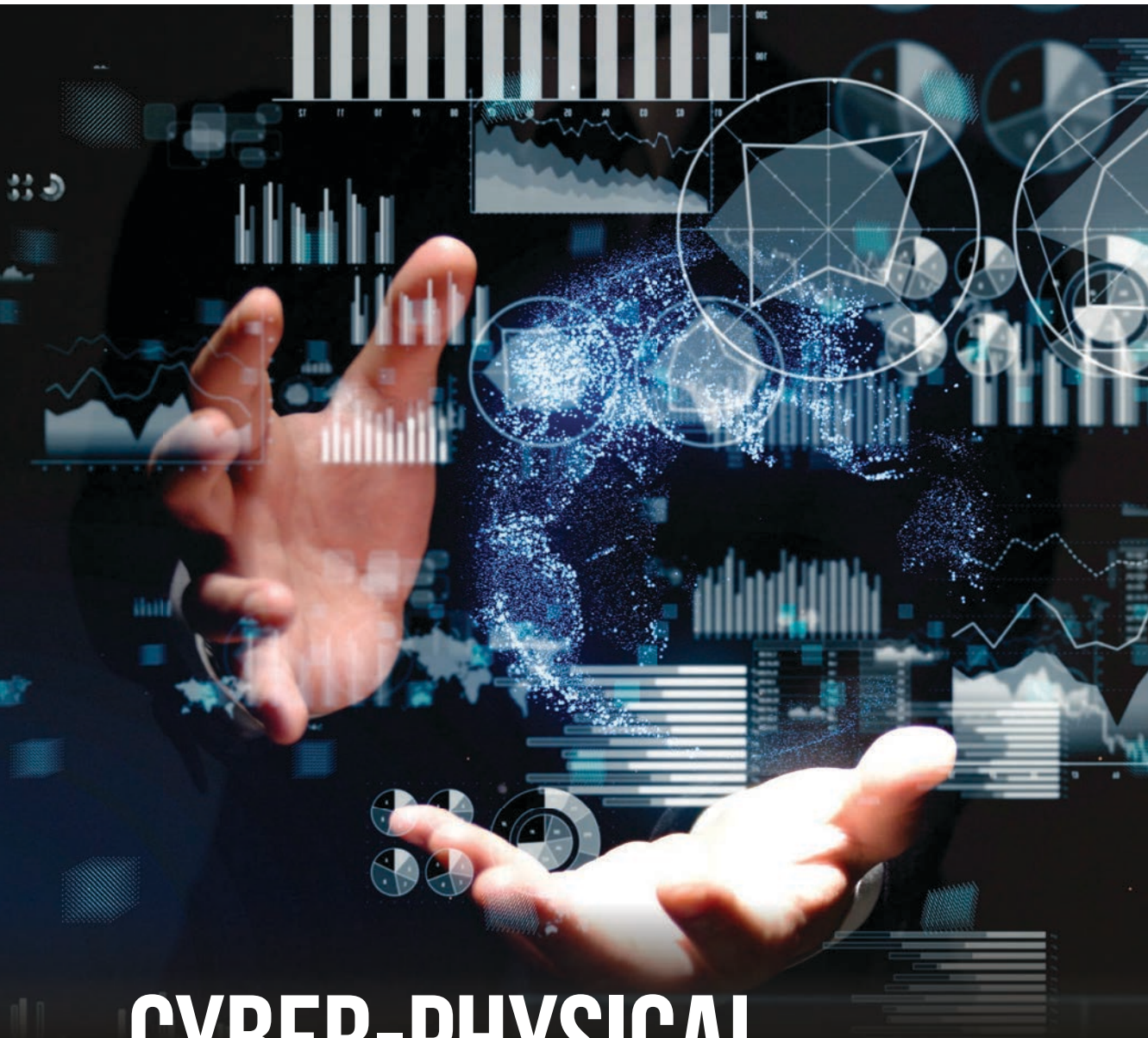
CYBIPHYS



CYBER-PHYSICAL SYSTEMS MODELLING AND SIMULATION

University of Cyprus

Imperial College London

Co-funded by the Erasmus+ Programme of the European Union

Cyber-Physical Systems Modelling and Simulation

Editors: Nikolas Flourentzou, Stella Hadjistassou, and Irina Ciornei

Contributors:

Natalia Morkun
Sergey Zaitsev
Volodymyr Kasymyr
Markos Asprou
Denys Kuznietsov
Iryna Zavsiehdashnia

Nikolas Florentzou
Lenos Hadjidemetriou
Lazaros Zacharia
Oleksandr Khropatyi
Oleh Lohinov
Andrei Varuyeu

Volodymyr Sistuk
Nadezhda Kunicina
Anatolijs Zabasta
Joan Peuteman
Boris Horlynski
Volodymyr Prystupa

Rasa Brūzgienė
Andrejs Romanovs
Igors Utesevs
Antons Patlins
Stella Hadjistassou
Irina Ciornei

RTU Press

2022

Cyber-Physical Systems Modelling and Simulation. Riga, RTU Press, 2022. – 239 p.

The textbook is devised for students specialising in applied physics and electrical engineering. It gives an overview of current computer control of electrical technologies and IT elements, as well as explains their operating principles. It can be useful for students and professionals focusing on applied informatics issues. The book provides the theoretical background and offers essential procedures for modelling realistic systems. The near to reality synthetic models are applied on respective simulation packages and platforms for harvesting accurate data for analysing the CPS.

Scientific Editors:

Nikolas Flourentzou
Stella Hadjistassou
Irina Ciornei

Language Editor Daina Ostrovska
Layout Design Ģirts Semevics
Published by: RTU Press
Phone: +371 67089123
e-mail: izdevnieciba@rtu.lv

ISBN 978-9934-22-675-5 (pdf)

Contents

Summary

Summary

Cyber-Physical Systems (CPSs) are interconnected real-time systems that combine discrete computation with continuous physical processes in a controllable way. CPSs are essential to form resilient critical infrastructures (CI), which significantly impact the quality of life, as they are assets (such as facilities, systems, sites, and networks) necessary for delivering the important services upon which daily life depends.

The ever-increasing use of CPSs extends the risk of attacks, which conceptually can become threat for the CI. Preparedness and protection of CPSs are important but not enough to keep them secure. Further analysis of CPSs is required to design robust and secure systems. CPSs need to be resilient in known and unknown threats, which can harm the CPS itself, the relevant infrastructures and most importantly, every vital societal function.

Over the years, many methodologies have been developed for safely and securely control and monitoring CPSs. This e-book not only provides the theoretical background – literature review, but also offers essential procedures for modelling realistic systems. The near to reality synthetic models are applied on respective simulation packages and platforms for harvesting accurate data for analysing the CPS. The information from these sets of data is therefore deployed for monitoring the CPS as part of CI at control centres.

Chapter 1:
Introduction: What are Cyber-Physical Systems?
Case Studies and Examples

Nikolas Flourentzou and Irina Ciornei
University of Cyprus

1.1. Cyber-physical systems

Cyber Physical Systems (CPS) are an emerging new category of interconnected real-time complex but controllable technical systems that incorporate both discrete symbolic computation and continuous physical processes in a sophisticated manner [1]. CPS widely adopts a combination of cyber (to be read as information and communication environment) and hardware infrastructure embedded with natural elements to sense, control, and actuate physical world behaviour, as well as human decision interaction. The innovation of the technologies in the cyber part is the foundation for intelligent and reliable behaviours of the whole system. Advances in CPS enable several features such as adaptability, scalability and resiliency on top of other traditional properties recognized for critical infrastructures such as safety, security, usability and capability, thus expanding the horizons of these critical systems.

Novel technologies significantly contribute to the way people interact with engineering systems. These interactions reflect the cyber-physical dimension of modern system analysis [2]. Information and communications technology (ICT) networks and computer systems are supporting the management and supervision of multitude of physical assets of critical importance due to their role in critical infrastructures such as large and interconnected power and energy systems, transportation and telecommunication networks, and water distribution systems to civil healthcare chains or modern industry production processes.

1.2. Thematic areas of CPS

CPSs can be classified according to their thematic areas. There are several areas of CPSs, but the most essential are the following.

1.2.1. Communications

Communication Systems is the essential tool for every CPS. It offers the capabilities to transfer signals for monitoring and control any physical system either from specific remote locations or from any location that is connected to the network. Due to these characteristics, communication systems are recognised as vulnerable because they are susceptible to cyber threats. Therefore, high security standards are required to minimise the risk of cyber-attacks. Risk assessments should be run frequently to define any awareness of password weaknesses of individual accounts, malware protection, software patching, secure configuration firewalls and internet gateways, among other security issues.

1.2.2. Power systems

Power System is the main part of the largest Critical Infrastructure (CI) – Electrical

Power and Energy System. Modern power systems consist of numerous CPSs that provide significant functionalities to the conventional power systems. These CPSs are additions on the existing power systems to give capabilities for the following characteristics:

- reducing pollution such as greenhouse gas emissions and wastes;
- adopting Internet of Things equipment;
- advancing smart metering and intelligent appliances;
- integrating renewable energy sources;
- improving energy efficiency (reducing losses);
- enhancing system reliability.

Although conventional power systems were controlled by Supervisory Control and Data Acquisition (SCADA) systems, these previous generation SCADA systems had limited capabilities in satisfying the current requirements for integrating ad-hoc or distributed power generation because of their highly centralized and isolated architecture. The next generation SCADA system looks into distributed and highly interconnected architecture.

1.2.3. Transportation systems

Transportation systems are also major CI. They are often considered as three CIs (for land, water, and air) and include categories according to what they transfer, such as public transport, private vehicles, freight (or other goods and animal transference), unmanned vehicles, etc.

With the increased electrification of this critical infrastructure, especially expected to grow for domestic road vehicles, or public transportation within cities, and their transition towards autonomy, an increased interdependency between power, transportation and telecommunication networks is anticipated. This increase comes with several challenges and vulnerabilities that need to be addressed from highly technical ones, such as interoperability between several independent actors and their cybersecurity level, to regulation and privacy concerns.

1.2.4. Water systems

Water networks are CIs that consist of two main categories – the clean water supply system and the drainage basin system. It requires several CPSs to be functional (such as water purification apparatuses, pressurising apparatuses) and pipe networks for

transferring water and, of course, connections to the sewers and storage reservoirs. All these systems of the water CI require a SCADA for operational monitoring and control.

1.2.5. Industry

The main objective of the fourth industrial revolution is to interact with every related sector, from the idea until the recycle of each product. As a result, each manufacturer is considered as an interconnected CPS.

1.3. Modelling and simulation

1.3.1. Case studies

The implementation of cyber-physical systems awareness is an essential tool for every simulation platform that has been developed for modelling and analysis of the combined impact (physical impact and cyber impact) of Critical Infrastructure Protection and Critical Infrastructure Resilience [3]. Within the book several real-life examples and case studies of cyber-physical systems will be detailed, such as (a) anomaly detection and prevention based on Big Data analysis techniques in metallurgy industry; (b) next-generation SCADA and Wide Area Monitoring and Control (WAMC) in large interconnected power systems, such as the Intercontinental European Power System; and (d) traffic analysis in road-transportation cyber-physical systems based on advanced image processing and video analysis techniques.

1.3.2. Cyber security

The Internet is the key medium which enhances the way data could be transformed into useful information beyond computers and mobile phones, ultimately leading to the creation of 'smart' cyber-physical systems. However, this also comes with a cost due to extended threats especially targeting high impact cost critical infrastructures. Specifically, remote exploitation of a whole host of new technologies installed or connected across the entire chain of several interconnected cyber-physical systems, with plethora of interaction actors, interactions of which among them and with the associated systems are facilitated by internet communication links, are also potentially vulnerable interferences [4]. Cyber security is critical to combat cyber threats. To realise this vision, the following objectives need to be achieved:

- **DEFEND** the CPS need to be safeguarded against evolving cyber threats. To achieve this goal, the CPS needs to effectively respond and mitigate incidents in such a way that both data and networks remain protected. The DEFEND goal, also refers to the ability and knowledge of the citizens and businesses to stand and remain resilient in case of threats.

- **DETER** refers to the aim of the CPS to detect, investigate, understand and

advise actions to disrupt hostile action taken against the system and society. This goal is a challenge in itself due to its related open question, such as what would be the rightful offensive action in cyberspace and what international body could regulate and supervise such action in case a party chooses to do so.

- **DEVELOP** refers to the growing innovation coming from cyber security industry supported by many world-leading scientific research and development institutes. This objective of the CPS is also sustained by the continuous transformation of the required skills for the personnel across the public and private sectors to adapt to a broad and integrated vision of system analysis and to overcome future threats and challenges.

1.4. Scope and objectives

The scope of this electronic book is to provide an integrated vision of cyber-physical systems with easy-to-follow examples addressed to both students and lecturers. The examples are intended to show real-world applications of either the domain specific critical infrastructure or as a general approach for systems, where the interdependencies between two or more critical infrastructures are emphasized.

The rest of the book is organized as follows.

Chapter 2 highlights the revolution of digitization, which extends the concept of cyber-physical systems in several application domains. The focus is on the concept of Big Data and the need for innovative techniques for fast and distributed data processing to create meaning out of this data. The chapter presents examples from the metallurgy industry.

Chapter 3 introduces the need for cyber-security assessment in this new digitized world. First, it defines the concept and frames its application. Then, it offers several classifications of the possible cyber threats and tools for detection, analysis and mitigation of the cyber threats.

Chapter 4 is dedicated to the power cyber-physical system in the era of Smart Grids (SG). It shows the application and architecture of SCADA as a predecessor of modern cyber-physical systems. Then, it goes into the details of modelling and analysis of wide area monitoring and control (WAMC) of large interconnected power systems with an emphasis on interdependences between the communication and the power CI.

Chapter 5 addresses the application domain of road transportation as a cyber-physical system. The focus of this chapter is on simulation and modelling techniques for road traffic analysis based on real-time image and video processing.

Chapter 6 is a comprehensive survey on the communication environments suitable for a broad range of cyber-physical systems. It analyses several communication protocols currently in place for the functionality and capability of several cyber-physical system components such as sensors, actuators, and controllers and shows how they might vary according to the demand of applications.

Chapter 7 presents an overview on embedded systems. It starts with the definition of the concept and explains the relation between embedded systems and the broad class of cyber-physical systems. Then, it surveys the most common architectures and principles for their realization. Several hands-on working-examples for the design and implementation of specific embedded systems are also provided in this chapter.

Chapter 8 dives into the specific design constraints of the communication layer in the case of the special class of cyber-physical systems such as the Internet of Things (IoT). Energy harvesting solution and design principles of such systems are detailed and exemplified in this chapter. The IoT serves a broad class of industry applications as well as all sorts of Cis, especially their edge side. Finally, in this chapter application of smart parging is exemplified.

Chapter 9 concludes the book. This chapter is dedicated to the most recent state of the art in hardware and software networks for cyber-physical systems, addressing the need for highly available, reliable and efficient use of shared communication resources (e.g., in the case of IoT, Big Data and wide area critical infrastructure networks).

Chapter 2:

Big Data and the Need for Data Processing

Natalia Morkun
Kriviy Rih National University

Co-authors:

Denys Kuznietsov
Kriviy Rih National University

Iryna Zavsiehdashnia
Kriviy Rih National University

Big Data is one of the key drivers of modern IT development. This trend is evolving fast and gaining momentum because of ongoing increase of data volumes to be analysed and synthesized. Major sources of data are social networks accumulating vast amounts of data on their users, Internet of Things (IoT) networks with the data on all attached objects and other data sources. They all someway accumulate information for further analysis.

Generally, Big Data is unstructured data of various volumes that require specific approaches to their storage and processing. Unlike traditional relativistic databases that store information according to internal structure and some processing algorithms, heterogeneous objects are quite difficult to consolidate, i.e., arrange central control and search among unstructured data including media-files, data from IoT devices, documents, etc. If these unstructured isolated data are not huge, no serious problems arise. Yet, when there are tens or hundreds of terabytes of such data and a great number of analysed objects, there can occur situations that make system data and object search impossible. Interaction among objects slows down and there are complications of navigation in multiple data directories and many other problems.

It should be noted that the Big Data concept conditions three major types of tasks, namely:

- arrangement of unstructured information composed of media-files, data from IoT sensors and microcontrollers, and other data types;
- analysis of large data volumes by using specific methods or ways to process unstructured data, generation of reports, introduction and application of forecast models [5];
- storage and management over large volumes of data with which ordinary relativistic databases cannot deal effectively.

Nowadays, the mining and metallurgical industry is no exception in view of this. In order to remain competitive on the mining and metallurgical market, most professionals of this industry are striving for a significant economic effect in the form of reduced expenses for data processing in production, logistics and management due to Big Data approaches and technologies.

Modern mining and metallurgical production are characterized by available complex technological equipment located at deposits of ore and coking coal, power engineering facilities, etc. The mining and metallurgical industry includes the following elements:

- by-product coke plants or shops for coal pre-treatment, coking and extraction

of useful chemical elements;

- blast-furnace shops for smelting ferroalloys and pig iron;
- underground mines and open pits for ore and coal mining;
- steel shops;
- rolling mill shops.

Automated control systems (ACSs) for metallurgical plants continually are creating data on the following processes:

- technological processes (TP ACS);
- logistic processes (the ACS of transport logistics);
- control processes (MES and ERP systems).

It is worth noting that the TP ACSs collect data from aggregate sensors on states and modes of technological processes. Control systems can provide data in the form of media-files on conditions of rolled bars, their defects, and ultrasonic control maps. The ACS of transport logistics contains information on vehicle and material movement, while the MES and the ERP control data on planning, orders, operating control over materials, order status and storage reserves.

In its turn, equipping production with up-to-date automation systems results in digitalization of all the received technological data. Here arises a problem of the data being difficult to access or sometimes inaccessible. It should be mentioned that while analysing an industrial situation like searching for a cause of defects, their dependencies and solutions to optimization problems, reports of completely heterogeneous systems should be collected, generated, and analysed. Accordingly, analysis of report data is extremely time-consuming, this resulting in the problem losing its topicality. The amount of obtained data from various sensors can reach tens of petabytes of data. The whole volume of data should be processed and analysed. Currently, available methods of data processing and analysis are unable to analyse and synthesize such huge data volumes. Application of Big Data approaches and technologies can be a solution to these problems.

Modern information technologies offer a set of approaches, tools, and methods of processing both structured and unstructured big data. There are three Vs used as characteristics of these approaches [5]:

- volume;
- velocity of processing and obtainment of results;
- variety of types of structured and unstructured data.

The characteristic data can be applicable to structured and unstructured data from mining and metallurgical production, including:

- multiple signals from control sensors;
- UT charts;
- images of rolled strips with defects;
- data on products and materials movements;
- data on orders and suppliers.

The Big Data technology enables combining data from the TP ACS, the ACS of transport logistics and the MES and the ERP systems into a single unit, spending much less time as compared with traditional approaches. Time saving will cause a proportional economic effect, thus making application of Big Data approaches profitable for mining and metallurgy. As a result, future Big Data use will enable the industry to perform predictive analysis of a certain portion of industrial wastage in mining and metallurgy on the basis of machine learning.

2.1. Dimension reduction of nonlinear dynamic process models in Big Data

2.1.1. Problems of Big Data analysis in the mining and metallurgical industry

As it was mentioned earlier, analysis and synthesis of big data generates certain problems for their processing, this being irrelevant for conventional data sets. These peculiarities cause significant problems for data analysis and therefore call for development and application of new statistic methods and approaches.

Unlike conventional data sets with the sample size greater than dimension, Big Data is characterized by massive or enormous sample sizes and big dimensions. These are typical problems associated with high dimensions:

Heterogeneity. Combining many data sources is the basis for creating Big Data that can belong to different subgroups. Any of these subgroups is able to reveal its

own unique properties which are not common for others. In standard conditions, for a small or average sample size, data points for small subgroups can be classified as deviations and they are difficult to model consistently because of the insufficient number of observations. Yet, this disadvantage can be treated as an advantage of better understanding of the heterogeneity of data using huge sample sizes. For instance, the model result of ore concentration mixture for big data sets calls for applying complex statistic and calculation methods. In small dimensions, we can apply such standard methods as the expectation-maximization algorithm for ore mixture and raw materials models. In case of Big Data, we should thoroughly order the evaluation procedure to avoid either overtraining or accumulation of noise and elaborate better calculation algorithms.

Noise accumulation. In case of Big Data analysis, simultaneous evaluation and verification of multiple parameters is required. Accumulation of estimation errors occurs when a forecast rule or solution or principle depends on a great many parameters like that. Noise accumulation process is an especially serious problem in large dimensions and can prevail over true signals. For instance, in multidimensional classification, the 'bad' classification result is determined by multiple weak points which do not reduce classification errors. It should be mentioned that choice of variables is of primary importance for the implementation of forcing the accumulation and increasing noise in the process of classification and regression forecasting. Yet, the choice of variables in big dimensionality is a complicated problem to be solved because of false correlation, random endogeneity, heterogeneity, noise accumulation and measurement errors.

False correlation. Availability of high dimension can contain false correlation. This may be due to the fact that numerous nondeterministic random values can have high correlation values in huge dimensions. Using a false correlation may lead to wrong statistics and as a result – to wrong scientific discoveries. It should be also mentioned that with big dimensionality, even for simple models, choice of variables is complicated due to available false correlation. In particular, if there is high dimensionality, some significant and most important variables can be enormously correlated with some false or wrong variables, which are not related from the scientific point of view. Also, besides the choice of variables, false correlation can cause wrong statistical results.

Random endogeneity. Random endogeneity is a problem occurring because of high dimensionality. While adjusting the regression of $Y = \Sigma d_j = \mathbf{1} \cdot \beta \cdot X_j + \varepsilon \cdot Y = \Sigma j = \mathbf{1} \cdot d\beta_j \cdot X_j + \varepsilon$ type, this term suggests that the residual noise $\boldsymbol{\varepsilon}$ can correlate with some predictors $\{X_j\}$. Unlike false correlation, accidental endogeneity can lead to process of current correlations between random variables. In the most general sense, endogeneity results from sampling bias, making inappropriate measurements and using non-response or false variables. This type of situation frequently arises in Big Data analysis mostly for two reasons:

- Due to new highly efficient methods of measurements, scientists can collect as many functions as possible and strive for that. In turn, this affects the possibility of some functions being correlated with partial or residual noise.

- As a rule, Big Data typically combines several sources that use different schemes or approaches of data generation. This may lead to an increase of the probability of systematic or permanent deviation of the data sample and as a result, it can lead to measurement errors. These results can lead to accidental endogeneity.

Thus, we can conclude that the problems associated with dimensionality reduction are quite important in Big Data analysis and synthesis.

2.1.2. Reducing the dimensionality of models of nonlinear dynamic processes of iron ore beneficiation

Relevant conclusions on quality of a technological process can be drawn only on the basis of a great amount of data on extraction of feed materials various in composition and sizes in some specific conditions of beneficiation. Information on such a complicated system of particles as slurry can be presented as a set of fraction extracts or differential numbers for each elementary category (class) of particles with the known parameter of separation and sizes. The larger are the classes of material particles fed to concentrate, the more complete will be the calculation data on behaviour of this material in the separation zone for a technologist or a designer. The obtained data can allow controlling beneficiation by changing the material feed rate (productivity), magnetic force (magnetic separation), the reagent mode (flotation), etc. in case of the set volume of extracting useful component particles into concentrate [6], [7].

Thus, information on the useful component content within the whole range of the granulometric characteristic of the processed ore may result in the most accurate assessment of technological units' performance. As this characteristic is a nonlinear function with a great number of input and output parameters, the model of these objects possesses high dimensionality, this making it difficult to analyse and apply to forming control over a technological process.

Dimensionality reduction is an effective tool for dealing with multidimensional data. Reduction is aimed not only at reducing calculation complexity, but also at structuring the obtained information and distinguishing major components of the process under study.

Being controlled objects, technological aggregates of beneficiation can be presented as some operators that transform vectors of input variables into those of output parameters [8], [9], considering controlling actions:

$$\overline{Y} = F(\overline{X}, \overline{U}), \tag{2.1}$$

where F is the operator representing a technological line of ore beneficiation; X is a state vector; and U is the vector of controlling actions.

Output parameters of separate stages of ore beneficiation (e.g. grinding) are to be considered as input ones for the subsequent stage:

$$\begin{aligned}
&\overline{X_1} = F_1(\overline{X_0}, \overline{U_1}),\\
&\overline{X_\ell} = F_\ell(\overline{X_\ell}, \overline{U_\ell}),\\
&\qquad \dots,\\
&\overline{Y} = \overline{X_L} = F_L(\overline{X_{L-1}}, \overline{U_L}),
\end{aligned} \tag{2.2}$$

where $F_\ell\, F_\ell$ are operators representing separate technological aggregates, stages or cycles of the ore beneficiation line $\ell = \overline{1, L}$; L is the number of aggregates (stages, cycles).

Elements of output parameter vectors $\overline{X_1} \dots \overline{X_L}$ of technological aggregates are their qualitative and quantitative indices. At the ℓ-th stage of the technological process, $\ell = \overline{1, L}$, there occur changes of characteristic functions of the set physical property including the mass fraction of particles $\gamma_\ell(\xi)$ in total weight and content of the useful component contained there – $\beta_\ell(\xi)$. For example, size functions $\xi \equiv d$ characterize granulometric composition of ore material particles $\gamma(d)$ and the useful component content in classes $\beta(d)$. Characteristics of the fed materials of the beneficiation line are denoted by index 0: $\gamma_0(d)$, $\beta_0(d)$, while those of the output product will have the index of the last technological operation $\ell = L$, $\gamma_\ell(d)$, $\beta_\ell(d)$.

$$\overline{X_\ell} = \{\gamma_\ell(d), \beta_\ell(d), \chi_\ell\}, \ell = \overline{1, L}, \tag{2.3}$$

where γ is the mass fraction of sizes, %; β is the mass fraction of the useful component in size classes, %; χ is specific power consumption of a given technological aggregate, kWh/t; d is the size of ore particles, mm; ℓ is the index of a technological operation; L is the number of technological operations.

Thus, the vector of signals from the iron ore beneficiation line (2.3) at moment t that in a general form can be presented like this:

$$\overline{X}(t) = (x^{(1)}(t), x^{(2)}(t), \dots, x^{(p)}(t))^T, p = \overline{1, P}, \tag{2.4}$$

should be transformed into vector $\overline{R}(t) = (r^{(1)}(t), r^{(2)}(t), \dots, r^{(p)}(t))^T, p = \overline{1, P'}$, where $P' << P$, in other words, into a vector of much smaller dimensionality composed of the most informative variables of the input vector.

Simultaneously, the elements of vector $\overline{R}(t)$ are determined by a set of input

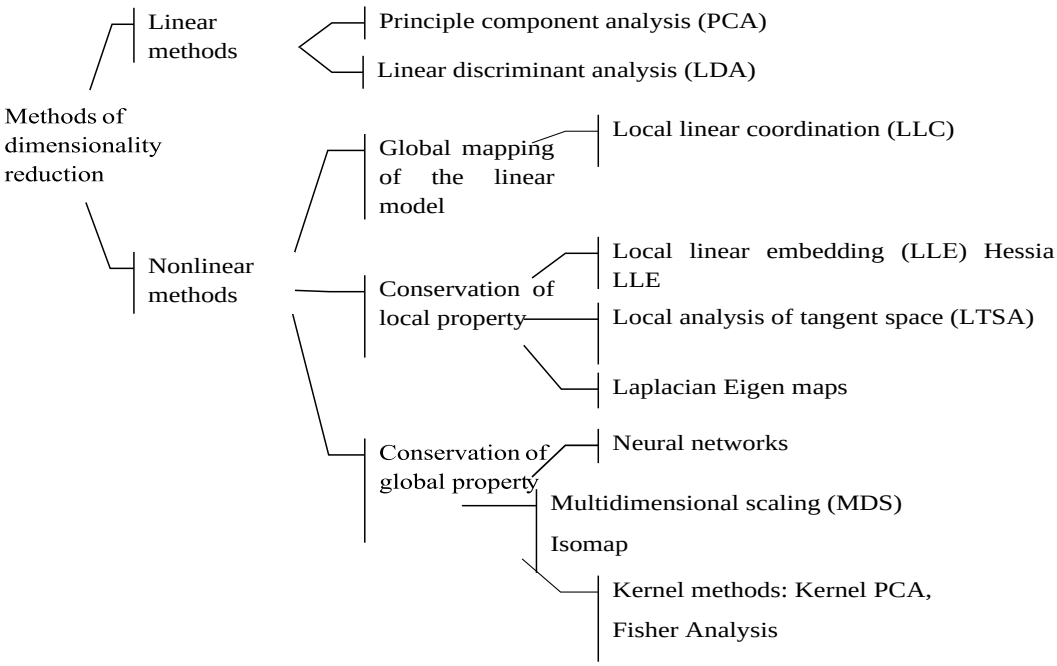


Fig. 2.1. Classification of dimensionality reduction methods

properties $\overline{X}(t)$, for example, by their linear combinations. The ratio between vectors $\overline{R}(t)$ and $\overline{X}(t)$ is presented as [9]:

$$\overline{R}(t) = \Lambda\{\overline{X}(t)\}, \tag{2.5}$$

where Λ is the operator of conversion to the smaller dimensionality space. Application of this approach will enable enhanced efficiency of identifying distributed processes of ore beneficiation. While forming control actions, one should perform inverse conversion to the input dimensionality space

$$\overline{X}(t) = \Lambda^{-1}\{\overline{R}(t)\}, \tag{2.6}$$

where Λ^{-1} is the operator of conversion to the initial dimensionality space.

From the viewpoint of the character of output data, methods of dimensionality reduction can be divided into two categories: linear and nonlinear. From the viewpoint of the degree of retaining the geometric structure of reduction, they contain a local and a global approach (Fig. 2.1).

Manifold learning methods can be divided into two categories [10]: global mapping of the linear model and non-linear method of conservation of the global

Table 2.1. Properties of Dimensionality Reduction Methods

#	Dimensionality reduction method	Parameterization	Free parameters	Calculation complexity	RAM volume
1	PCA	yes	absent	O(D3)	O(D2)
2	MDS	no	K	O((nk)3)	O((nk)3)
3	Isomap	No	K	O(n3)	O(n2)
4	LLE	No	K	O(pn2)	O(pn2)
5	Hessian LLE	No	K	O(pn2)	O(pn2)
6	Laplacian EM	No	k, σ	O(pn2)	O(pn2)
7	Diffusion maps	No	σ, t	O(n3)	O(n2)
8	LTSA	No	K	O(pn2)	O(pn2)

property. Main differences of the global and local methods are defined by the local structure.

Table 2.1 contains basic characteristics of analysed dimensionality reduction methods: the parametric character of mapping between multi- and small-dimensional spaces; free parameters to be optimized; calculation complexity of the algorithm; and the required volume of random-access memory (RAM) [10].

Analysis of the properties given in Table 2.1 reveals the following:

1. Most methods of dimensionality reduction are not parametric – do not define direct mapping from the multidimensional space into the small-dimensional one (or vice versa).

2. The functions of most nonlinear dimensionality reduction methods possess free parameters to be optimized – those that directly affect the value function (authenticity losses) which is optimized. Non-convex methods of dimension reduction have additional free parameters like the learning rate and the maximum permissible number of iterations. The LLE method uses the regularization parameter of weight reconstruction.

- Calculation complexity of the dimensionality reduction algorithm is of primary significance for its applicability. Calculation complexity is determined by the properties of data sets: the number of points of input data n and their dimensionality D.

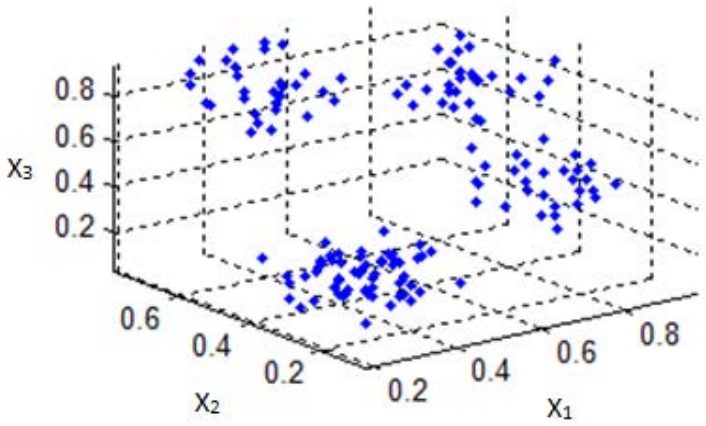


Fig. 2.2. The results of input data clustering

Parameters of the method are: the number of the closest neighbours κ (for graph-based methods); the number of iterations I (for iteration methods); the ratio of non-zero elements in the scattered matrix to the total number of elements p; the number m of local models of factor analysers in LLC.

Let us consider the results of the research of basic approaches to dimension reduction taking the three-dimensional set of standardized characteristics of ore materials as an example (Fig. 2.2).

To compare efficiency of various dimensionality reduction methods, we use data on magnetic separation of ground particles in iron ore slurry. Each mineral and technological type of slurry is characterized by several physical-mechanical and chemical-mineralogical parameters. A data set is a totality of fraction extracts or differential numbers for each elementary category of particles (a size class) with a known parameter of separation and size.

Application of principle components analysis (PCA) results in presentation of the smaller dimensionality data which describes the trend of the greatest changes in input data [9]. The obtained linear conversion of T maximizes the expression

$$F = T^T \, cov_{X-\overline{X}} \, T \to max, \tag{2.7}$$

where $cov_{X-\overline{X}}$ is the covariance matrix of X data centred as to the origin of coordinates.

Application of the PCA method enables a correct mapping of input data in the smaller dimensionality space for 1.04–1.15 seconds. The result of projecting the input data onto the first two major components (k_1, k_2) by means of PCA is presented in Fig. 2.3.

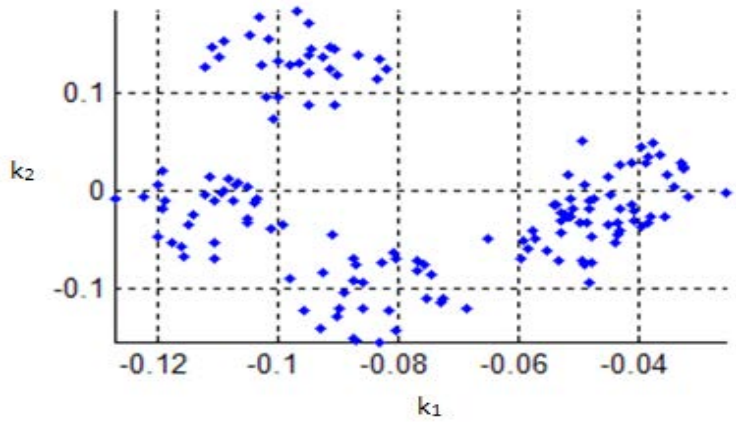


Fig. 2.3. The result of dimensionality reduction by means of the PCA method

The non-linear method of multidimensional scaling (MDS) [11], [12] in mapping the multidimensional data view into the smaller dimensionality space conserves pairwise distances between data points. The function which evaluates the difference of pairwise distances in the initial multidimensional view and the obtained view of the smaller dimensionality describes the mapping quality. The voltage function is an example of a function of this type.

$$F(Y) = \sum_{ij}(\|x_i - x_j\| - \|y_i - y_j\|)^2, \tag{2.8}$$

where $\|x_i - x_j\|$ is the Euclidean distance between data points of the bigger dimensionality; $\|y_i - y_j\|$ is the Euclidean distance between data points of the smaller dimensionality.

An alternative of the considered function is the cost function of Sammon accentuating conservation of smaller distances first:

$$F(Y) = \frac{1}{\sum_{ij}\|x_i - x_j\|}\sum_{ij}\frac{(\|x_i - x_j\| - \|y_i - y_j\|)^2}{\|x_i - x_j\|}. \tag{2.9}$$

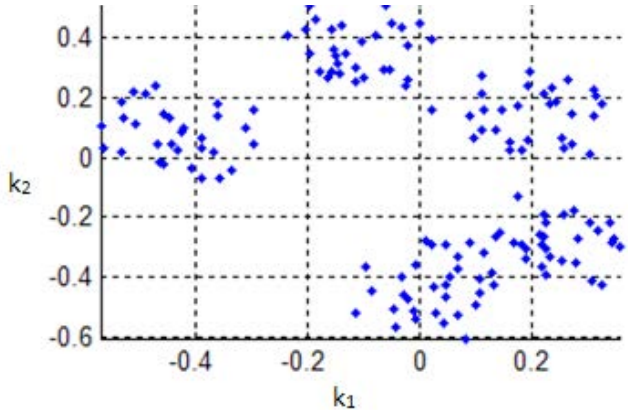


Fig. 2.4. The result of dimensionality reduction by means of the MDS method

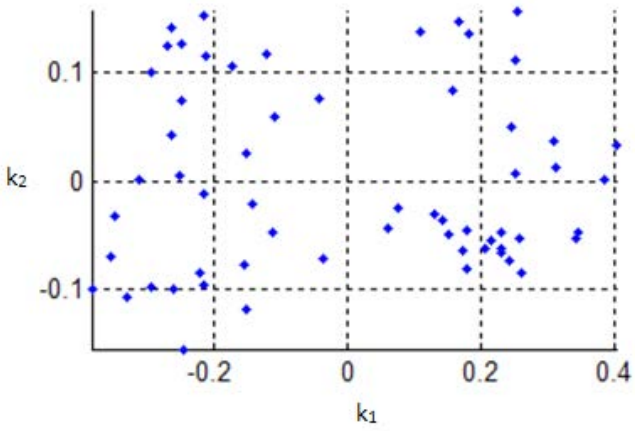


Fig. 2.5. The result of dimensionality reduction by means of Isomap

The voltage function is reduced by using the conjugate gradient method [11]. This method enables correct mapping of input data into the smaller dimensionality space within 21.07–22.86 seconds. The result of projecting input data onto the first major components (k_1, k_2) by applying the MDS method is presented in Fig. 2.4.

This method does not consider distribution of conjugate points, as it is grounded on the Euclidean distances, this being a drawback of the multidimensional scaling method [10]. For instance, if multidimensional data are of the curvilinear multiform, the distance between them can be much greater that the Euclidean one. In this case, we should apply the Isomap method that considers the curvilinear distance between data points. In Isomap, geodesic distances between x_i of data points are calculated by building a graph in which each xi point is associated with its k closest neighbours x_{ij} in the X data set. The shortest distance between points of the graph evaluating the curvilinear distance between these two points is determined by the Dijkstra algorithm [13]. The results of projecting input data onto the first two components (k_1, k_2) by using the Isomap method are presented in Fig. 2.5.

The diffusion MAP method is based on analysing the Markov random transposition on the data space. Under these circumstances, the proximity degree of point data is determined for time steps. By using this procedure, the diffusion distance is determined. In the smaller dimensionality space, the obtained diffusion distances are conserved.

The first stage of the diffusion map method (DM) [14] includes the formation of the data graph. The graph edge weights are calculated by using Gaussian kernel functions resulting in forming matrix W. Then, on the basis of matrix W, the standardized matrix P(1) is calculated:

$$p_{ij}^{(1)} = \frac{w_{ij}}{\sum_k w_{ik}}. \tag{2.10}$$

The obtained matrix $P^{(1)}$ is considered a stochastic matrix that conditions the matrix of forward transition probability for dynamic processes. Thus, the $P^{(1)}$ matrix represents probability of transition from one data point to another per time unit. The forward transition matrix for the t-th time unit $P^{(t)}$ is determined by $(P^{(1)})^t$. On the basis of transition probabilities $p_{ij}^{(t)}$, diffusion distance is defined.

$$D^{(t)}(x_i, x_j) = \sum_k \frac{(p_{ik}^{(t)} - p_{jk}^{(t)})^2}{\psi(x_k)^{(0)}}, \quad (2.11)$$

where $\psi(x_i)^{(0)} = \frac{m_i}{\sum_j m_j}$ is the factor providing more weight to graph elements with increased density; and $m_i = \Sigma_j p_{ij}$ is the node degree.

The given equation reveals that pairs of points with greater transition probability have smaller diffusion distances. The diffusion distance implies that many graph paths provide better noise resistance than that of the geodesic distance. Views of a smaller dimensionality Y that conserve diffusion distances are formed from d nonzero major eigen vectors calculated from

$$P(t) \cdot Y = \lambda \cdot Y. \quad (2.12)$$

As the graph is fully-connected, the maximum eigen value is zero, i.e., $\lambda \cdot 1=1$, and its eigen vector is ignored. In the smaller dimensionality view, eigen vectors are standardized by their own values:

$$Y = \{\lambda_2 \cdot v_2, \lambda_3 \cdot v_3, \dots, \lambda_d + \mathbf{1} \cdot v_d + \mathbf{1}\}. \quad (2.13)$$

The projection result of input data for the first two principle components (k_1, k_2) by using the DM method is given in Fig. 2.6. The quality of the obtained data of

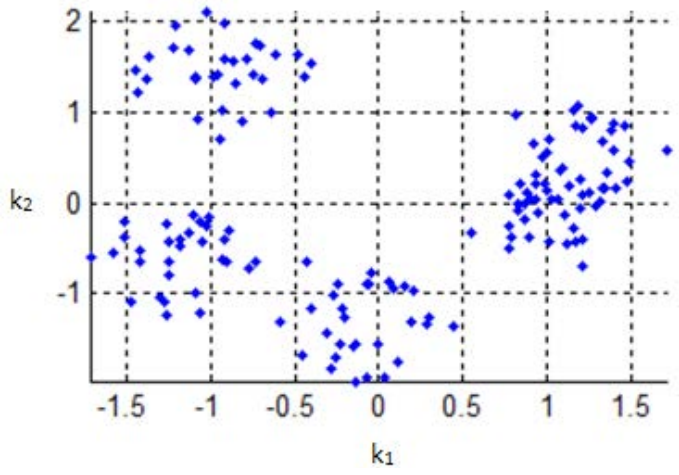


Fig. 2.6. The result of dimensionality reduction by means of the diffusion map method

smaller dimensionality is compared by assessing the degree of conserving the local data structure in the following two ways [15]:

- by assessing the classification error of the *k*-nearest neighbour (k-Nn) for the classifiers learning from low-dimensional data views; the data were classified in the low-dimensional space with $k = 12$;

- by assessing the adequacy and continuity of smaller dimensionality views.

The adequacy degree is determined as follows:

$$T(k) = 1 - \frac{2}{nk(2n-3k-1)} \sum_{i=1}^{n} \sum_{j \in U_i^{(k)}} (r(i, j-k)), \tag{2.14}$$

where *r(i, j)* represents the rank of the data point J in the low-dimensional space according to calculated pairwise distances; and variable U indicates the set of points that are the K nearest neighbours in the low-dimensional space.

The continuity measure is determined as follows:

$$T(k) = 1 - \frac{2}{nk(2n-3k-1)} \sum_{i=1}^{n} \sum_{j \in U_i^{(k)}} (r(i, j-k)), \tag{2.15}$$

where *r(i, j)* is the rank of the data point J in the high dimensionality space according to the calculated pairwise distances between the data points of high dimensionality; and variable V indicates the set of points adjacent to that of K-closest neighbours in the high dimensionality space.

Errors of generalization, probability, and continuity enable assessment of the conservation degree of the local data structure. Quality assessment based on generalization, probability and continuity errors is preferable to assessing the reconstruction error, as to visualize or classify data well enough, their local structure should be conserved.

Table 2.2 shows generalization errors of k-Nn classifiers and relevant adequacy values obtained by applying various methods of dimensionality reduction.

The given results reveal that space reduction by nonlinear methods appears to be most efficient for reducing dimensionality of data on iron ore beneficiation. In this case, consideration of space topology of ground particle distribution on the basis of physical-mechanical and chemical-mineralogical properties enables more precise specification of beneficiation parameters.

Such nonlinear methods as Isomap and Laplacian Eigenmap are efficient for dimensionality reduction, yet they do not provide optimal conditions of classification after data transformation. Therefore, besides nonlinear space transformation,

Table 2.2. Analysis of Data Dimensionality Reduction Methods

#	Method	Parameter setting	Errors of generalizing classifiers	Adequacy T (12)	Continuity C (12)
1	PCA	no	5.15 %	0.95	0.92
2	MDS	$5 \le k \le 15$	4.68 %	0.82	0.96
3	Isomap	$5 \le k \le 15$	4.98 %	0.85	0.73
4	LLE	$5 \le k \le 15$	4.97 %	0.71	0.75
5	Hessian LLE	$5 \le k \le 15$	4.11 %	0.52	0.63
6	Laplacian EM	$5 \le k \le 15$ $\sigma = 1$	4.65 %	0.85	0.79
7	Diffusion maps	$10 \le t \le 100$ $\sigma = 1$	3.86 %	0.93	0.94
8	LTSA	$5 \le k \le 15$	3.62 %	0.59	0.64

we should increase the discriminatory power of properties, for instance by extra application of the Fischer discriminative analyser or principle component analysis. The results of analysing dimensionality reduction methods, by using the Sammon cost function, are provided in Table 2.3.

Among the dimensionality reduction methods considered, the diffusion map method reveals the best results with the Sammon error of $\varepsilon = 2.6$ %, $\sigma(\varepsilon) = 0.011$ and the average execution time $T = 0.127$ sec.

2.2. The concept of working with Big Data within a mining and metallurgical enterprise

2.2.1. An overview of the principles of working with Big Data within an industrial enterprise

At industrial enterprises of the mining and metallurgical industry, Big Data is generated through introducing the industrial IoT technology. During this process, basic units and parts of machines are equipped with sensors, control devices capable of doing edge (fog) computing. While an industrial process is taking place, current data are collected and processed on a regular basis. Application of analytical platforms enables online processing of Big Data. Also, analytical platforms present results in the most convenient way to perceive and store them for further use. The conducted data analysis allows concluding on the technical condition of equipment and its efficiency, as well as the necessity to introduce any changes into technological processes.

Table 2.3. Analysis of Dimensionality Reduction Methods

#	Method	Error ε	MSD error σ(ε)	Average execution time T, s
1	PCA	0.028	0.012	0.18008
2	MDS	0.027	0.011	0.90678
3	Isomap	0.084	0.014	0.65598
4	LLE	0.241	0.018	0.36894
5	Hessian LLE	0.381	0.015	0.23029
6	Laplacian EM	0.264	0.019	0.14284
7	DM	0.026	0.011	0.12676
8	LTSA	0.317	0.012	0.06873

The following basic principles and approaches of working with Big Data should be mentioned:

- **Horizontal scaling**. This approach is central to Big Data processing. As there may be any amount of data, any system applied to processing Big Data should be capable of scaling. For instance, if the data volume increases three-fold, one should increase the amount of hardware in a cluster proportionally. With that, data processing should not affect the efficiency of the process.

- **Fault-resistance**. In many cases, the Big Data concept implies application of a great many calculation nodes (tens of thousands) in a cluster. According to the horizontal scaling principle, their number will continue increasing in the long run. In its turn, there appears an increasing probability of units' failure. That is why, applied Big Data methods should consider these situations and provide some preventive measures.

- **Data locality**. As data are divided according to large quantities of calculation units, in case they are directly located and kept in a single physical server and processing takes place on another server, data transmission expenses could be unwarrantedly great. For this reason, data processing should be conducted in the cluster node in which they are conserved.

Thus, provision of fault-resistance and locality principles combined with increasing horizontal scaling of nodes is essential while working with Big Data.

Table 2.4. Comparison of Network Client-server DBMSs

DBMS								
No.	Criteria		MySQL	Firebird	InterBase	MS SQL Server	Oracle	IBM DB2
1.	OC	Windows	√	√	√	√	√	√
2.		Linux	√	√	√		√	√
3.		BSD	√	√				
4.	Information reliability and security level		Medium	Medium	Medium	Very high	High	High
5.	Capabilities of SQL query language and DB tools		Weak	Medium	Very weak	Powerful	Powerful	Powerful
6.	Hardware system requirements		Medium/ Low	Low	Medium /Low	Medium/ High	Very high	Very high
7.	Operation and administration complexity		Medium	Low	Low	High	High	Very high
8.	Rates of development, improvements, potential		Fast	Slow	Very slow	Fast	Fast	Slow
9.	License cost, 000 $		Free	Free	6.5	35	30–80	>50

2.2.2. Big Data clusters

Selection of the technological Database Management System (DBMS) for the decision-support system should be based not only on technical characteristics, reliability, efficiency, and fault-resistance of the database (DB) but also on development rates, documentation completeness, and quality. At present, the market offers quite a great variety of DBMSs (e.g., IBM DB2, MySQL, MS SQL Server, Oracle, InterBase, Informix, Firebird) with a large proportion of free products [16]. Comparison of these DBMSs is presented with main parameters in Table 2.4.

The detailed analysis of advantages and drawbacks of the above-mentioned DBMSs shows that the MS SQL Server is the most industrially practical. Being cheaper than Oracle and IBM DB2, it provides a high level of data security, complete implementation of capabilities of the SQL query language and service support of Microsoft. Longstanding experience of applying MSSQL Server 2012 at mining enterprises of Kryvyi Rih, in particular at the Northern GZK, supports the decision made. The system efficiently operates on a four-way server with 16 GB RAM, while the processors possess 4 cores each. The DB consists of 100–120 tables, the main fields of which are 8-byte-sized real numbers. Within 1 minute, MS SQL Server processes over 10000 complex requests to various data tables, and the peak load on the server station does not exceed 70 %. An average workload on processors is about 15–20 %. Control system sensors can measure some process flow parameters every second that results in a rather intense data flow. Such discreteness can result in increase of the DB size at the rate of 16 GB/year.

Figure 2.7 presents a typical model of integration of the intellectual decision support system based on Big Data principles into the information structure of an enterprise.

An information-processing block comprises a step-down transformer, an ADC and an Ethernet data transmission device. The DB server stores all the necessary information on each electric motor and intermediate data.

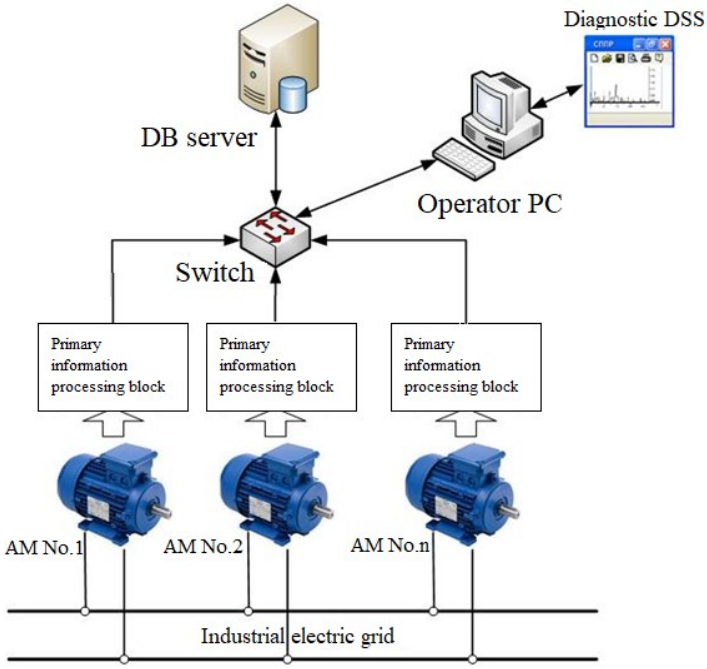


Fig. 2.7. DSS integration into the typical information structure of an enterprise

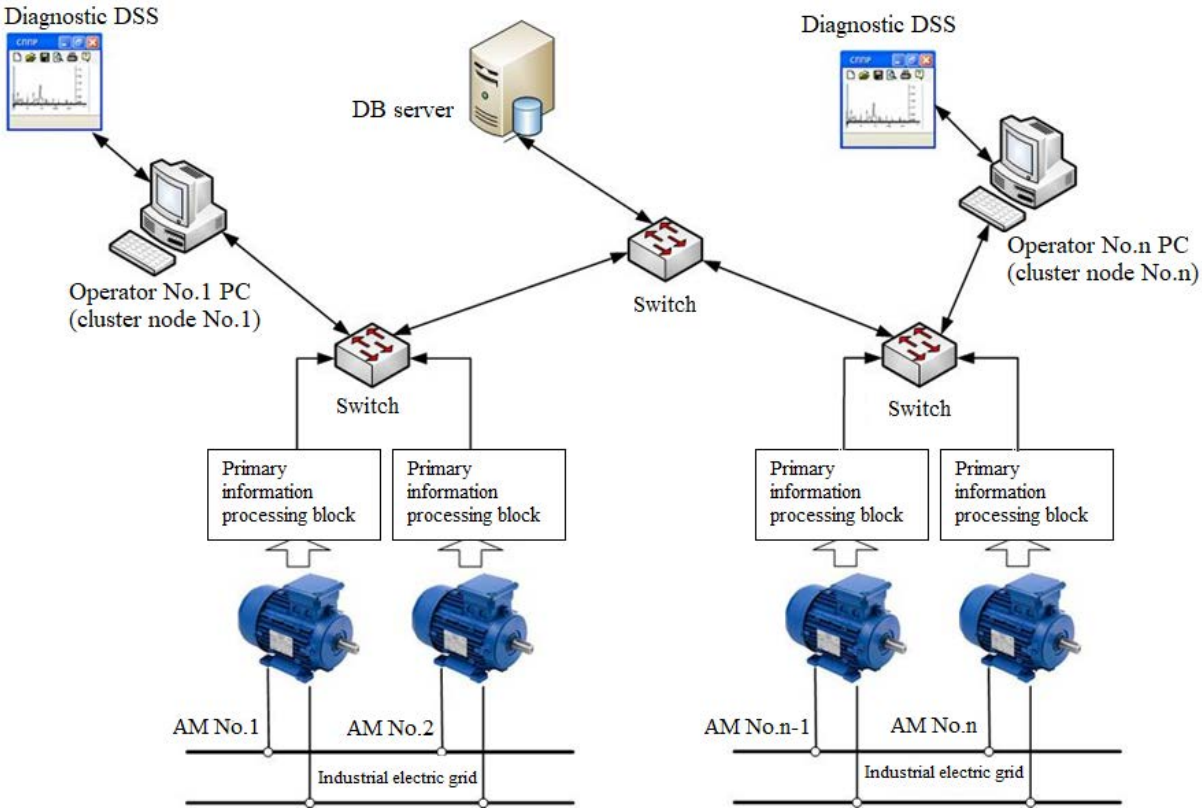


Fig. 2.8. Logic-functional scheme of cluster DSS implementation and integration into the most common and typical information structure of a factory or enterprise

In the case of applying a decision support system in the form of a cluster structure, a typical and most common model of integration of the intellectual DSS into the information structure of an enterprise will appear as shown in Fig. 2.8.

In case of using the cluster DSS and MS SQL Server, Kubernetes can be used. It should be noted that Kubernetes is an open-source system for managing the container clusters of Unix/Linux as a unified single system. Docker containers can be started and configured using Kubernetes. In turn, Kubernetes allows the management and run of containers through host clusters. Kubernetes also provides co-location and support for increased reliability due to the replication of a large number of containers. This approach was created and supported by Google, but it is also is supported by many companies, including Amazon, Docker, Ubuntu, IBM, and Microsoft.

2.3. Construction features of a modular cyber-physical system for preventive diagnostics of mining and metallurgical power equipment

2.3.1. Basic information

At present, electric drives are main operating power units at modern enterprises. In most cases, their basic elements are polyphase asynchronous electromotors (AMs). The fact that electric motors of this type consume about 40 % of the world power proves that they are the most widely used [17], [18]. Their main feature is high failure rates, in particular, their average performance without overhauls makes 10–15 years. Delayed detection of emergency modes of AM results in interrupted technological processes, defective products, electromotor restoration and repair costs, increased power consumption, accidental breakdown, etc. Improving the reliability of equipment is of great importance for the mining and metallurgical industry, and random equipment with electric motor failures may practically suspend a significant part of the production [19].

Modern tools and methods of electric equipment diagnostics are primarily based on the application of various hardware technological sensors directly bound to a technological (production factory) object (i.e., direct methods) most of which are used during scheduled maintenance and routine diagnostics [20]. It should be noted that the authors understand that scheduled diagnostics and maintenance of energy equipment can lead to partial or complete stoppage of the production or to a decrease in some of the main technological processes in which energy equipment under analysis is involved. Preventive or routine diagnostics means monitoring of the object under study without stopping the technological process in which it is involved. The main advantage of the preventive diagnostics over the planned period consists of a short time (several minutes) to obtain results. This type of diagnostics helps prevent emergency halt modes or situations of the total stop of industrial (factory) and household energy (with electric motor) equipment performance. Application of modern methods, approaches, and information technologies to process monitoring and analyse the current technical parameters and technical state of energy equipment (electric motors) is a highly topical task, because it enables supporting and realizing modern requirements of interacting, monitoring and analysing operation of electric plants, pumps, mills, conveyors, robotic systems, which are part of the Industry 4.0 concept [21], [22].

The modern information infrastructure of enterprises provides equipment, automation and control processes as well as the interconnection of integrated information subsystems and some kind of equipment, in particular microcontrollers, smart sensors, etc. [23]. Control and monitoring of industrial and household equipment through the global IoT network, which is a promising trend of Industry 4.0, are widely practised nowadays [24]. Here, devices are integrated on the basis of specialized control and data processing centres. Thus, Ericsson Mobility Report [25]

mentions about 16 Bn Internet-connected devices in the world. In 2020, the number of such devices makes nearly 29 Bn with 18 Bn belonging to the IoT network.

Modern information systems and technologies based on IoT approaches should contain and process constant information exchange between enterprise devices without human involvement, thus enabling automated accumulation and analysis of information. IoT allows to create an information system with the ability to self-organize and adapt to current conditions both at the enterprise and at the global level. That means that in regard to monitoring, diagnostics and preventive assessment of the current state of electric motors (energy equipment), it is possible to develop, implement and use a self-learning information system capable of adaptation and self-diagnosis.

A typical and important problem of development of modern information systems for monitoring, diagnostics, and preventive assessment of the current state of electric motors (energy equipment) involves the elaboration of adaptation, self-diagnosing and self-learning technical systems included into the global IoT network. This elaboration also includes the development of modern methods, approaches ,and information technologies for developing, integration, and using elements of Smart Boxes and Smart Apps. The development of a special access network implies the development and use of a platform and information technologies for the IoT monitoring, their preventive assessment, and as a result leads to the control of Smart Box and Smart App.

2.3.2. Analysis of existing solutions

Creation and integration of the IoT network requires application of its four key elements:

1) Smart Box as a hardware device;

2) Smart App as a software program;

3) special access networks for realizing the process for data transmission and exchange;

4) special platforms for IoT element control.

To access IoT elements, unified dedicated standards are used [20]. It should be noted that eMTC (enhanced Machine-Type Communication) works on GSM network, and eMTC is deployed on the basis of mobile networks LTE, and GSM-IoT (GSM – Internet of Things). Also, the most popular and highly demanded is the special standard for IoT-NB-IoT (Narrowband IoT). The peculiarity of NB-IoT is that it can be deployed and used in both LTE and GSM networks. As a result , it can be concluded

that when designing and creating Smart Box hardware devices, abovementioned unified communication standards should be considered to implement a future platform for interaction of devices.

Many companies-manufacturers of energy equipment are engaged in research into innovations for IoT systems. Thus, LG has created the intellectual self-diagnostics technology Smart Diagnosis and the power management system Smart Grid Ready [26]. These technologies enable the latest models of household appliances to perform self-diagnostics and inform the customer about minor problems (e.g., ice generator turnoff or the emergency mode of a washer motor) using Wi-Fi, NFC and sound signals. This helps preventive diagnose faults and electric equipment (electric motors) malfunction. In systems of this type each company can apply its own protocols and use "smart" operating modes. However, this approach does not allow the system to be unified for all other companies.

Another rather widespread information system for the IoT in the sphere of monitoring and diagnostics is Winnum [27], which is an integrated environment providing many powerful operations for large data volumes, such as collection, storage, and processing of Big Data. It also allows monitoring operation of the system nodes, in particular the current status and monitoring of the technical condition of the electric motors (energy equipment). There is possibility to restore all data of the events before failure or halting. However, the main disadvantage of the systems of that type is the small number of supported electric devices, in particular, the only industrial man-program controlled machines and a small number of operating AMs.

For digital diagnostics of AM as a component of electric equipment, information technologies, methods and technical hardware exist that perform the spectrum-current analysis of the electric grid to which the investigated objects are connected. The method of spectrum-current analysis enables examining AM without direct connection to the object under study [28]. The spectrum-current analysis often involves direct Fourier transform to obtain, for instance, amplitude-frequency characteristic for monitoring in real time. The presence and use of a large number of sensors and the need for expert assessment to determine the final state of the equipment examined is the main disadvantage of spectrum-current analysis.

[29] suggests a structure of a cyber-physical information system as big part of the smart production. This structure assumes that the IoT is enabling communication between elements under study. That is, electric equipment can operate and monitor the technical condition of its units simultaneously. The main disadvantages include a large amount of non-determined information resulted from the technological process and the use of a fairly large number of sensors for each electric motor.

Based on [15]–[29], it can be concluded that the structure of development and deployment of information self-learning and self-diagnostic systems as part of the IoT is a very important and popular task. The main disadvantage in all developed

methods and devices that are commercialized is that the technologies are sensitive and used only within a certain enterprise. Thus, there is no generalized, free and unified system that allows the use of methods and techniques for diagnosing electric equipment. Disadvantages include greater costs of implementing this kind of system at the enterprise due to the necessity of using measurement sensors for each unit of electric equipment under study. As for the implementation of the IoT network technology, there is one major problem – the absence of a unified information management system.

With the aim to decrease the number of used sensors, [18] and [30] suggest applying the indirect method of diagnosing the AM complex. This method uses spectrum-current analysis to diagnose electrical equipment that is connected to the electrical network. This method helps create and integrate Smart Box devices as hardware elements of the general IoT network. It allows using a smaller number of diagnosing sensors and implementing a control platform for different Smart Box modules.

2.3.3. The structure of a modular cyber-physical system for preventive diagnostics of mining and metallurgical power equipment

Figure 2.9 presents a typical logic-functional circuit of an enterprise operating on Industry 4.0 principles, the IoT in particular.

In Fig. 2.9, an enterprise resource planning (ERP) is presented, which is the integrated management of main business processes flow and processes bound with various production technologies and various software. Management enterprise system (MES) is automated computerized systems that are used in technological production in order to monitor the current state of an object, implement documentation, as well as control the process of converting raw materials into finished products. A Smart Box is represented as a hardware-software device. It is used for monitoring and reading the current information from the energy equipment (electrical motor) under study, e.g., spectrum characteristics of the equipment, applying current sensors. EQ_1, EQ_2,..., EQ_n are power equipment over which the current state is currently being monitored.

Management Enterprise System provides technological information to help operators, workers and manufacturers-decision makers understand how current equipment operation modes can be enhanced, thus enabling increase of production efficiency on the basis of optimal performance with minimized expenses on diagnostics and further maintenance.

Thus, based on the suggested logic-functional circuit (Fig. 2.8), the typical circuit of a Smart Box device integration for process implementation of determining the technical condition of the electric motor (energy equipment) under study looks as shown in Fig. 2.10.

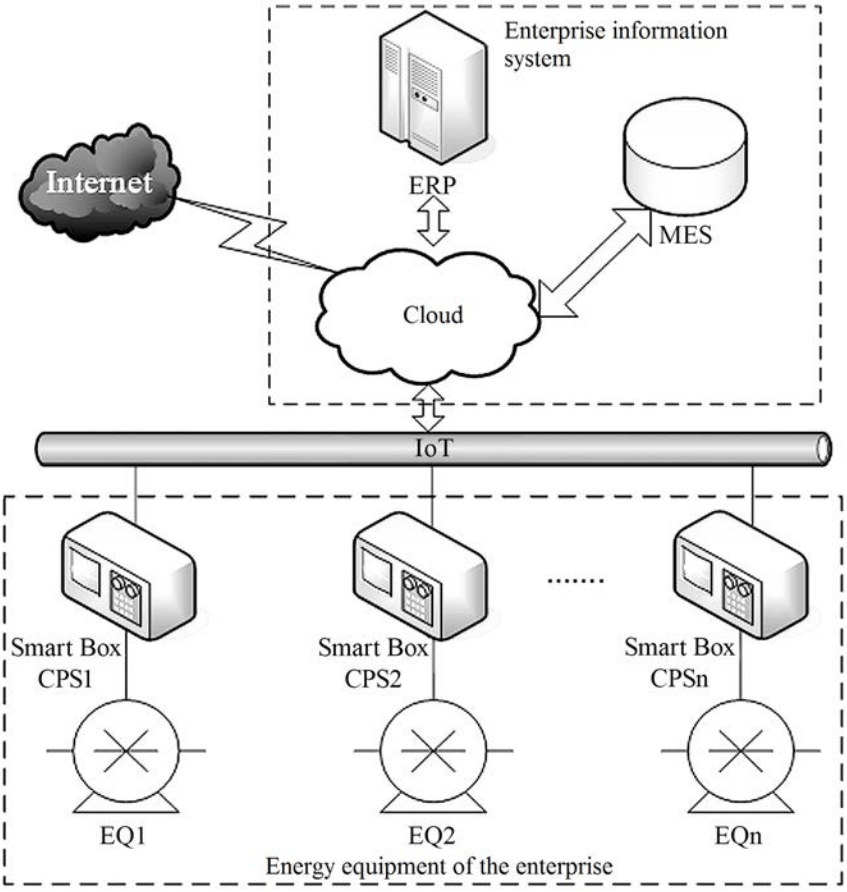


Fig. 2.9. A logic-functional circuit of an enterprise with Smart Box diagnosing devices

In Fig. 2.10, $A = \{a_1(f), a_2(f),..., a_m(f)\}$ is the vector of noise spectrum which generated in the electric grid; $I(t)$ is the current of electric grid; $J=\{j_1(t), j_2(t),..., j_n(t)\}$ is the vector of higher harmonics generated by electric equipment (electric motors) in the electric grid; x_c is the decision on the current condition of the electric motor; $\alpha(t)$ is the special value – the process variable; $B=\{b_1(t), b_2(t)...b_n(t)\}$ is the vector of higher harmonic generated by other non-studied electric motors; $z(t)$ is the energy equipment (electric motor) workload variable; $\varphi(t)$ is the non-studied object operation character; and μ is the vector of the Smart Box setting.

In general, Smart Box device consists of two main and important parts: hardware and software. Hardware includes microcontrollers, various sensors and switches with the data transmission medium. Software processes primary information and does its partial analysis. Direct and more extended analysis is performed on the level of the IoT network. Additionally, the analysis is performed by the enterprise (MES).

Prototypes of current decision support systems are represented on the

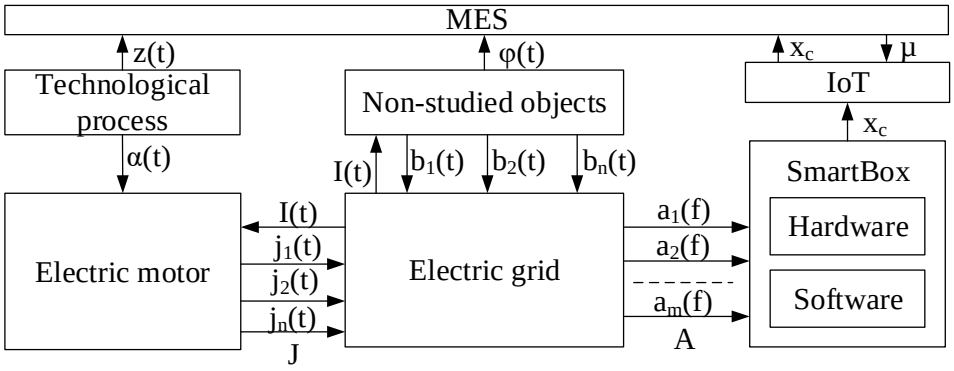


Fig. 2.10. The circuit of a Smart Box device integration for process implementation of determining the technical condition of the electric motor (energy equipment) under study

considered circuits (Figs. 2.9 and 2.10). The main difference is the possibility of remote diagnostics, controlling and management of studied devices' operation at households and factories (enterprises) [20], [29], [30]. Also, the majority of IoT device manufacturers create and use specialized information systems and platforms of their own design. This leads to the fact that there is no possibility to unify the processes of designing, creating and using Smart Box devices. As a result, current software and hardware solutions cannot be used and implemented without the specialized knowledge of experts. Also, the analysis of technical information of the enterprise is based on the use of specialized software and hardware.

[30] and [31] suggest an improved structure of a modular cyber-physical smart system for preventive monitoring and diagnostics of electric motors (energy equipment) based on applying and deploying IoT approaches. It allows the creation of a process for group spectrum analysis. A Smart Box is the key element of any IoT network.

There are several versions of Smart Boxes for enterprises. Besides tasks to be solved, Smart Boxes can be characterized by utterly different architecture of hardware and software that can cause additional expenses on their integration into the general IoT network of the enterprise due to the use of extra hardware and software. That is the main reason for suggesting the singling out each Smart Box as an individual module capable of analysing adjacent Smart Boxes in addition to self-learning, preventive diagnostics and self-analysis.

The model of a cyber-physical smart system applying a Smart Box device for preventive monitoring and self-diagnostics of electric motors (energy equipment) is presented in Fig. 2.11.

In Fig. 2.11, the AMs under study are connected to a power supply network.

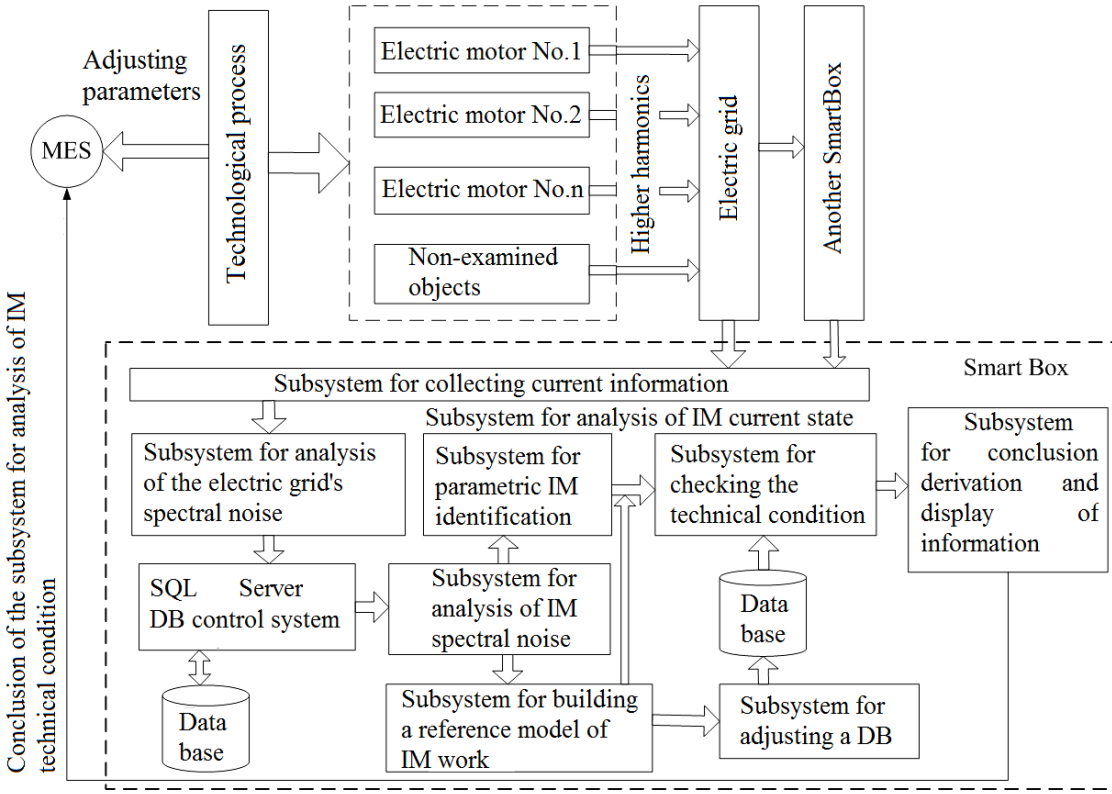


Fig. 2.11. The model of a cyber-physical smart system applying a Smart Box device for preventive monitoring and self-diagnostics of electric motors

Power network can be a single- or three-phase. According to existing research [18], in the course of operation and due to their design features, AMs create higher harmonics in the electric grid. To analyse higher harmonics which are produced by the equipment, the current technological information collecting subsystem converts an analogue signal into a digital one. Then, a process for analysing the spectral noise of the electrical network is created for analysis [26]. An analogue-digital converter can be the main or the only element of this subsystem. The main and most important task of the Database Management System (DBMS) is storing, receiving, and managing all the necessary and important technological data to provide efficient and fail-safe operation of the modular Smart Box device. It should be mentioned that technological data can consist of special technological data responsible for conservation of the standard of the studied AM operation and data responsible for current parameters of the studied AM operation. The decision-making subsystem and information output is an expert system. Also, the resulting and final decision on the current technical state of the object under study should be accepted by the MES because of the technological enterprise flows and process features and possible errors in calculations that can lead to mistakes when concluding about the technical condition of the equipment.

The above model (Fig. 2.11) enables use of one Smart Box for several electric motors of the same type. To identify each motor as an individual object under study and distinguish it among the group of similar motors, the spectrum analysis is performed [29]. Object identification is one of the main aspects of the IoT concept [19], [20].

The suggested approach enables monitoring of the current condition of the equipment applying fewer Smart Boxes. In general, the number of Smart Box devices is dependent on the type and quantity of the electric motors (energy equipment) used, features of the technological enterprise process, and the type of communication between them. Thus, the logic-functional circuit of the enterprise with module-based diagnostic Smart Boxes can be presented as shown in Fig. 2.12.

Also, each module can contain only one Smart Box device. This Smart Box device can be connected to one or more electric motors connected to the electric grid. Together, all modules represent an information computing cluster. Arrangement of objects into a cluster results in their efficiency and reliability.

Also, Smart Box device cannot make accurate and correct conclusion data solely on the basis of testing (current) technological data, as it is impossible to consider all possible situations of the system (e.g., errors in sensor measurements (ADC), unavailability of an accurate mathematical model of defect appearance regularities and sensor faults). That is why in the Smart Box operation a mismatch of situation classifications can occur. To solve this problem, one can use elements of fuzzy logic, soft computing ANFIS, genetic algorithms in neural networks, mathematical methods of discriminant analysis, and other methods of classification.

As a result of the considered types of information technology systems [30], to build an expert Smart Box system, it is suggested to apply a multilevel fuzzy-neuro network hybrid system which consists of subnetworks of various architectures (neural network and fuzzy logic). In particular, in a diagnostic system, a set of possible situations can be split into the set of routines (S1) and the set of exceptions. Then, it

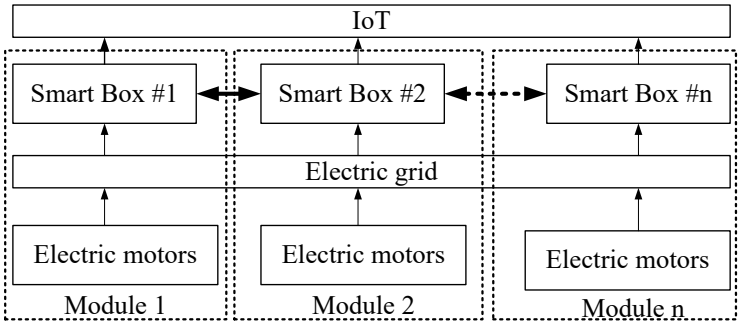


Fig. 2.12. The module network of diagnostic Smart Boxes

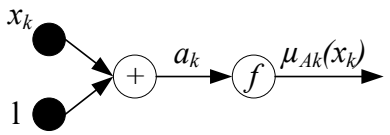


Fig. 2.13. The structure of the FNS membership function for the property x_k of a specific defect k

is necessary to make a correct decision and possible to correctly determine the type of this situation and attribute it to set S1 *(t)* or the set of situations S2 *(t)*. The correct solution of the task consists of building the set of rules which identify and determine the current technological situation and can be involved in the process of calculating the input function (using a system based on soft computing).

The dependence of the functioning of neurons can be represented as follows:

$$y = f(s) = f(\Sigma_{i=0}^{n} x_i \cdot w_i), \tag{2.16}$$

where y is the variable that represents neuron output; *f(s)* is the result of works activation function; x_i is inputs of neural network; and w_i is weighting factors.

Using a fuzzy neuronetwork system (FNS) is a soft computing paradigm. FNS should contain clear vectors of input data and fuzzy levels of influence of each input on a possible situation (output). Then the FNS can represent a 3-layer structure. The set of levels will allow controlling (Fig. 2.13) the current state of the *l*-th component (defect) of the Smart Box of the device ($l = \overline{1,k}$), where the spectral characteristics of the equipment (input data) represent the first level [x_0;x_n]; the second level filtrates situational attributes $C_j(j = 1,N)$ – separation of noise from a wanted signal; the third level identifies the situation (defect/no defect).

Fuzzy datasets x_k at the first level are network weights (the range of amplitude fluctuations on the corresponding frequency), vector A_k is the result of generalizations.

The weights of the first layer of the neural network are fuzzy sets $A_{ki}(k = \overline{1,N_l}, j = \overline{1,N_c})$; $\mu_{A_k}(x_k)$ is the activation function calculated by the formula

$$\mu_{A_{kj}}(x_k) = \frac{1}{1+e^{-akj}}. \tag{2.17}$$

The weight coefficients of the second level of the NFS network take values within [0; 50]. Also, weight coefficients are the average deviation of amplitude fluctuations on the corresponding frequency.

Figure 2.14 presents the general case of the NS structure.

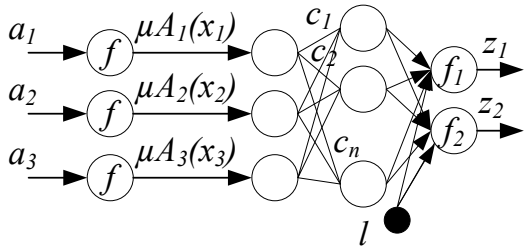


Fig. 2.14. The fuzzy neuro network system structure

Considering each level, fuzzy neuro network functioning looks as follows:

$$z_m^h = f_m^s(\sum_{j=0}^{N_c} w_{jm}^s \cdot \mu_j^s(x_1,..,x_N)), \tag{2.18}$$

where s = 2 and s = 3 are the network levels; h = 1 and h = 2 are the corresponding levels of the resulting state; and f_m^s determines the value of the activation function of the output layer of the NFS network. Based on the use of expert judgment, the membership function is determined [32].

Thus, the suggested fuzzy neuro network system as the logic inference apparatus for the Smart Box enables monitoring the current condition and preventive diagnostics of electric motors (energy equipment) in real time.

2.3.4. Results of modelling the operation of a modular cyber-physical smart system for preventive diagnostics of energy equipment

Adequacy and fitness of application of the suggested modernized structure of the modular cyber-physical smart system of preventive monitoring and diagnostics of electric motors (energy equipment) were analysed by simulation and applying computer-aided Monte Carlo analysis. This method is statistical, i.e., it imitates real distribution of sample statistics provided by a great number of experiments. The random number generator used was the one that is part of technology Net.Framework 5.0 based on the computer clock generator.

According to the research conducted [18], while monitoring the current condition of AM, 6 and more characteristic frequencies (CF) should be used for AM identification. It should be mentioned that as characteristic frequencies are a learning set for the FNS network, the total number of CF frequencies can impact the learning time and the total time of the information system response.

Impacts of the CF number on the learning time were analysed through computer modelling in Matlab software using a special Fuzzy Logic module and using the app m-function ANFIS where CF were used as input variables. As an ANFIS learning algorithm, the error back propagation method was chosen to study the parameters of the membership

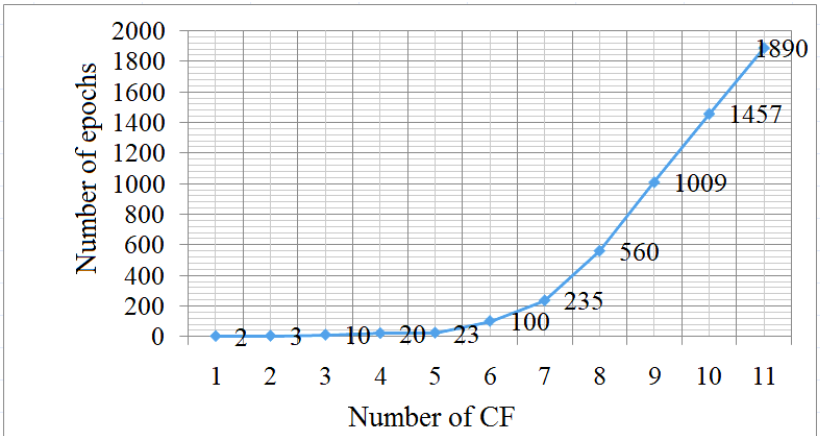


Fig. 2.15. Results of modelling impacts of the CF number on the learning time

function. This method is based on the gradient method of the fastest descent. Figure 2.14 presents the results.

As is seen from the results of computer modelling, the smallest learning time can be obtained using 1–6 characteristic frequencies.

In the course of research for each type of Smart Box device system reaction time, τ, to the object being studied was measured. The average reaction time is considered as the time required by a Smart Box to conclude about the technical condition of asynchronous motors. At each experiment level, the total number of asynchronous electric motors under study changed from 1 to 5. The simulation of the occurrence of a defect in the electric engine was derived from the random generation of spectral noise. The input sample for the neural network comprised 6 CF.

Reactions of standard and modular Smart Boxes are compared in Fig. 2.16.

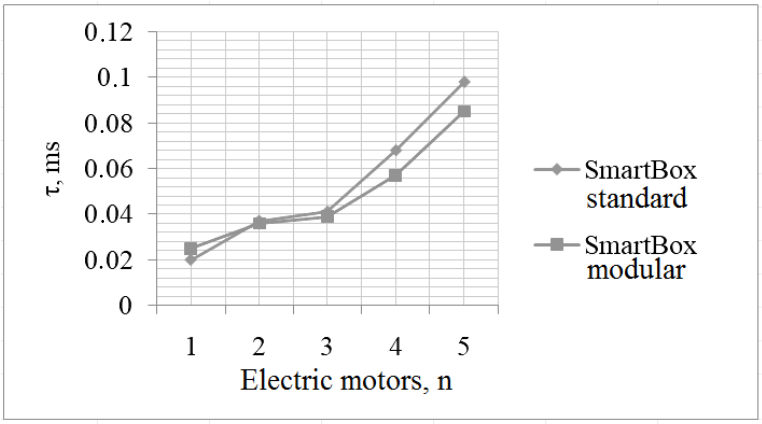


Fig. 2.16. Results of modelling the process of preventive diagnostics in standard and modular Smart Box

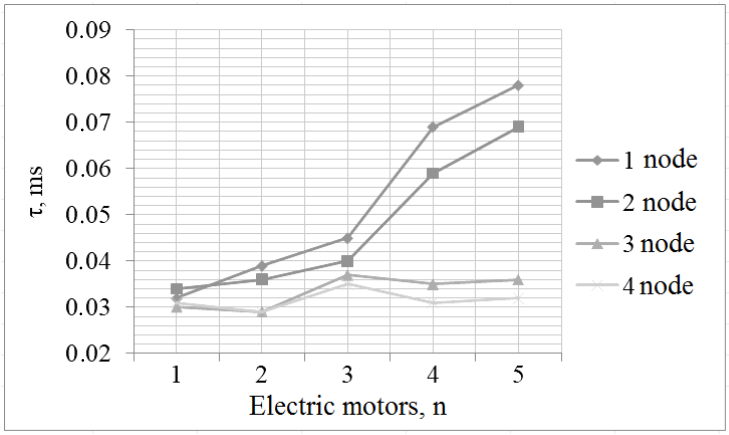


Fig. 2.17. The results of modelling the process of preventive diagnostics in cluster systems of modular Smart Box

The modular structure of a Smart Box was analysed and tested through the increase of diagnosed energy equipment (asynchronous electric motors) from 1 to 5 and an increase of the total number of computing cluster nodes from 4 to 1. The Smart Box module was used as a computational cluster node.

A model of HPC-type (high performance computing clusters) system was used as a cluster type. Figure 2.17 presents the testing results.

Thus, from the testing results it can be concluded that the ability to react in the cluster Smart Box system is higher by approximately 30–34 % as compared with the standard one provided where 3 or 4 cluster nodes are used.

At a real enterprise, obtained results can obviously differ due to appearance of various situations. Yet, the quality of detection of electric motor defects will be more efficient in the case of deploying a modular Smart Box system.

As a result of the performed calculations, mean square deviation made Sa = 2.77 Hz. Also, absolute error for 95 % reliability (Student's coefficient $t\alpha$ = 1.984 at α = 0.05 and n = 100) made Δx = 5.78. The ratio error was calculated with the following value ea = 3.46 %. Repeatability of experiments (homogeneity of variances) was tested Gp = 0.3293 at the threshold Gk = 0.369.

2.3.5. Analysis of the research results of the proposed structure of the modular cyber-physical system

The main task of electric motors (energy equipment) technical preventive diagnostics at the enterprise was solved by application of the concept of cyber-

physical systems, i.e., combination of Smart Boxes with smart production (MES), where every modular Smart Box device is capable of preventive diagnostics both individually and in a complex. Use and deployment of modular Smart Boxes reduces the number of diagnostic devices. The possibility of creating clusters of Smart Box modules increases fail-safe and high-speed performance of the IoT system as a whole.

In case of absence of integration to the general IoT network of the enterprise, the suggested modernized structure of a cyber-physical smart system for preventive monitoring and diagnostics of electric motors (energy equipment) may be part of the IoT of an enterprise, a factory or a private household. Smart Box can be a part of a "Smart house" system and can operate independently (e.g. to diagnose household appliances – hair dryers, washers, and vacuum cleaners).

The main advantage of the suggested cyber-physical smart system is the ease of customization, as it allows the use of a fuzzy output system as a subsystem of decision-making information output. The structure of the fuzzy system enables easy change of parameters of term-sets and increased quality of defect detection.

The possibility of creating a cluster of modular Smart Boxes is another advantage of the system, thus enabling work in combination with other modules and additional information exchange with adjacent Smart Boxes.

Disadvantages of this type of intellectual systems include inability to work with power plants not equipped with AM; probability of errors connected with the electric grid quality – voltage changes, higher harmonics caused by several-fold more powerful equipment. The prospect of development of this direction consists in creating the modernized structure of the alternative global IoT network for technological energy equipment. This will increase the quality of device communication and enlarge the information base for diagnosing.

Chapter 3:

Cybersecurity

Sergey Zaitsev
Chernihiv Polytechnic National University

Co-authors:

Boris Horlynski
Chernihiv Polytechnic National University

Volodymyr Prystupa
Chernihiv Polytechnic National University

3.1. Basic terms of cybersecurity

The benefits of today's digital world facilitated by the exponential development of information technology are counterbalanced by the increase of new threats to national and international security, especially in cyberspace.

Cyberspace is defined as a virtual environment (virtual space) that provides opportunities for communication and/or implementation of public relations. To achieve this goal, cyberspace makes use of the Internet and/or other global data networks or operates other compatible (connected) communication systems provided by electronic communications.

The prefix "cyber" is used to give the word a meaning of something that belongs to the age of computers – the Internet and digital technologies (e.g., cyberspace, e-sports, cyberculture, cyber-attacks, cybercrime, and cybersecurity).

The basic definitions of words with prefix "cyber" are as follows:

- ***Cyberattack*** – directed (deliberate) actions in cyberspace, which are performed using electronic communications (such as software, software and hardware, or any other communication technologies) with the scope of achieving one or a combination of the following objectives: (a) to get unauthorized access to confidential or integrity information, or to make available to third parties of processed electronic information resources (transmitted or stored) via communication and/or technological systems; (b) breach in the security of regular operation of communication and/or technological systems affecting or not their sustainable and reliable operation; (c) use of the communication system or other electronic communications and their resources to carry out cyberattacks on other objects of cyber defence.

- ***Cybersecurity*** – protection of vital interests of one or more citizens, a business entity or the state in the cyberspace, which ensures the sustainable development of the digital communication environment (and the information society) by preventive measures, timely detection and neutralization of real or potential threats in cyberspace.

- ***Cyberthreat*** – existing and potentially possible phenomena and factors that create danger in cyberspace, have a negative impact on the state of cybersecurity of the state, cybersecurity and cybersecurity of its objects.

- ***Cyber defence (1st meaning)*** – a set of organizational, legal, engineering and technical measures including cryptographic and technical actions needed for the protection of any sensitive and non-public information. The main scope is the prevention of cyber incidents, while in case of cyberattacks to offer

means for timely detection and protection as well as to eliminate possible costly consequences. It also ensures restoration of stability and reliability of communication and technological systems.

- ***Cybercrime*** (computer crime) – a criminal act in cyberspace that is potentially dangerous for the society and social life, liability for which is provided by law.

- ***Cybercrime*** – a set of words related to cybercrimes:

    - ***cyber defence (2nd meaning)*** **–** a set of political, economic, social, military, scientific, scientific and technical, informational, legal, organizational and other measures carried out in cyberspace and aimed at protecting the sovereignty and defence capabilities of the state, preventing armed conflict and repelling armed aggression;

    - ***cyber intelligence*** – activities carried out by intelligence agencies in cyberspace or using it;

    - ***cyberterrorism*** – terrorist activity carried out in cyberspace or using it;

    - ***cyber espionage*** – espionage carried out in cyberspace or using it.

Along with incidents of natural (unintentional) origin, the number and power of cyberattacks motivated by the interests of individual states, groups and individuals is growing. There are cases of illegal collection, storage, use, distribution or destruction of personal data, or illegal financial transactions, fraud or other sort of theft on the Internet. Cybercrime is now becoming transnational and can cause significant harm to the interests of the individual, society, and the state. Extensive use of digital technologies in economic, scientific, technical, and information spheres, in the field of public administration, in defence-industrial and transport complexes, in communication systems, security and defence sector makes them vulnerable from the point of view of intelligence and subversive activities by individual states, their intelligence services or hacker groups in cyberspace.

The potential for damage from cyberattacks at the personal or systemic level is enormous. According to some estimates, organized for profit or just to create a mess, cyberattacks now cost the world economy about $ 400 billion a year – this amount exceeds the GDP of about 160 of the 196 countries on our planet [33].

The most destructive cyberattack in history is considered to be the NotPetya virus, which on June 27, 2017 attacked numerous computer systems of government and commercial institutions. The NotPetya virus, which belongs to the family of malware, infects computers running the Microsoft Windows family of operating systems (OS), namely encrypts IP, and overwrites and encrypts the data needed to boot the OS.

As a result, IPs become inaccessible. For decryption and restoration of access to the IP, the program requires a ransom in "bitcoins". In total, according to Microsoft and ESET, the cyberattack affected at least 65 countries, while the overall damage from this attack is estimated at $ 1.2 billion [34].

The real manifestations of cyberattacks are unpredictable, and their result is significant financial and economic losses or unpredictable consequences of disruptions in the functioning of information and telecommunications systems, which are directly affecting the state of security and defence.

Analysing the definitions of cybersecurity, we can conclude that they are different in meaning. Cybersecurity is measures related to and aimed at the security of systems and information resources; cybersecurity is also a broader concept that includes actions and measures related to the security of individuals and public relations, including law enforcement, intelligence, counterintelligence, operational and investigative, as well as political, informational, public education, and direct cyber security measures. At the same time, the development of "cyber defence" should clearly prevail over the development of "cyberattacks".

To ensure comprehensive protection of the basic properties of information (confidentiality, integrity, accessibility) from cyber threats in modern ITS, it is necessary to distinguish the following main areas of cyber defence: organizational protection, legal protection, engineering protection, cryptographic protection, and technical protection.

However, information in its transmission may also be unintentionally affected by industrial, stationary or natural disturbances, and various other factors. In this case, noise-tolerant coding is used. It should be noted that the complex use of cryptographic methods of information protection and noise-tolerant coding methods complement each other and allow to ensure the effective functioning of information and telecommunication systems. Thus, noise-tolerant coding methods occupy a certain niche among the methods of ensuring the integrity of information. Until recently, satellite communication systems, digital television standards, mobile communication systems, and other relevant systems have increasingly used turbo codes as the most powerful noise-tolerant codes. However, as of today, low-density parity-check codes (LDPC codes) have been used instead of turbo codes in Wi-Fi (IEEE 802.11 standard), WiMAX (IEEE 802.16 standard), DVB-T2, DVB-S2, etc. This is due in particular to the following factors:

- rapid development of technologies, in particular the power and speed of electronic computing, which allows the implementation of LDPC codes that have high resource consumption and computational complexity of algorithms and are more complex than turbo codes;

- unprotected patents of LDPC codes in contrast to turbo codes;

- the ability to approach with the help of LDPC codes to the Shannon limit (0.6–0.8 dB);

- the expediency of using LDPC codes in information and telecommunication systems, where energy savings or signal-to-noise ratio are very small.

LDPC codes were proposed by Gallagher and studied in scientific works of domestic and foreign authors, including: McKay, Lee, Radford, Zolotarev, Ovechkin, Bashkirov, Novikov, and other scholars.

Modern information and telecommunication systems use noise-tolerant codes to increase the reliability of information transmission: Heming codes, Bose-Chowdhury-Hawkingham codes, Reed-Solomon codes, cascade codes, convolutional codes, turbo codes, and LDPC codes.

Third-generation, fourth-generation, and fifth-generation 5G wireless data transmission systems use adaptive power management, modulation, and coding parameters to increase the reliability of information transmission. Thus, for adaptation one-level schemes are used, for example, for adaptation of encoders of the turbo code or LDPC of the code, only the coding speed changes.

3.2. Basic approaches to the classification of cyber threats

The latest information technologies can be both means of production and cyber weapons. Cyberspace has now become an active arena of information confrontation between different states.

Hacking program… Virus… Worm… Trojan… DDoS attack…

These terms refer to attacks using virtual weapons , but we are just beginning to realize all the consequences of this phenomenon.

Increasingly, we find in the news information about cyberattacks in the information resources of financial institutions, energy supply and transport companies, and government agencies that guarantee security, defense and protection against emergencies.

In 2020, due to the spread of coronavirus and the mass transition of companies to remote work, hackers developed the following viruses:

- A «coronavirus map» that allows attackers to access sensitive information stored on an infected computer. It is distributed on the Internet as software that provides real-time information about the spread of coronavirus in the world. After installation, it launches a number of interactive processes on the infected PC, which provide interception and transmission of private information (usernames,

passwords, credit card information, etc.), search and transfer of files needed by the attacker, substitution of e-wallet addresses for cryptocurrencies.

- "Black Water", hidden in the file "Important - Covid-19.rar", which is information about coronavirus COVID-19 and is distributed on the network. When the recipient reads the contents of the file, a virus is launched and provides remote access to the PC.

- "HawkEye", hidden in phishing emails on behalf of the Director of the World Health Organization. The emails contain an archive with the file "Coronavirus Disease (Covid-19) CURE.exe". When the recipient reads the contents of the file, a keylogger is launched, which registers the various actions of the user – keystrokes on the computer keyboard, movements and keystrokes.

The following are the main approaches to protecting ITS in cyberspace and protecting against cyberattacks.

To ensure comprehensive protection of the basic properties of information (confidentiality, integrity, accessibility) from cyber threats in modern ITS, the following main areas of cybersecurity must be separated: organizational protection, legal protection, engineering protection, cryptographic protection, and technical protection (Fig. 3.1).

The set of measures shown in Fig. 3.1 should be aimed at:

- cyber incident prevention;

- detection and protection of ITS from cyber attacks;

- elimination of the consequences of cyberattacks in ITS;

- restoration of stability and reliability of functioning of communication and/or technological systems.

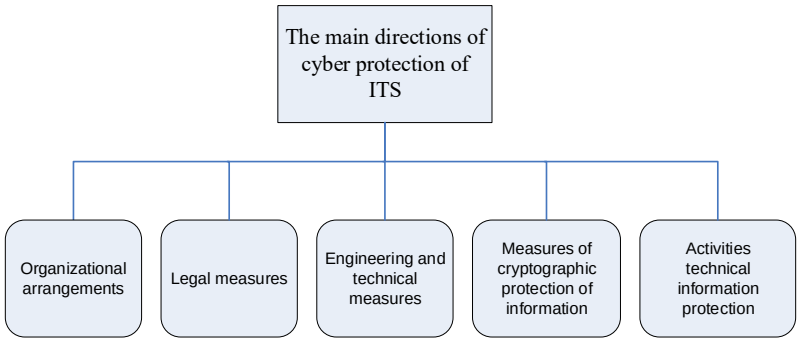


Fig. 3.1. The main directions of cybersecurity of ITS

To protect information in information and telecommunication systems, it is necessary to use a comprehensive information security system (hereinafter – CISS) with confirmed compliance.

One of the stages of creating a CISS is, in particular, the definition of threats (potential and significant) for information, based on the analysis of which the choice is made of basic solutions to combat all significant threats, the formation of general requirements, rules, restrictions, and recommendations, which regulate the use of secure information processing technologies in ITS, certain measures and means of information protection, and the activities of users of all categories.

Building reliable cybersecurity in ITS is impossible without prior analysis of the properties of information, classification of the AS in which information circulates, as well as possible threats (as an interconnected set of information and AS).

The properties of information according to the law should be understood by three main components:

- ***confidentiality*** – the property of information to be protected from unauthorized access;
- ***integrity*** – the property of information to be protected from unauthorized distortion or destruction;
- ***accessibility*** – the property of information to be protected from unauthorized blocking.

Researchers O. Yudin and S. Buchyk in [35, *95*] give the general system of threats, DIR (see Fig. 3.2), which can be used for the general classification of cyber threats.

Attention should be paid to the classification of threats by the nature of the direction (regulatory, organizational, and engineering direction) and additional principles of classification of threats by strategic or tactical nature.

Information security threats are classified in the following information security standards:

- Trusted Computer System Evaluation Criteria (TCSEC);
- European Information Technology Security Evaluation Criteria (ITSEC);
- Federal Criteria for Information Technology Security (FCITS);
- Canadian Trusted Computer Product Evaluation Criteria (CTCPEC);

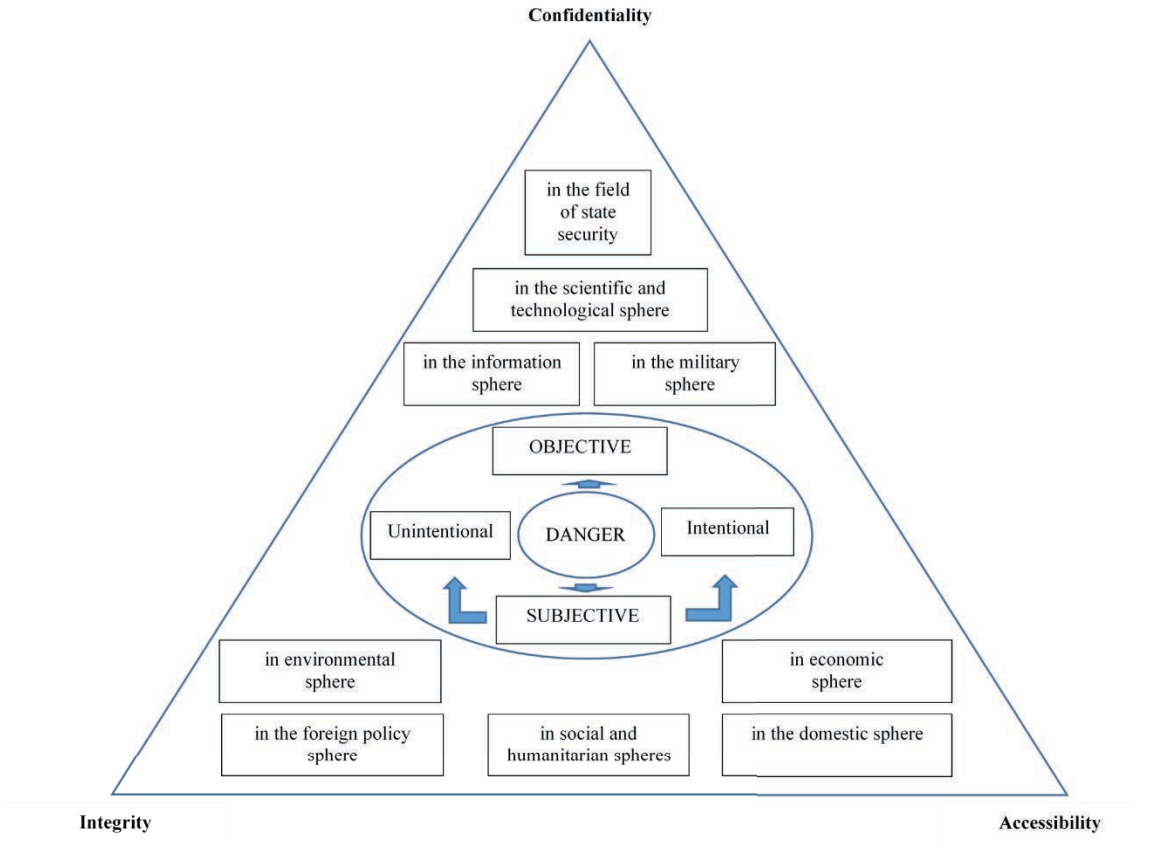


Fig. 3.2. General threat system

- General Criteria for Information Technology Security Evaluation (ISO / IEC 15408);

- Criteria for assessing the security of information in computer systems from unauthorized access (ND TZI 2.5-004-99).

As a result of the impact on information and the system of its processing threats are divided into four classes:

1. *Violation of confidentiality of information* (gaining access to information with limited access).

2. *Violation of the integrity of information* (complete or partial destruction, distortion, modification, imposition of false information).

3. *Violation of information availability* (partial or complete loss of system

performance, blocking access to information).

4. *Loss of observation or controllability of the processing system* (violation of procedures for identification and authentication of users and processes, giving them authority, exercising control over their activities, and refusal to receive or send messages).

"Observation" means a property of the system that allows a user to record the activities of other users and processes, the use of passive objects, as well as unique establishments of the identities of users and processes involved in certain events to prevent security breaches and/or ensure responsibility for certain actions.

This classification can be used in the field of cybersecurity.

Let us consider Recommendation X.800 (Security architecture for Open Systems Interconnection for CCITT applications) [36], which defines the levels of the reference seven-level OSI model [37] on which security functions can be implemented, security mechanisms are used as well as security administration.

Common security features (services) of the X.800 Recommendation are:

- *Authentication* provides authentication of communication partners and authentication of the data source. Partner authentication is used when establishing a connection and periodically during a session; it serves to prevent such threats as *a "masquerade"* and a repeat of a previous communication session.

- *Access control* provides protection against unauthorized use of resources available on the network.

- *Data confidentiality* provides protection against unauthorized receipt of information separately highlighting the confidentiality of traffic, i.e. information that can be obtained by analyzing network data flows.

- *Data integrity* is divided into subspecies depending on the type of communication used by the partners – with or without a connection, whether all data or only individual fields are protected, whether recovery is provided in the event of a breach of integrity.

- *Non-repudiation* provides two types of services: failure with confirmation of the authenticity of the data source and failure with confirmation of delivery.

The following special mechanisms and their combinations can be used to implement security services (functions):

- encryption;
- digital signature mechanisms;
- access control mechanisms;
- ensuring the protection of data integrity (data integrity mechanisms), which include cryptographic control functions;
- authentication (authentication exchange mechanisms);
- padding traffic mechanisms;
- routing control mechanisms;
- notarization;
- trust functionality, event detection, security control.

The relationship between services and special security mechanisms is shown in Fig. 3.3.

Consider also the main approaches to the classification of the AS. According to the set of characteristics of the AS (configuration of OS hardware and their physical location, the number of different degrees of restriction of access to processed information, the number of users and user privileges), it is possible to distinguish three hierarchical classes of AS, namely:

Class 1 – a single-machine single-user complex that processes information of one or more degrees of access restriction (for example, an autonomous personal computer, access to which is controlled using organizational measures).

Class 2 – a localized multi-machine multi-user complex that processes information of various degrees of access restriction (for example, a local area network).

Class 3 – a distributed multi-machine multi-user complex that processes information of various degrees of access restriction (for example a global network).

The Council of Europe Convention on Cybercrime [38] identifies the following groups of threats to the confidentiality, integrity and availability of computer data and systems implemented through:

- unauthorized access to the information environment (illegal intentional access to a computer system or part thereof, bypassed by security systems);

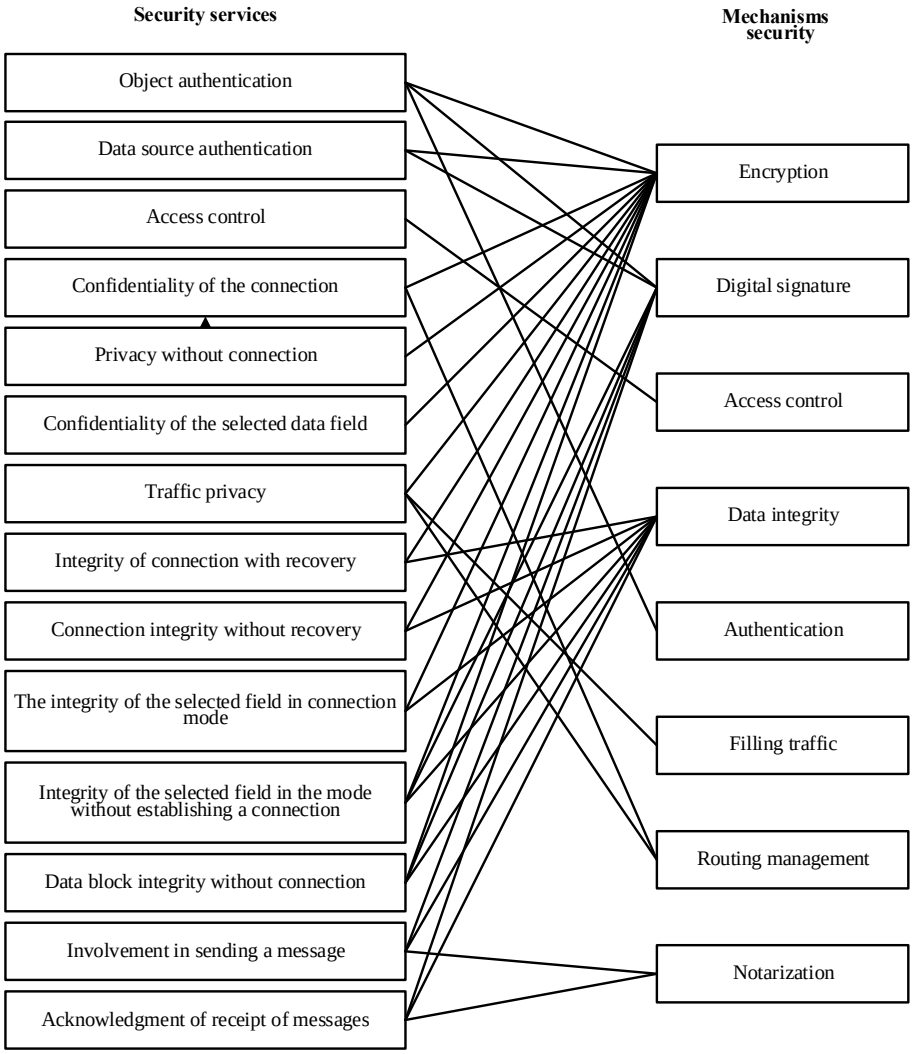


Fig. 3.3. Relationship between services and special security mechanisms

- illegal interception (unlawful intentional audiovisual and/or electromagnetic interception of computer data not intended for public access);

- data interference (unlawful alteration, damage, deletion, distortion or blocking of computer data and control commands by means of cyber attacks on information systems, resources and networks of state and military administration);

- interference in the operation of the system (illegal violation or obstruction of the computer system through the development and dissemination of

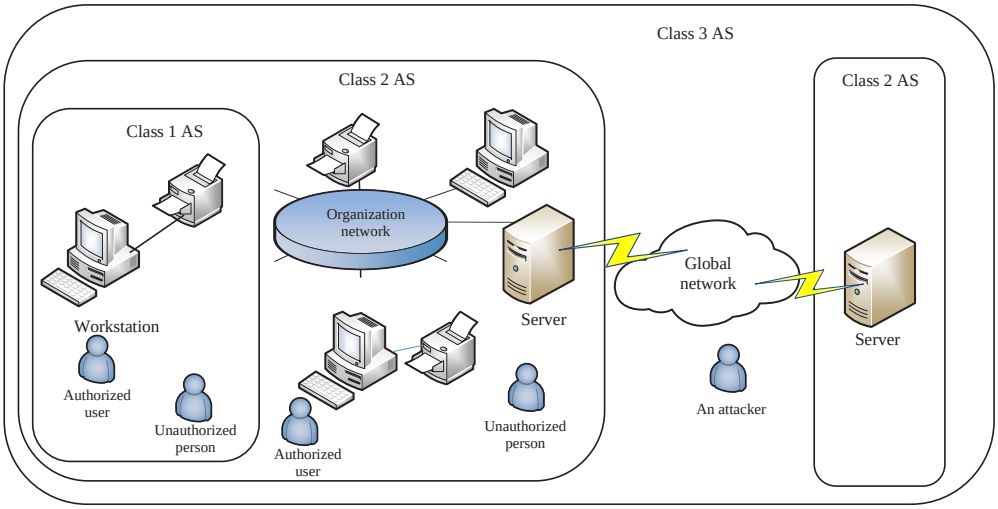


Fig. 3.4. Examples of AS Classes 1, 2 and 3

viral software, the use of hardware bookmarks, electronic and other types of influence on technical means and systems of telecommunications IP protection systems, systems and networks, software and mathematical software, data transmission protocols, addressing and routing algorithms);

- illegal use of computer and telecommunications equipment.

According to the method of spreading, cyberattacks can be divided into mass and targeted attacks [39].

Massive cyberattacks are aimed at the global spread of malware (malware – short for –"malicious software"), which can disrupt the system, delete important files or damage them.

The most common malware examples are [40]:

- Traditional (file) viruses – viruses that infect programs, making them inoperable, or delete certain files from your computer.

- "Trojan horses" (Trojans) – viruses disguised as useful programs. The capabilities of such programs vary from simply tracking an infected computer to destroying certain files with a remote command or timer.

- "Worms" – used to send spam and create botnets (entire networks of infected computers). They can enter the computer both through the user's fault (by clicking on certain links in e-mails) and without user's participation – using

"vulnerabilities" in security programs.

- "Anti-antivirus" – viruses specifically designed to infect antivirus programs.
- “Sniffers” – "spyware" programs designed to intercept and further analysis, or only the analysis of network traffic intended for other nodes.
- “Rootkits” (suite of a superuser that has the right to perform all operations without exception) – a "spyware" program or a set of them to hide traces of the presence of an attacker or malware in the system.
- “Miner viruses” – programs or scripts used to extract cryptocurrency without the knowledge of the system owner, while reducing its computing power.
- “Macroviruses” – viruses written in the language of macros, which is used to automate some processes in programs such as Microsoft Word. Macroviruses infect the program itself through an open document with a "surprise", and then the infected program infects all documents, which are opened.
- “Droppers” – less harmful viruses (or relatively harmless programs) designed to "distract" the antivirus system at the time of infection with a more harmful virus.

Unlike mass cyberattacks, targeted (targeted) cyberattacks are pre-thought-out actions to destroy the information systems of a particular organization.

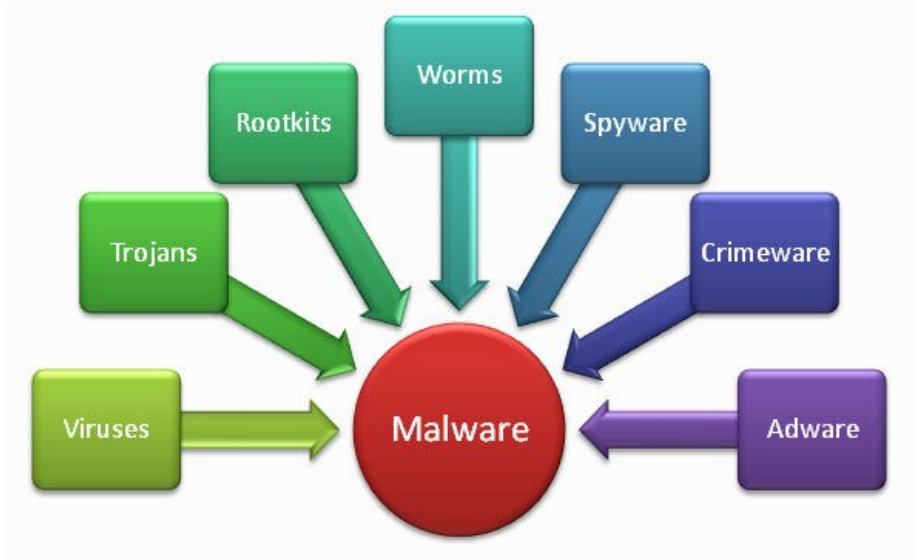


Fig. 3.5. The most common types of malware

Traditionally, a targeted cyberattack is carried out in several stages:

- research, during which the analysis of the state of penetration into the IP is done;
- exploitation of vulnerabilities with installation on the device of the victim of remote control;
- fixing in the system with suppression of means of protection, blocking of control systems and destruction of traces of penetration;
- installation of target software and its operation.

Among the targeted attacks, the so-called cyberattacks APT (Advanced Persistent Threat) are a type of complex cyberattacks to gain unauthorized access to information systems of the «victim» and establish covert access to it in order to use or control in the future.

The following cyberattacks are included in the class of developed constant attacks [41]:

- Operation Shady RAT, which lasted several years and killed more than 70 companies, governmental and non-governmental institutions, and international organizations.
- Operation Aurora – a cyberattack against the information systems of a number of companies in the technology and security sector. The attackers attacked software configuration management systems that contained information from Google, Adobe, and a number of Fortune 100 companies over the past few months.
- Hacking of the information systems of the European Commission on the eve of an important summit.
- Hacking of the information systems of the French government on the eve of the G20 meeting.
- Hacking of the information systems of the International Monetary Fund.

The following sources of threats in cyberspace can be identified:

- Botnet operators are hackers who, instead of gaining access to the system, take control of several (from tens to hundreds or thousands) systems to

coordinate attacks and to spread phishing schemes (phishing is a form of fraud in which an attacker masquerades as a reputable entity or person in email or other forms of communication), based on users' ignorance of the basics of network security, the purpose of which is to extract from gullible or inattentive users of the network critical data, spam and attacks aimed at denying service Denial of Service – DoS-attacks.

- Criminal groups seeking to attack systems for monetary gain; in particular, organized crime groups use spam, phishing, malware for identity theft and online fraud.

- Foreign intelligence services can use cyber tools as part of their espionage activities. In addition, some countries are actively working to develop the doctrine of information warfare (including cyberwarfare).

- Hackers hack networks for the thrill of "victory" or as an attempt to increase their status in the hacking community, or for financial gain. While remote access to a computer requires little knowledge, hackers can download attack scripts and protocols from the Internet and run them against the victim's site. Thus, while malware becomes more sophisticated, it also becomes easier to use.

- Insider – a person who is dissatisfied with the work in the organization or the organization as a whole. Such a person can become a major source of cybercrime. An insider does not need a high level of knowledge about hacking the system because his/her knowledge of the target system allows to gain almost unlimited access to damage the system or to steal confidential system data.

- Terrorists seek to destroy, neutralize or use critical infrastructure in order to endanger the national security of the victims, to cause mass casualties, to weaken the economies of entire countries or individual organizations.

To carry out a cyber attack, the attackers use the following methods:

- interception of passwords of other users;

- "social engineering";

- use of software errors and software bookmarks, as well as errors in user identification mechanisms and imperfections of data transmission protocols;

- obtaining information about users by standard means of operating systems;

- blocking the service functions of the system under attack.

The most vulnerable in any infrastructure are resources open to external access – websites, application servers, databases and others. To attack an organization's internal resources, an attacker needs to find infrastructure vulnerabilities and overcome network security; and attack on external resources requires much less effort. One of the most popular types of attacks on public services is a distributed attack aimed at denial of service (Distributed Denial of Service (DdoS)). Its essence is to generate a large amount of parasitic traffic that is sent to target servers, such as to generate traffic which exceeds the bandwidth of the channels that connect the servers to the Internet, leading to traffic processing to be blocked.

3.3. Modern digital cybersecurity tools

Existing threats in cyberspace require the implementation of comprehensive measures aimed at ensuring cyber protection. Under these conditions, the state must create adequate legal, organizational, technical and other means and methods of cyberspace protection that exist, which are a reflection of the state policy of information security and cybersecurity.

Knowledge of the principles and mechanisms of cyber attacks allows ITS owners to select and implement effective cyber defense tools that will allow (1) to prevent cyber incidents; (2) to detect cyberattacks; (3) to protect against cyberattacks; (4) to eliminate the consequences of cyberattacks; and (5) to restore the sustainability and reliability of ITS. The most widely used digital tools to protect against cyberattacks are shown in Fig. 3.6.

By digital tools we mean software and hardware – hardware devices (tools) designed to ensure the confidentiality, integrity, availability and monitoring of information by blocking unauthorized access to IT computing resources, vulnerability detection, response to cyber threats, analysis and prevention of cyberbullying and cyber incidents.

An electronic signature (hereinafter – ES) is the digital equivalent of a signature (seal, stamp, etc.), the presence of which in the message allows you to accurately determine the source of the message (document) and legally prove that, with some probability, only the sender could create and sign this document. ES is used to provide authentication services and integrity protection for which the subject of the verification of the signed data is unknown in advance. With a certain choice of controlled parameter ES can be used in the implementation of the confirmation service as well. ES mechanisms use "public" keys, which are generated by the sender of data and verified by the recipient. Asymmetric encryption techniques can be used

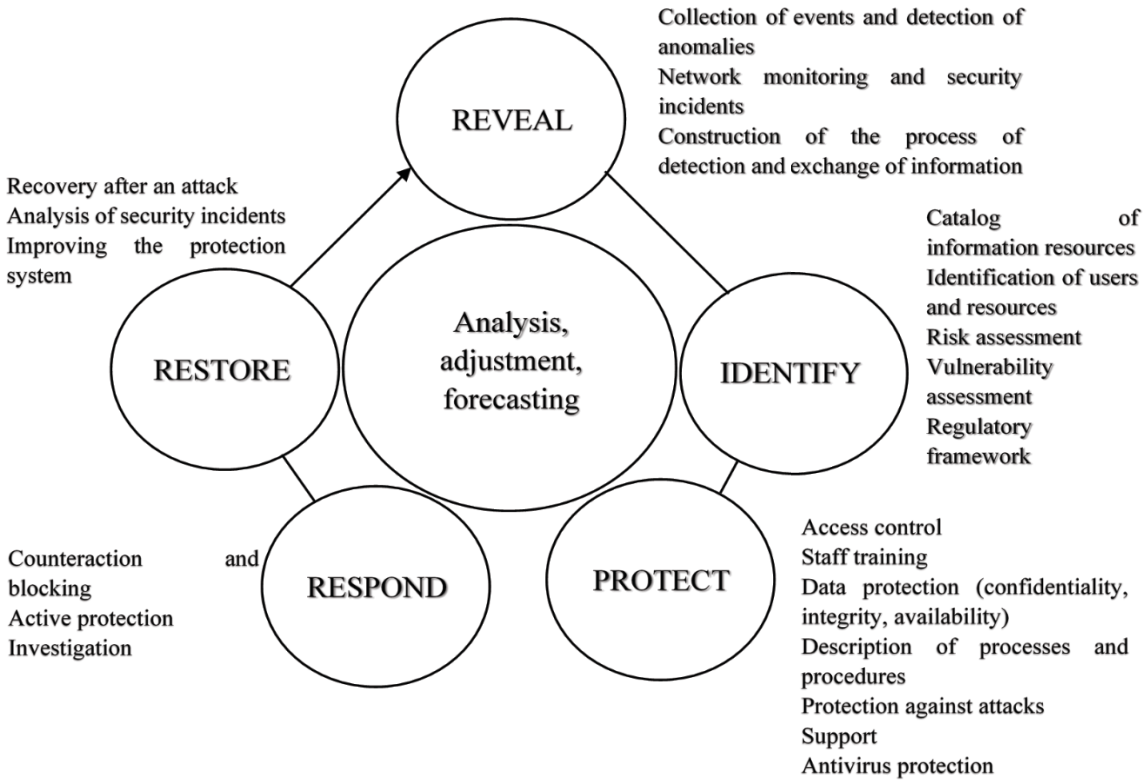


Fig. 3.6. Active cycle of cyber defence

to encrypt the checksum of the message being signed. ES consists of two procedures: the procedure of imposition (generation, formation) of the signature and the procedure of removal (verification, verification, destruction) of the signature. The use of "public" keys for the ES is intended to confirm the origin of the message but does not control the recipient of the message. Authentication tokens can be used to ensure user information security and secure remote access to information resources by an authorized user.

The means of technical protection of information (TPI) include hardware and software, the main functional purpose of which is to protect information from threats of leakage, integrity and blocking; technical means, in which, in addition to the main purpose, information protection functions are provided; tools specially designed or adapted for searching for embedded devices or for assessing the security of information. In the context of this definition of TPI, consider a number of modern digital cybersecurity tools. Firewall (in German – *brandmauer*) – a software or firmware device that monitors and filters network traffic passing through it following a set of pre-specified rules. Firewalls ensure the integrity of data or procedures that can only be accessed by authorized parties. The most common place to install firewalls is the perimeter of the local network to protect internal hosts from external

attacks (Fig. 3.7).

Intrusion Detection System (IDS) is a software or firmware designed to detect unauthorized access to a computer system or network, or unauthorized management, primarily over the Internet. The most common types of IDS are: (a) network intrusion detection system (NIDS) – a system that analyzes incoming network traffic; and (b) host-based intrusion detection system (HIDS) – a system that tracks important operating system files. IDSs can also be classified according to threat detection methods. The best known are: (a) signature-based detection (recognition of bad patterns, also known as malware) and (b) anomaly detection (detection of deviations from "correct" traffic, often through machine learning).

A firewall differs from an IDS in that it restricts access to a host or subnet to certain types of traffic to prevent intrusions and does not track intrusions that occur within the network. IDS, on the other hand, passes traffic by analysing it and signalling when suspicious activity is detected. An Intrusion Prevention System (IPS) is a software or hardware that is designed to detect, block, or tamper with a computer system or network. IPSs, unlike IDSs, are not limited to alert notification, but also take measures to block the attack (for example, disconnection). Information about any malware activity or malfunctioning provided by IDS or IPS is centrally collected by the Security Information and Event Management (SIEM) system. The SIEM-system solves the following typical problems:

- collection, processing and analysis of security events coming from various sources;
- real-time detection of attacks and violations of security criteria and policies;
- prompt assessment of protected information, telecommunications and other critical resources;
- security risk analysis and management;
- conducting investigations into incidents;
- making effective decisions on information protection;
- formation of reporting documents.

Antivirus software is a software designed to protect ITS objects/resources from being damaged by computer viruses. Antiviruses are classified according to the features that antivirus protection technologies are using: 1) classic antivirus (use only signature method of virus detection); 2) antivirus with proactive protection; 3) antivirus with combined program functionality (antivirus protection, additional spam

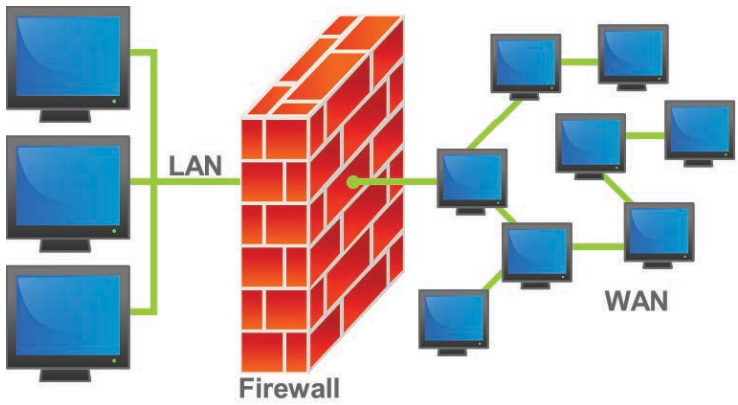


Fig. 3.7. Firewall

filtering, encryption and data backup); 4) target platforms (to protect workstations, file servers, mail and Internet gateways, virtualization servers), among many others.

Vulnerability scanners are software or firmware tools for diagnosing and monitoring network computers that allow to scan networks, computers, and programs for possible security issues, and to assess and resolve vulnerabilities. Vulnerability scanners allow to check for possible "holes" in the system that can be used by attackers. Low-level tools, such as a port scanner, can also be used to detect and analyse the applications and protocols running in the system at any instance. Such scanners are widely used in the so-called "penetration test", a method of assessing the security of a computer system or network by partially simulating the actions of external or internal intruders to penetrate it. This process includes actively analysing the system to identify any potential vulnerabilities that may arise due to incorrect system configuration, known and unknown hardware and software defects, or operational delays in procedural or technical countermeasures.

Honeypot ("trap") (from English "pot of honey") – a resource-bait for cybercriminals. The task of honeypot is to be attacked or unauthorized, which will later analyse the strategy of the cyber attacker and determine the list of means by which cyberattacks can be inflicted on real resources. Honeypot is actually a resource that does not perform any actions without any influence on it. Honeypot collects a small amount of information, which analyses the statistics of the methods used by cybercriminals and determines the availability of any new solutions that will later be used to combat them.

Sandbox is a mechanism for securely executing programs used to run untested code, unverified code from unreliable sources, and to run and detect viruses. Typically, a sandbox is a tightly controlled set of resources for running a guest program, such as disk space or memory. Some antivirus software developers now use sandbox in their products as a means of proactively protecting users from as yet unknown threats.

In the near future, Microsoft Corporation also intends to build into its OS protection against malware – tool Windows Sandbox [42].

We should also mention the classic software or firmware that implements the appropriate sets of requirements, rules, restrictions, and recommendations aimed at achieving and maintaining the state of information security, namely:

- tools with data leakage prevention technology (Data Loss Prevention, DLP);
- Identity and Access Management (IAM);
- address filtering tools (Uniform Resource Locator Filtering, URL Filtering) and content filtering software;
- tunnelling tools (Virtual Private Network, VPN);
- tools with technology for checking and filtering packets (Deep Packet Inspection, DPI), etc.

Based on the results of the analysis of the principles of operation of the digital cybersecurity tools discussed above and the mechanisms for ensuring their protection of information and the network, a generalized assessment of the directions of their applicability was made (Table 3.1).

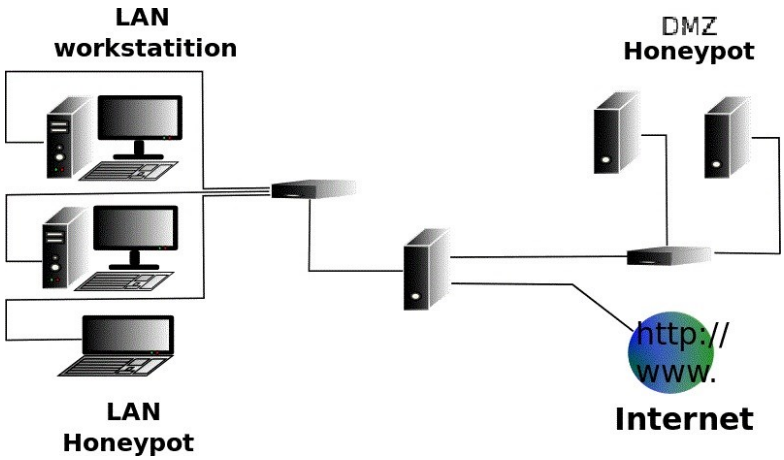


Fig. 3.8. Resource-bait "honeypot"

Table 3.1. Applicability of Modern Digital Cybersecurity Tools

Digital cybersecurity tool	Property			
	Confidentiality	Integrity	Accessibility	Observation
Cryptographic protection of information	+	+	+	–
Authentication token	–	–	+	–
Firewall	–	–	+	–
IDS	–	–	+	+
IPS	–	–	+	+
SIEM	–	+	+	+
Antivirus	–	+	+	+
Vulnerability scanner	–	–	–	+
Honeypot	–	–	–	+
Sandbox	–	–	–	+

3.4. The latest trends and prospects for the development of digital cybersecurity tools

The variability of the cyber threat landscape, the frequency of their occurrence, the complexity and target nature of cyberattacks require the evolution of existing cyber security rules and the transition to a combination of technologies to prevent, detect, and respond to cyberattacks.

Classic digital cybersecurity tools usually allow the detection of known cyberattacks, but their capabilities, unfortunately, do not always stop unknown attacks specifically designed to bypassing the existing protection system by changing signatures and patterns of behaviour. Traditional cybersecurity solutions in most cases have a protection based on signature files, detection mostly of known cyber threats, based, as a rule, on the behaviour of old cyber threats and cyberattacks (do not carry out in-depth monitoring of activity causal analysis), and fail to provide information about cyberattacks.

As a result, we see a growing gap in the detection of unknown cyber threats and cyberattacks. In particular, according to the Verizon Data Breach Investigations Report 2016, there is a steady increase in cyber incidents in which, for the system

to compromise, took a day or even several hours, and the threat can compromise the system in minutes or hours, while the system owner usually reacts after weeks, months or even years [43]. This suggests that cybercriminals are becoming more effective, and malware is becoming more sophisticated. Techniques for cyberattacks are also being developed, and cyberattacks have become targeted, coordinated, and are using a variety of vectors.

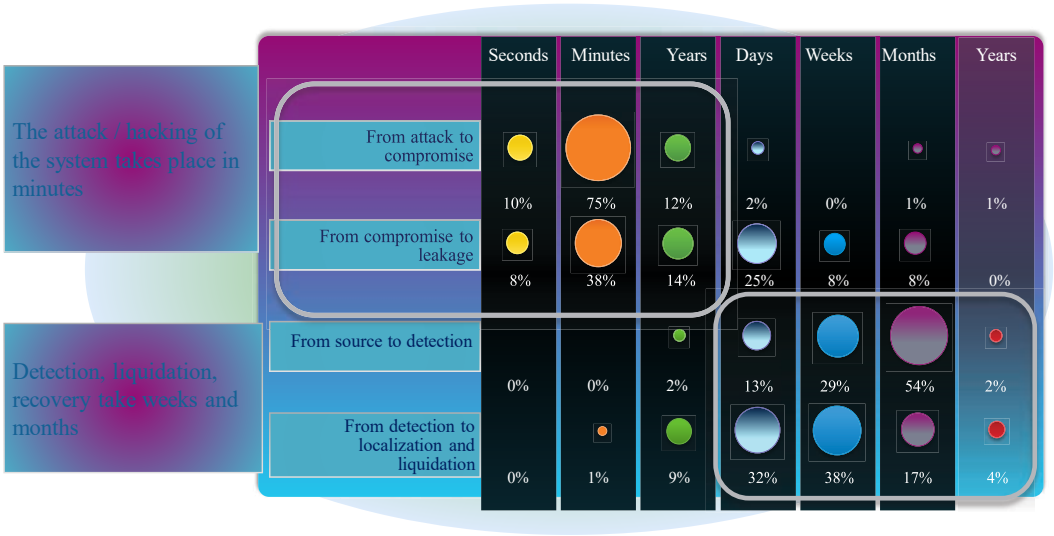


Fig. 3.9. Time scale of events in percetage of the total number of fractures

Therefore, we hear more and more about the solutions of cybersecurity "next generation", which include, in particular:

- Next Generation Firewall (NGFW);
- Unified Threat Management (UTM) systems;
- Next Generation Intrusion Prevention System (NGIPS).

Such systems are a security product that can perform multiple security functions within a single device, such as: network firewall, network intrusion prevention, antivirus gateway, anti-spam gateway, VPN, content filtering, load balancing, data leakage prevention and reporting (Table 3.2).

In the field of protection against modern cyberattacks, according to experts [44], one of the latest trends is the use of cybersecurity analytics systems using machine learning and artificial intelligence systems on large amounts of data (e.g., Machine Learning-based Security Analytics using Big Data). These systems allow the

Table 3.2. Comparative Analysis of Functions of NGFW, UTM and NGIPS Systems

NGFW	**UTM**	**NGIPS**
Firewall	Firewall	–
Program and relationship analysis and full stack visibility	Program and relationship analysis and full stack visibility	Program and relationship analysis and full stack visibility
Intrusion Prevention (IPS)	Intrusion Prevention (IPS)	Intrusion Prevention (IPS)
Unlock encrypted sessions, check encrypted packets, detect and block threats (SSL and SSH inspection)	Unlock encrypted sessions, check encrypted packets, detect and block threats (SSL and SSH inspection)	–
Packet checking and filtering (DPI)	Web content filtering	Content awareness
	Antispam, antivirus and antispyware protection	
Reputation -based malware detection	–	Content awareness
Quality of service (QoS) functionality	–	–
–	Traffic shaping / bandtwith control	–
–	Data Loss Prevention, DLP	–

detection of abnormal behaviour of systems or users and thus detect most dangerous cyberattacks. The use of such systems for cybersecurity tasks in the form of UEBA (User Entity and Behaviour Analytics) platforms is promising. Modern so-called detectors of abnormal staff activity increase the likelihood of detecting a motivated insider, which reduces the risk of cyber incidents.

There are also purely unique digital tools of cybersecurity, among which the most noteworthy are:

- Endpoint Detection and Response (EDR) systems [45];
- ESET LiveGrid technology [46];
- Arbor Peakflow SP technology to prevent DDoS attacks [47];

- technology of extended protection against malicious programs (Advanced Malware Protection, AMP) [48];

- software for cyber threat hunting [49];

- Cyber Threat Defense (CTD) technology based on network telemetry analysis to provide protection against cyber threats [50].

As you can see, there is a number of effective digital cybersecurity tools on the market.

Chapter 4:
Power Systems: SCADA System, Smart Grid, Simulation and Modelling, Wide Area Monitoring and Control

Nikolas Flourentzou
University of Cyprus

Co-authors:

Markos Asprou
University of Cyprus

Lazaros Zacharia
University of Cyprus

Lenos Hadjidemetriou
University of Cyprus

4.1. Introduction

The power system is an essential critical infrastructure which could be viewed as the backbone of any modern society impacting many of our everyday life activities [51]. Thus, it is heavily important to ensure that its operation is reliable and secure. The problem of controlling and managing the power system is becoming more and more difficult because its size (due to the growing demand) and complexity (due to the penetration of renewable energy sources, the use of electric vehicles and the complication of the electricity market) are increasing. Thus, developing models to accurately emulate its behaviour, it is required to predict failures and physical or cyber-physical attack scenarios [52], [53].

This chapter provides a review of SCADA system, smart grid, simulation and modelling, wide area monitoring and control.

4.2. SCADA systems

The Supervisory Control and Data Acquisition (SCADA) is an industrial-level system used for automated control, for monitoring and analysing the grid, and generally, for supporting Critical Infrastructure Protection (CIP).

SCADA systems have a centralised architecture approach and communicate with sensors and actuators to provide remote access to the operators through a specialised interface. Synchronised information is collected in databases and managed at the control room.

SCADA systems are formed by the following apparatuses:

- Human Machine Interface (HMI) as Graphical User Interface (GUI);
- supervisory system and Master Terminal Unit (MTU);
- Remote Terminal Units (RTU) such as sensors and attenuators;
- Programmable Logic Controllers (PLC); and
- communication infrastructures using specific protocols.

A simple diagram of a typical SCADA system architecture is shown in Fig. 4.1, where the main parts of the system are demonstrated.

4.2.1. Human machine interface

The HMI processes all the data collected by the SCADA system through the Remote Terminal Units (RTU), Programmable Logic Controllers (PLC) and other control devices for transforming them into valuable information, which can be easily understood by humans. HMI presents these data to operators which can observe the status of the infrastructure in real-time, in terms of custom mimic displays, faults, alarms, warnings, and trends to make decisions for adjusting any controls and/ or settings. HMI is also connected to other technologies, such as data servers, to allow for historical trending and further analysis. The HMI can generate report logs

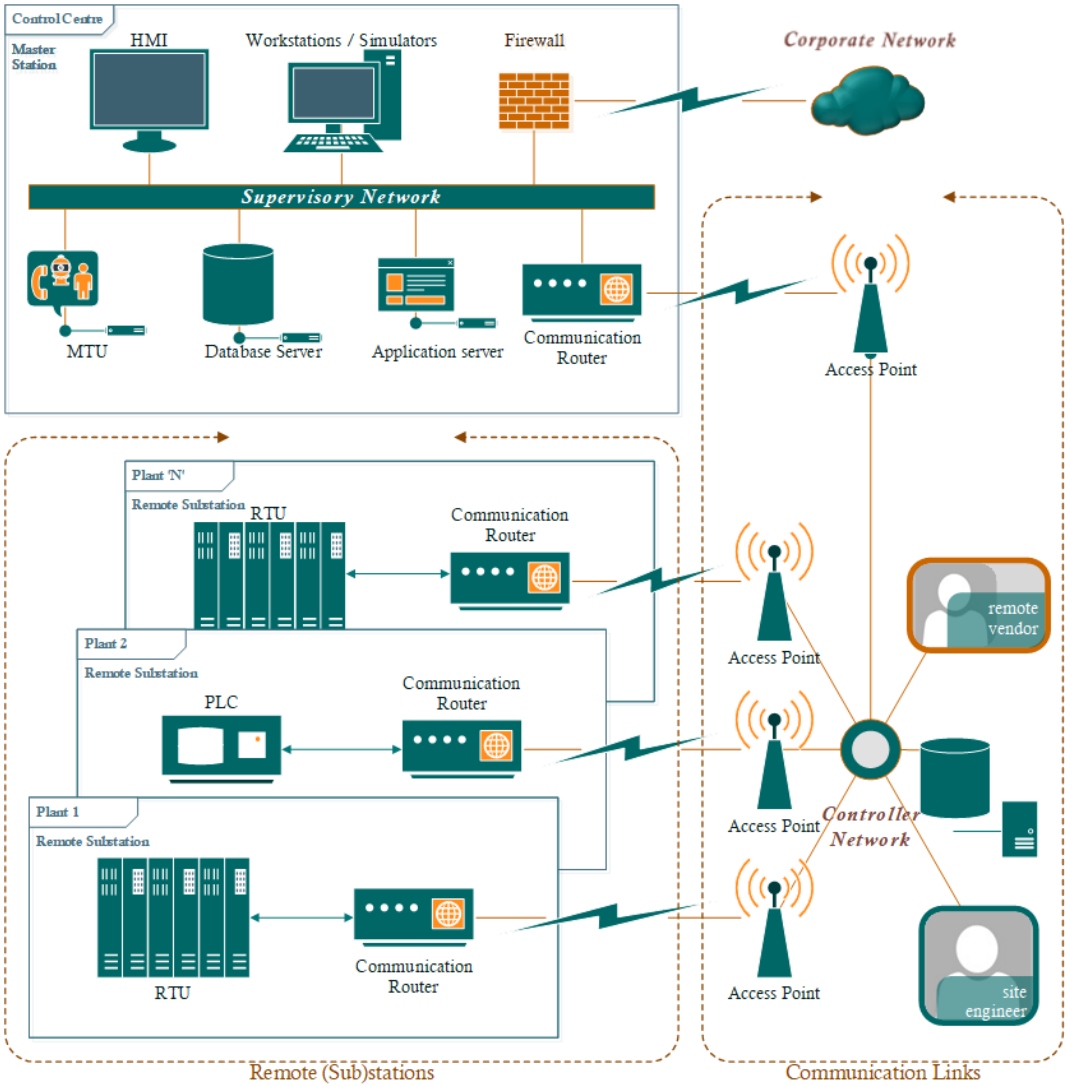


Fig. 4.1. SCADA system architecture

that summarise these historical trends to indicate possible future problems of the infrastructure and its network. The HMI can also offer geospatial data representation (information about geographic locations that can be stored in and used with a geographic information system), quick troubleshooting, filtering of nuisance alarms (that can desensitise operators to alarm critical reports), longer uptime of equipment, among other additional functions.

4.2.2. Supervisory system

The supervisory system provides the control logic of the SCADA system by gathering data on the process and sending commands to the process. The supervisory system contains distributed software applications, multiple servers usually distributed in several locations to enable recovery in case of disasters. Multiple servers are configured with redundancy in a hot-standby formation, which uninterruptedly monitor and control the field devices in case of a server failure to increase the integrity of the system.

The supervisory system includes the Master Terminal Unit (MTU), which is responsible for gathering data from and sending commands to the remote stations and substations. The MTU transmits the data to the HMI and the servers of the control centre for further exploitability by the operators. Each set of data is stored. The supervisory system of large SCADA systems may have the ability to handle complex control algorithms for performing model predictive control, nuisance alarms filtering and other functions.

4.2.3. Remote terminal units

Remote terminal units (RTUs) receive measurement signals from and send control signals to the intelligent electronic devices (IEDs). RTUs exchange data and commands with the MTU through the communication links of the controller network. RTU has the capability to communicate on a peer-to-peer network with other RTU and programmable logic controller.

4.2.4. Programmable logic controller

Programmable logic controllers (PLCs) are used for gathering data from sensors and actuators. PLC includes a microprocessor that can provide control logic.

To realise continuous, reliable, and efficient communication between the SCADA components, specific communication protocols have been developed. The main considerations for designing the specifications of the protocols are the processing capabilities of the components and the communication requirements of the required applications. Each SCADA network implements a precisely defined set of protocols. The most commonly used types of communication protocols are Modbus, IEC 60870-5 series, Distributed Network Protocol (DNP3), IEC 61850 series, and EtherNet/IP.

The special need of these protocols is their urge for strong cyber-security measures, such as industrial firewalls, dedicated VPNs, cryptography, elimination of cyber and cyber-physical threats, blocking unauthorised access, multiple-level authentication, specialised protocols, proprietary interfaces.

4.3. Smart grid

Environmental concerns about carbon dioxide (CO2) emissions impose several changes to the traditional operation of power systems. Therefore, the power generation needs to be based on distributed renewable energy sources (RES) in order to minimise the use of fossil fuels by the traditional generation plants. Furthermore, the electrification of the transportation sector, by the massive deployment of electric vehicles (EVs), in combination with the electrification of thermal loads, by the introduction of air-conditioning and heat-pump devices can minimise the environmental impact; however, the load significantly increases the demand side of a power system. All these changes impose several challenges to the power system operation, and thus, the infrastructure needs to be upgraded with wire and non-wire solutions.

4.3.1. Wire and non-wire solutions

Wire solutions include traditional investments in infrastructure to upgrade or replace wires, poles, and power transformers for increasing the stability limits of the systems and relieving the operating conditions. However, non-wire solutions are alternative solutions to avoid traditional infrastructure expansion and increase the RES hosting capacity of the grid [54]. In the framework of smart grids, these solutions include changes in the operating practices, introduction of intelligent schemes for monitoring and control of the power system, exploitation of the advanced functionalities by the power electronics converters, and utilisation of flexibilities provided by the energy storage systems in order to minimise the congestion, relieve the operating conditions, increase the utilisation of existing grid capacity, enhance the stability and power quality, and increase the hosting capacity of RES without causing problems in the operation of power systems.

4.3.2. Structure and communication requirements

In the framework of smart grids, the development of advanced monitoring and control schemes requires that the system operator should be able:

- to receive measurements for the system operating conditions;
- to process these measurements based on intelligent and automated monitoring and control schemes to enhance the situational awareness and to improve the operational capabilities of the system; and

- to send coordination signals to the flexible actuators (i.e. generators, transformers, RES, load, etc.) to manage the overall system operation.

At the transmission system, the exchange of measurements and coordination signals is achieved through the SCADA system, as already explained in Section 4.2. In this case, in each substation of the transmission grid, there are RTUs, PLCs, protection relays, and Phasor Measurement Units (PMUs) that concentrate voltage and current waveform measurements at each bus and each line of the system, process these measurements to calculate the voltage and current phasors, the active and reactive power, and the frequency. The resulted quantities can be used locally to enable some intelligent functionalities to regulate the operation of the system or protect the system against grid failure, or can be reported to the control centre of the SCADA to allow the operator to manage the overall operation of the system by sending back some coordination actions to the substation. Communication between the control centre and the substation is facilitated through dedicated or shared network communication infrastructure. It should be noted that since the number of the primary substations (HV/MV) in the transmission level is limited compared to the number of the secondary substations (MV/LV) in the distribution level, it is feasible to have communication infrastructure between all the primary substations and the control centre of the Transmission System Operator (TSO). However, this is not the case in the distribution level.

In distribution level, the monitoring and control of all the resources requires communication with all the secondary substations (MV/LV), which are hundreds to thousands of times more than the primary substations. Additionally, if the Distribution System Operator (DSO) wants to be able to monitor the operating conditions within the Low Voltage (LV) distribution grids, then there is a need to concentrate measurements from millions of smart meters installed at each consumer. Furthermore, if the DSO wants to take control and coordination actions for regulating the operation of its system, the communication between the DSO control centre and each distributed resource (i.e., wind power plant, large photovoltaic systems, residential photovoltaic systems, energy storage systems, EV charging station) should be established. Therefore, the need for communication between the DSO control centre, each secondary substation, consumer, and flexible resource is a major challenge for the deployment of the smart grid concept that needs to be overcome. Several solutions that include power line communication, radio communication, wired and wireless communication exist, where in combination with Information and Communication Technology (ICT) this communication can be enabled for smart grids. However, in an environment where measurement and set-points are exchanged between millions of devices, the cyberattack surface increases, and therefore, security solutions should be introduced to ensure the security, integrity, and reliability of the power system infrastructure.

4.3.3. Challenges and opportunities

Since the communication in the transmission grid is already established in the majority of the cases, the introduction of smart grid solutions to increase situational awareness and stability of the power system is feasible. Especially with the introduction of PMU technology in transmission level this can be achieved by wide area monitoring and wide area control solutions that will be further explained in Sections 4.4 and 4.5, respectively. Furthermore, at the transmission level, the TSO should be able to control the active and reactive power by large-wind power plants or large-scale Energy Storage Systems (ESSs) to provide flexibilities to the system operation. The reactive power control of wind power systems can enhance the voltage stability of the system, while the active power control is only applied in form of curtailments, in case of emergencies, to ensure the system's balance and stability. Furthermore, the flexibility provided by ESS can be used to compensate the intermitted and unpredicted nature of RES and to provide ancillary services to the system for enhancing its stability and flexibility. For example, an ESS can be combined with a wind power plant to allow the combined system to generate a controllable power profile that is not creating disturbances to the system operation. An ESS can also provide frequency reserve services and in case of emergency (e.g., frequency drop) can discharge in order to support frequency of the system. Furthermore, an ESS can also be used for energy shifting applications to shift a part of the demand or generation in another time window within a day, which can be beneficial for the unit commitment and economic dispatch of the traditional generation plants, and as a result electricity costs can be reduced. The abovementioned provide some opportunities where RES and ESS can be intelligently utilised in transmission level to improve the operation of the system within the framework of smart grids.

In the distribution side, there are significantly more challenges and opportunities in the case of smart grids. First of all, the majority of distribution grids are not observable by measurements. There are cases where the Medium Voltage (MV) distribution grid is observable by traditional measurements obtained by the secondary substations and concentrated in a SCADA system; however, this is not the case in the majority of distribution grids. Furthermore, the deployment of smart meters at the consumers side is still limited. Even in cases where these smart meters have been massively deployed, these measurements are typically used for billing purposes. In the smart grid concept, the massive deployment of smart metering technology by consumers provides a great opportunity to increase the situational awareness of both LV and MV distribution grids. Information and communication technology can be used to automate the concentration of measurements by the smart meters. State estimation [55] and power flow [56] schemes can then be used to monitor the LV distribution grids. The monitoring of LV distribution grid can be further used to partially monitor the MV distribution grid, since the power absorbed by each secondary substation and the voltage at the LV side of the distribution transformer can be utilised for calculating the operating conditions at the MV distribution grid.

Another important aspect in the operation of future distribution grids is the intense utilisation of power electronics converters. Power electronics-based converters (i.e. inverters, rectifiers) are used for the integration of PVs, ESS, EVs and efficient electrical appliances. These power electronics inverters offer increased controllability to the operating conditions and can be enhanced with additional functionalities to support the distribution grid operations. Furthermore, converters are Internet of Things (IoT) enabled devices and can be used for obtaining measurements for the PV generation or the load consumption and for the voltage conditions at the point of coupling. The IoT capabilities can also be used to send coordination signal to control their operation and as a way to manage the overall operation of the grid by the operator. The PV or ESS inverters, for example, can be coordinated by a centralised scheme to provide reactive power compensation [57] in a LV distribution grid or a micro-grid to achieve a unity power factor. In a more intelligent approach, advanced inverters can be used to provide both reactive power compensation and phase balancing services [57], [58] to additionally symmetrise the operation of distribution grids, which often is highly asymmetrical since the majority of loads are single-phase connected. Such intelligent schemes can maximise the efficiency, power quality, and utilisation of the distribution grid.

Another important aspect to enable the voltage and frequency stability of the power grid is to utilise the power electronics-based PV and ESS to support the stability of the system under normal operation and under grid disturbances and failures. Under normal operation conditions, different reactive power injection schemes for PV inverter can be used to compensate the voltage variation across the feeder. These schemes are well known as cosφ(P), cosφ(P,V), and Q(V) [59], [60] and are able to regulate the reactive power injection by PV according to the power generation and/or the voltage conditions at the point of coupling, based on local measurements. Other optimisation schemes propose also the online coordination of the reactive support by each inverter by a central controller, where increased communication and observability requirements are needed in order to achieve global optimal conditions for the distribution grid. In either way, the distribution grid voltage variation imposed by the reverse power flow, which is caused by the intense distributed generation, can be compensated by avoiding violation of the permissible limits of the voltage at the distribution feeders. Under grid disturbances (i.e., a voltage sag event), the PV or ESS should be able to keep the synchronisation, remain interconnected to the system, and provide adequate voltage support to the system to maintain the voltage stability of the grid [61]. Especially in the case of ESSs, this support scheme can also be used to provide both voltage and frequency support in cases where there are disturbances in either the voltage and/or the frequency [62]. Such fault ride through support scheme increases the system capability to maintain the stability under grid disturbances, especially as the penetration of distributed PV and ESS is increasing.

The ESS provide flexibilities to the power system and can be used in both the transmission and the distribution grid. Besides the use of ESS in combination with RESs to smooth out the variation imposed by the unpredicted nature of RES, in the

distribution grid, the ESS can provide energy shifting and peak shaving services [63] at local level to relieve the operating conditions of the distribution grid and potentially to increase the hosting capacity of RES. Additionally, the ESS can be used at the consumer level to increase the self-consumption in buildings [64] that are equipped with photovoltaic systems, in a way to reduce the interaction with the grid and to enable higher penetration of PVs in distribution grids. In these cases, proper optimisation schemes can be developed for ESS [65] to minimise electricity cost for consumers, especially under a variable electricity pricing scheme. The methods mentioned above [64], [65] highlight the fact that the ESS can also be used to extend the lifetime of power electronics converters to increase the competitiveness of such green technologies.

Another important aspect that can impose significant challenges on the distribution grid operation is the electrification of the transportation sector by the massive deployment of EVs. This will introduce a tremendous increase on the load for charging the electric vehicles, which can cause significant violation on both thermal and voltage limits at the distribution grid. Therefore, instead of upgrading the infrastructure with traditional wire investments, several methods have been proposed [66] to allow the intelligent coordination of EV charging station to avoid violation of the permissible limits of the distribution grid.

The smart grid concept introduces several solutions that are not based on traditional investments for upgrading the infrastructure in order to enable the secure and reliable operation of the grid under new circumstances. These solutions exploit new functionalities and flexibilities by the key component of a future power system to increase the efficiency, power quality, and utilisation of existing grid capacity. The integration of these solutions under the framework of smart grids is facilitated through information and communication technology and can address the main challenges that are faced by future power systems towards a green and sustainable evolution of the power system infrastructure.

4.4. Wide area monitoring

The advent of Synchronised Measurement Technology (SMT) and its rapid deployment in the power system enables the transition from the existing (SCADA/EMS) system to a Wide Area Monitoring (WAM) system. The difference between the WAM system and the conventional schemes of power system condition monitoring is that the WAM system consists of evolutionary applications able to track the dynamics of the power system; therefore, information about the operating conditions of the power system is available in quasi real time. The output of the WAM applications is forwarded to the control applications for controlling the power system. In the case of a fault occurrence, control applications are responsible for mitigating the fault by taking preventive actions. Thus, the importance of the WAM system in the conservation of the power system operating situation in proper limits implies that consistency, accuracy and robustness should be the main features of a WAM system.

The implementation of a WAM scheme is not an easy task, since the WAM design poses economic, operating and technical challenges. This subchapter aims to provide the reader with a general idea about the components of a WAM system and to outline the steps for the successful and cost effective WAM system implementation.

4.4.1. Components of a WAMC system

4.4.1.1. Phasor measurement unit

The key element of a WAM system is the phasor measurement unit (PMU), which is currently the most advanced measurement device in the power system measurement layer. The configuration of a generic PMU that was first introduced at Virginia Tech is shown in Fig. 4.2. Although commercial PMUs today may have different structure or components, the main idea for extracting the phasor measurements remains the same. According to Fig. 4.2, the voltage and the current analogue quantities entered in the anti-aliasing filters after passing through the voltage and current transformers for stepping down their magnitude. The anti-aliasing filters are used before the Analog to Digital (A/D) converter to prevent the aliasing of the signal when it is reconstructed by its discrete samples. To satisfy the Nyquist criterion, the cut-off frequency of these filters is usually chosen to be half of the sampling frequency of the A/D converter. The filtered analogue quantities then pass through the A/D converter for being discretised. The discretisation process is dictated by the signal received by the Global Positioning System (GPS). The PMU is equipped with a GPS receiver

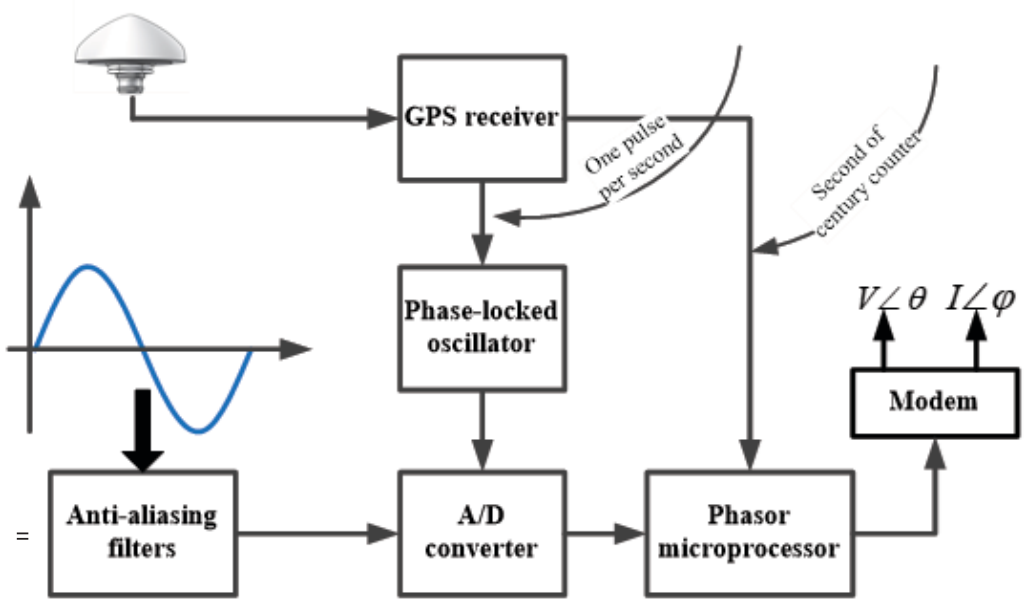


Fig. 4.2. Generic PMU configuration

in order to receive the one pulse per second that is provided by the GPS. Essentially, the GPS signal achieves the synchronisation of the PMUs that are situated in different locations, since all the PMUs receive the same pulse per second with a 1 μs accuracy. More specifically, one pulse per second sets the beginning of the signal sampling for all the PMUs, facilitating the phase comparison of the current and voltage quantities. Initially, the sampling rate of the analogue signal was 12 samples per cycle, and therefore, a synchronised sampling procedure was required (i.e., sampling the signal at the same time instants). Recently, the sampling rates have been significantly increased, and thus the accuracy of the calculated phasor measurements has also been improved.

The samples generated by the A/D converter are the input to the phasor microprocessor that calculates the voltage and the current phasors. The algorithm used in the phasor microprocessor may differ depending on the manufacturer; the principle is usually based on the Discrete Fourier Transform (DFT) or the Fast Fourier Transform (FFT), although some recent works propose the use of Kalman filter approach [68]. Special attention should be given to the window of samples that is used for applying either the DFT or the FFT algorithm, since either the voltage or the current analogue signal changes according to the power operating condition. In the phasor microprocessor, the frequency of the sampled signal, as well as the Rate of Change of Frequency (ROCOF), is estimated. An additional advantage of the PMU over the conventional measurement devices is the provision of time stamped measurements. The time stamp of the phasor measurements is done on the phasor microprocessor, which receives the second of century, and the fraction of second signal from the GPS receiver. The Second of Century is a 4-byte time message that contains the number of seconds that occurred since January 1, 1970. The time stamp of the phasor measurements contains also the time quality of the measurements. Since the phasor measurements of the PMUs are synchronised and time stamped, they are also referred to as synchrophasors in the literature. Finally, PMU measurements are sent either directly to the control centre or to Phasor Data Concentrator (PDC) via the modem that is accommodated in the PMU [67].

4.4.1.2. Phasor data concentrator (PDC)

A phasor data concentrator gathers and time aligns the measurements provided by the PMUs. As soon as a phasor measurement set that contains the phasors with the same GPS time stamp is completed, the PDC forwards the phasor set to the central control centre for usage by the WAM applications. A data concentrator can be found either at a local control centre for collecting phasor measurements from regional PMUs or at the central control centre of an electric utility for communicating and collecting the phasors from the regional PDCs as shown in Fig. 4.3 [67]. The communication between the PMUs and the PDCs in the same region is achieved through a LAN (Local Area Network), which in some cases is bidirectional, while the PDCs utilises a WAN (Wide Area Network) for communicating with each other or with the central control centre, and the communication flow is bidirectional. Further

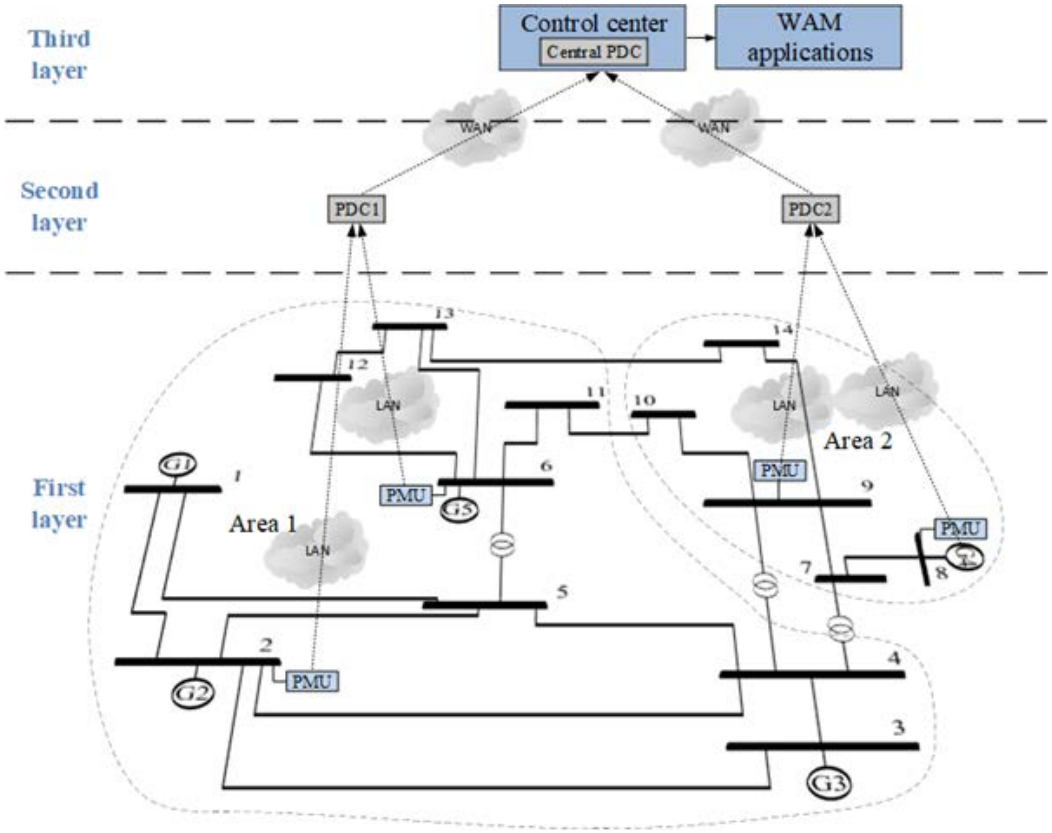


Fig. 4.3. Measurements system comprising both PMUs and PDCs

discussion regarding communication issues is given in the following section. In addition to gathering and aligning the phasors, PDCs sometimes contribute to other applications, i.e., archive the phasors utilised for offline applications and check for errors in the derived synchrophasors, before forwarding the set to the upper level.

4.4.2. Communication infrastructure

Communication infrastructure plays a crucial role in the implementation of a WAM system. The rapid advancements in the technology of fibre optics increase the data transfer capability in terms of amount and latency, enabling the transfer of large amount of data between the power system field and the central control centre in quasi real time. Nevertheless, in the context of communication infrastructure issues regarding the communication protocols, the amount of data transferred, the data quality, and the data latency need to be addressed carefully in the design of a WAM system, since different WAM applications have different communication capability requirements.

4.4.1.3. Communication network

The communication network of a WAM system has great impact on the overall performance of the system. One of the main concerns in the design of a WAMC system is the choice between two orthogonal approaches regarding the communication network; the first is the utilisation of the already existing Internet network as a basis of an intra power network and the second is the design of a scalable communication network dedicated to the power infrastructure. Electric utilities facing the above dilemma must consider the advantages and disadvantages of the two proposed communication architectures and decide accordingly.

The choice of utilising the Internet network poses some drawbacks that could prove disastrous in the power system operation. The first challenge that needs to be addressed carefully is the security of the measurements provided by the PMUs. Most of the PMU measurements are often transmitted unencrypted, which can be easily intercepted by internet attackers during the transmission. Of course, this could be compensated using encryption schemes; however, this solution implies more bits in the measurement frame, and therefore, the amount of transferred data increases. Moreover, some WAM applications need reliable and real time data without large delays; therefore, it is arguable whether an internet communication channel can deliver reliable and constrained time delay measurements to the WAM applications. Besides, knowing the functionality of internet-based protocols (TCP, UDP, RTP), which drop or delay data packets randomly, the reliable and consistent delivery of measurements in monitoring applications of a WAM system based on an internet communication channel is doubtful. Further, internet communication channels are often highly congested, and as a result they can lose lose data packets. Although in a WAM system the loss of measurements is expected, with an internet-based communication network, the measurements loss may occur occasionally, bringing forth large gaps of time discontinuities between two sequential measurements arrived in the control centre.

The aforementioned drawbacks of an internet-based power communication network can be overcome using a dedicated power communication network in which channel congestion, data reliability and delay issues can be handled by the electric utilities that own the particular WAMC system. On the other hand, the expansion of a WAMC system in a large territory implies increased cost in the design from scratch of a dedicated communication network for a WAMC system; while the cost of implementing an internet-based communication network is considerably lower. It is therefore not an easy decision for the electric utilities to choose between the two network categories since there is a trade-off between cost implementation and reliable, consistent and secure measurement delivery [69].

Table 4.1. Associated Delay of Different Communication Links [70]

Communication link	Approximated delay-one way (ms)
Fibre-optic cable	100–150
Digital microwave link	100–150
Power line	150–350
Telephone line	200–300
Satellite link	500–700

4.4.1.4. Data latency

The great advantage of the WAM system over the existing SCADA system is the response in power oscillations, voltage and frequency instability in real or quasi real time. It is therefore of great importance in the acquisition of real time data from the power field with minimum delay since some contingencies need to be prevented within milliseconds from the time of their occurrence. On the other hand, other relative communication issues such as data transfer capability, communication protocols, and data quality are implicated directly to the determination of the data latency. Therefore, achieving optimal data latency in a WAM system is like tuning the aforementioned communication issues.

The sources of data delay in a WAMC system until data reach the control centre are the transducer delays when the potential or current transformers measure the rms current and voltage of the busbar, the windows size of the DFT used in the PMU for digitising the voltage and current analogue signal, the phasor calculation time, the size of the PMU data output, the phasors aligning in the PDCs and the latency due to the communication link.

Many of the above delays are unavoidable, particularly those that come from the devices' hardware, however, the delay due to the communication link can be minimised by choosing the link whose data transfer delay is at the minimum. In Table 4.1 several communication links used in communication networks are tabulated along with their associated delay when they are used in a WAM communication network. As it is shown, the fibre optic and the digital microwave link exhibit the smallest data latency among the other mediums of data transfer. The rapid advancement in the manufacturing technology of communication links promises further improvement in the data transfer delay, and perhaps in some links the delays in the data transfer will be negligible. Therefore, the power system operators need to choose between the capabilities that each communication link offers and the cost associated to each link.

In both the communication delay and communication network section, it can be concluded that there is a basic trade-off between the cost of the communication infrastructure and the capabilities provided by it. It is, therefore essential in the

creation of a roadmap by the electric utilities for a successful and cost-effective WAM implementation.

4.4.3. Roadmap for a successful WAM system implementation

The implementation of WAMC system requires a well-designed plan, perfect knowledge of WAM applications, awareness of new applications and their immediate integration in the existing WAM system, and affordable WAM implementation. Therefore, the policy that many stakeholders follow today is the gradual implementation of a WAM system according mainly to the available budget and to the applications needed to be executed by a WAM system. For a successful and cost-effective implementation of a WAM system, electric utilities must always design a roadmap containing not only the immediate actions needed but also a future plan indicating the future design requirements. In this part, some essential issues are presented that need to be addressed carefully by the power system operators before implementing a large area WAMC system..

4.4.3.1. Consistent performance of all measuring units

It is quite possible in a WAMC system to have installed PMUs by different manufactures or to have different types of PMUs from the same vendor. This could happen mainly because the WAM system expands in a large territory incorporating more than one stakeholder. Each stakeholder follows its own purchasing decisions by satisfying certain electric utility policies, being consistent with an agreement with a particular vendor, or declaring a purchasing competition and selecting the most affordable offer. Hence, the inconsistent performance of the PMUs designed by various manufacturers in a WAMC system certainly deteriorates the overall performance of a WAMC system.

To address adequately the performance inconsistencies of PMUs by different vendors, two IEEE standards were established. The latest version of Standard IEEE C37.90 was published in 2005 containing issues regarding relay and relay systems used for power system protection and control. The IEEE C37.90 provides the standard conditions of services, ratings, performance and testing requirements for relay and relay systems. The IEEE C37.118-2005 substitutes the original IEEE 1344-1995 standard; it is complementary to IEEE C37.90 in the sense that the IEEE C37.118-2005 addresses relays with synchronised measurement capabilities. An important aspect that was not addressed in IEEE C37.118-2005 was the dynamic behaviour of PMUs in the power system transient conditions; therefore, an updated standard, namely EEE C37.118-2014, was published in 2014. Hence, it is a necessity for electric utilities to incorporate PMUs in their measurement systems that conform to the IEEE standards. Deploying PMUs from different manufactures that are “speaking the same language” ensures the improvement of the WAM system’s overall performance.

4.4.3.2. WAM design strategy

The design of the WAM system requires future planning. It is essential to foresee future applications that can be integrated into the already existing WAM system without major changes in its structure. The architecture of the WAM should not focus only on the "low hanging fruit" applications ignoring the future power system protection requirements. For example, the deployment of PMUs in the power systems is still restricted compared to the power systems scale. It is, however, certain that in the next few years, with the PMU cost decreasing, the integration of PMUs to the power systems will rapidly increase. Therefore, stakeholders should consider the large availability of PMUs in the future and design their WAM system accordingly.

The power system operators should design the architecture of a WAM system to support multiple applications. It is essential to know the exact requirements for each application and act accordingly. As it was already mentioned in the section on the communication issues, a dedicated communication network with communication channels of minimum delay and large transfer capability is the ideal communication infrastructure for a WAM system, but it is an expensive investment for the electric utilities. Therefore, a good strategy for electric utilities is to classify the intended applications for implementation in a WAM system according to their communication requirements. It is then easier to decide for which applications a dedicated network is a more appropriate solution and which applications can be implemented with a shared communication network. Further, it is advisable to incorporate all mature technologies in a WAM system since they add reliability and consistency to the overall system performance. However, stakeholders should be aware of any upcoming technological advancements and if necessary, to adapt them appropriately in the system. Therefore, a WAM system should be highly scalable and flexible to changes either concerning its infrastructure or its technology.

4.4.3.3. Efficient data management

One of the basic differences between the PMU and a conventional relay is the large amount of data provided by a PMU. According to IEEE C37.118-2014, each PMU has the capability to provide up to 100 frames per second; hence in a WAM system incorporating hundreds of PMUs, the amount of data flow from the field to the central control centre is huge. Therefore, the management of data efficiently is critical for the successful operation of the system. In this direction, the fact that many applications make use of data from the same locations should be taken into consideration by electric utilities. The classification of applications that use the same phasor data (i.e., from the same locations) could contribute to the effective data traffic management by limiting the times that a batch of data is transferred to the control centre. In addition, the data storage in the control centres poses a challenge to the design of a WAM system. Since the amount of data received by the control centre of a WAM is remarkably large, the design plan of a WAM system should include large storage facilities. It is therefore preferable to classify the data that will be used in real time applications such as voltage instability assessment, power system oscillations and phase angle monitoring, and

post mortem applications such as power system model validation and power system dynamic recordings. This data classification should determine the type of data that will be used directly and the type of data that should be stored.

4.4.3.4. Wide area monitoring power system applications

The WAM applications make use of measurements provided by PMUs in order to provide a quasi-real-time visualisation of the power system operating condition. The main objective in the implementation of the WAM applications is to overcome the weakness that conventional monitoring applications exhibit in tracking of dynamic behaviour of power system during the fault occurrence.

Furthermore, some monitoring applications require a fully observable power system to provide reliable results. Therefore, the observability analysis prior to the monitoring applications execution is necessity. In this section, a general overview of the main WAM applications, such as real time state estimation, frequency instability assessment, and voltage instability assessment, will be conducted.

- **State estimation:** The state estimation application constitutes the cornerstone of a wide area monitoring system. It makes use of redundant measurements in order to conclude with an estimation of the power system state vector. A strict prerequisite in the state estimation procedure is that the power system should be fully observable by the acquired measurements; if not, the state estimator could not obtain a unique solution of the state vector [67]. The state vector contains the states of the system (i.e., bus voltage magnitude and bus voltage angle), and its determination leads to calculation of real and reactive power flows for each transmission line, net real and reactive power injected to the buses, as well as to monitoring the voltage magnitude of the power system buses to be within their operational limits. Further, the state estimation results could be used for detection and elimination of gross errors contained in the field measurements before their usage by other applications. Therefore, the reliance of many applications on the state estimation results indicates the importance of the state estimator and in parallel necessitates the presence of a robust, consistent and accurate state estimator in the WAM system. The high reporting rate of the synchronised phasor measurements has given the initiatives to the development of a real time state estimator able to observe the dynamic behaviour of the power system operation.

- **Frequency stability assessment:** The deviation of the frequency from its nominal value can be caused mainly by an abrupt generator loss, and therefore the power demand cannot be satisfied (under frequency situation) or load loss where electricity production exceeds the power demand (over frequency situation). In both cases, the frequency becomes unstable leading to many potential problems in the power system operation. The aim of the frequency stability assessment is to identify whether the power system will

avoid cascading failures resulting from the sustained energy unbalances. The late detection of frequency instability could lead to a total blackout. The main drawback associated with the conventional frequency stability assessment applications accommodated in the SCADA systems is that the information extracted by the steady state view is limited. A common approach used in many SCADA systems to compensate frequency deviation is the Under Frequency Load Shedding (UFLS). However, some assumptions, such as the system inertia, load frequency sensitivity and voltage sensitivity, are known, thus making the UFLS relays operate unreliably. In this sense, real time monitoring of the frequency deviation is an immense need in modern power systems.

- **Voltage stability assessment:** The voltage instability in a power system occurs by the attempt of the load dynamics to restore power consumption while the power generation and transmission system is not able to compensate this consumption. The voltage drop in the power system occurrs mainly because of inability of the power generation system to satisfy the reactive demand. A power system is assessed as voltage stable when all the voltage magnitudes of the buses remain within the acceptable limits after a disturbance. The voltage instability could either occur in a region of a power system affecting only the voltage of the regional buses or it may have a cascading effect, causing a voltage drop in other buses below the acceptable limits. The importance of keeping the power system stable in terms of voltage is evident by the recently occurred blackout. The main objective of the voltage instability assessment application is to monitor the voltages of the buses in real time providing an indication of whether a region is critical to become voltage unstable.

4.5. Wide area control

Inter-area oscillations are causing major challenges that modern power systems have to confront since their appearance is increasing due to the multiple changes and constant expansion of the power systems. The appearance of inter-area oscillations creates many issues, such as the degradation of the power quality, the limitation of the transmission system capacity, and in several occasions, it can even lead the system to instability. Typical local controllers utilised for damping any undesirable oscillations, fail to compensate them due to the lack of global observability.

Due to the abovementioned issues, the necessity to develop new practices and methodologies, which will be able to firstly detect and then to compensate the inter-area oscillations, has emerged during the last few years. The detection of the inter-area modes was not possible until the advent of the synchronised measurement technology. More specifically, the deployment of Phasor Measurement Units (PMUs) in the transmission level and their capability in providing synchronised, near real-

time measurements from remote locations across wide areas (e.g., more than one TSO jurisdiction) has made the observation of inter-area oscillations observable and achievable. In addition, the widespread deployment of PMUs in power systems has laid the foundations for the development of Wide Area Monitoring and Control (WAMC) applications. The exploitation of the synchronised measurements to provide feedback control and coordination to the system, with the aim of compensating all the inter-area modes, has led to the introduction of the Wide Area Control (WAC) concept. The important advantage gained by utilising a wide area controller into the power system is that it establishes a common system objective and a correlation among the independent local controllers. Therefore, the necessity for the successful development of highly effective wide area controllers is becoming more apparent recently.

4.5.1. Wide area control structure

The rapid deployment of wide-area measurements in the power system led to the introduction and implementation of several design methods for the development of wide area controllers. Various methodologies and WAC structures have been proposed until now to ensure the effective damping of the inter-area oscillations and to enhance the small-signal stability. Control structures can be classified into quasi-decentralised, centralised, and hierarchical control architectures [71]. These are proposed with the aim to update the current decentralised control architecture where local controllers (which utilise only local information) are considered. Quasi-decentralised structure refers to a case where the local controllers receive some global information only regarding the state of the power system, while the majority of the data are collected and processed locally [72]. The disadvantage of this control structure is that it cannot ensure the complete compensation of all inter-area oscillations due to its limited global information. The centralised control scheme consists of only one central controller (e.g., located in the control centre), which utilises global measurements to derive direct control inputs for all the active parts of the power system, such as generators and RES. In this way, this control structure bypasses and cancels out the operation of all local controllers (such as the PSSs) of the power system [73]. The major drawback here is that this control scheme is highly dependent on the communication infrastructure and any failure or delay can have catastrophic consequences on the operation of the power system. Hierarchical control refers to a multilevel control structure, where controllers of a higher level coordinate the controllers in the lower level [74], [75]. This is the most desirable architecture since it utilises the global information to damp effectively all inter-area oscillations and at the event of losing a higher level, the system can still operate by utilising the controllers of the lower level [76].

4.5.2. Coordination of system components

Various methodologies for the development of a wide area controller have been investigated in the literature. The WAC schemes could be classified based on

the components of the power system that the wide area controller is intended to coordinate. Specifically, the components which are commonly considered in the available literature for coordination are the generator local controllers (exciter, governor, and PSS), the High-Voltage Direct Current (HVDC) lines, the renewables and the FACTS devices. The proposed schemes are either proposing the WAC of a specific component type or a combination of the above components.

Exciter: Considering the coordination of synchronous generators, the existing methodologies propose the WAC of its local controllers, i.e., exciter, governor, and PSS. The published schemes consider the coordination of either one or more types of local controllers. In the case of the exciter, a supplementary WAC signal is derived and applied to its input (in addition to the one provided by the PSS, when they are also included in the power system), having as a goal to increase its small-signal damping capability [71].

Governor: Many WAC methodologies surveyed propose the coordination of the exciter only (directly or indirectly, through the PSS). One of the main reasons is that the excitation systems have a faster response compared to the governor. Further, the availability of the PSSs feedback signal on the input of the exciter adds on the motivation for this choice. However, when someone wants a fast simultaneous damping of local and of inter-area oscillations, then the coordination of the governor with the control of the generators' frequency might be beneficial [77].

Power System Stabiliser: The WAC of PSS is a very popular approach to increase the power system's small signal stability. This is mainly due to its capability of generating an electric torque component able to compensate the local component of the low frequency oscillations [78]. Nevertheless, some studies suggest the coordination of the PSS by providing a supplementary signal on its output signal. Others are trying to upgrade the conventional PSS in such a way that its parameters will change according to the stochastically changing operational point of the system [79].

High-Voltage DC lines: HVDC systems are becoming interesting and important components of future power systems. This is mainly due to their low electrical losses over long distances compared to the respective AC systems [80]. HVDC lines are utilised either to deliver power from renewable sources to the main power system or to transfer large amounts of power between AC systems. Due to their power electronic-based technology, they are flexible to be controlled according to the needs of the system. For this reason, the HVDC lines are constantly gaining attention in participating into the WAC schemes along with PSSs and FACTSs. To damp the targeted inter-area oscillation, the HVDC line modulates accordingly its active power. However, the weakness of coordinating an HVDC line to damp a specific mode is the fact that it ignores the impact that this action has on the other inter-area modes that may exist [80].

Flexible AC Transmission System: The participation of FACTSs into the WAC

scheme is also a very effective option for the successful compensation of the inter-area oscillations. Like the HVDC lines, these are power electronic converter-based devices, meaning that they can be flexible for coordination by a wide area controller. Therefore, various works have been published proposing wide area control designs for regulating the FACTS operation.

Renewable Energy Sources: During the last decades, the system stability has been threatened mainly due to the large penetration of intermittent sources (renewables) in the system, which displace conventional controllable energy sources such as the synchronous generators. Based on this transition, the system's inertia and dynamics are changing [81]. While these changes in the dynamics of the system are widely recognized, the methodologies proposed so far for WAC design do not consider the influence of RESs on the performance of this controller. These approaches were motivated by the relatively low level of RES penetration. For this scenario a common practice among system operators was to consider RES as negative loads. However, as the RES penetration percentage becomes significant, and thus the system dynamics changes faster due to their presence, the negative load model for RES becomes obsolete. This motivates to design such WAC that takes into consideration the effect of RESs, a position in line with the smart grid of the future, where RESs are expected to be a large portion of the power generation. It is worth mentioning that use of the negative load model for RES also disregards their possible contribution to the overall stability of the power system [82]. To summarize, there is a need for a WAC scheme which leverages the capability of renewables to contribute to the system stability along with the synchronous generators. This scheme shall capture a common coordination between these two types of generators to effectively contribute to the damping of all the power oscillations. Amongst other benefits, the successful compensation of the inter-area modes can even enable the very high penetration of RESs into the grid [83].

4.5.3. Data delays and data dropout

Communication networks are utilised to transfer the synchronised measurements from the PMUs to the WAMC applications and the coordination signals from the WAC to all the local controllers of the system. However, communication networks are far from ideal, and this is one of the main obstacles in the implementation of real-time applications. The reason is that communication networks are characterised by communication delays and data dropout. These are considered to be the most known factors responsible for the degradation of the WAC damping capability, which in the end could make the system unstable [84]. Communication delays may appear during data exchange between the devices involved in the communication link environment (e.g., PMUs, PDC, etc.). They are caused by the long-distance transportation of feedback signals, the type of environment where the communication takes place, the data buffering, and the time required to send each bit of information. Even though a common literature approach is to consider the communication delays as constant or time-varying, they are actually random in nature because of the multitude of

possible communication channels. Besides communication delays, the operational/processing delays occurring from other components of the real-time applications (e.g., PMUs, PDCs), is also a significant issue. More specifically, [85] and [86] provide a detailed explanation of the delay types, which are segregated as follows:

a. PMU Delay: It is defined as the difference between the time the timestamp of a packet is set and the time the packet is at the output of the PMU. This delay type is directly dependent on the PMU operation (e.g., reporting rate, filtering).

b. Communication Delay: It is the time required for the measurement to travel from the PMU to the PDC. This delay has the most variability mainly due to the conditions of the communication network (e.g., traffic, medium length, routers, switches, etc.).

c. Control Processing Delay: The time period needed by the application to process the measurements and derive the control signals.

d. Command Delay: The time required for the coordinated component to react to the feedback control signal.

Data dropout is an unpredictable phenomenon and it can occur due to the appearance of noise/uncertainties in the communication network, transmission errors and data congestion, which lead to buffer overflow. To overcome this issue, many network protocols make use of transmission-retry mechanisms, which ensures that the data are re-transmitted after a certain time period. When this period passes and no re-transmission takes place, then the data is dropped. Data dropout is also a contributor to the degrading performance of the system, along with the delays. Apart from the data dropouts due to network issues, data drop phenomena can occur due to cyber-attacks, such as denial of service (DOS), which can terminate the flow of measurements for several instants [87].

Various studies have been proposed to address the impact of network delays and data dropouts on the WAC performance. A common solution to address these problems is the use of predictors, which could enhance the design of the wide-area controllers.

4.6. Modelling and simulation

Many of today's power systems are operated close to design limits and under heavily stressed conditions due to a mix of significant changes from planning and daily operation due to electricity market implementation, generation technologies, and an increase in demand all based on a legacy design for which these changes were not anticipated. This operation environment is challenging the reliability and security

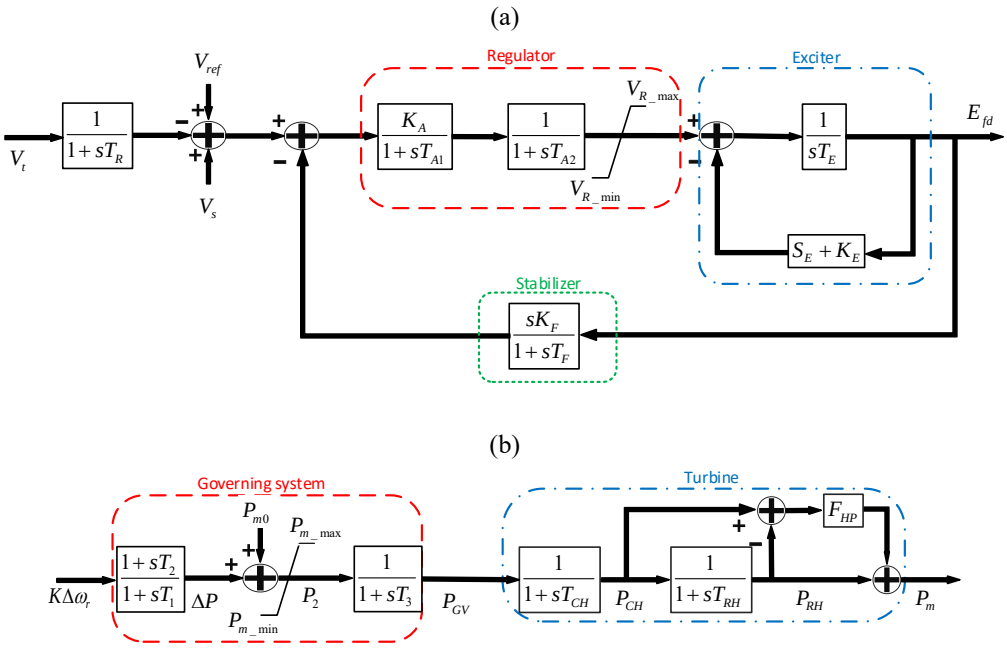


Fig. 4.4. Block diagrams of (a) DC2A exciter and (b) general purpose governor

of the power systems. Severe faults and disturbances are also encountered more often due to aging and natural related causes. Their timely detection is necessary for appropriate actions to be taken.

The power system is formed by a few sectors, of which each power system operator assesses the stability of its own sector to prevent disruption or unavailability of operation. Operators in the control centres of the power systems make use of transient analysis tools to further enhance their situational awareness of the system. These tools could help operators to adequately plan for any remedial actions that might be necessary for maintaining the stability of the system in case of disturbances. To do that, an essential part for the successful implementation and validation of the WAMC methodologies is the development of analytical and realistic simulation. The accurate representation of the system's behaviour is required specifically when dynamic conditions take place where the WAC contribution is essential. Another critical use of these simulation models is the identification and addressing of all potential threats which can affect the WAMC performance in actual conditions, obtaining in that way robustness along with high performance.

4.6.1. IEEE dynamic test systems

It is imperative to acquire a dynamic simulation environment, which is able to perform accurate Electro-Magnetic Transient (EMT) simulations intended for

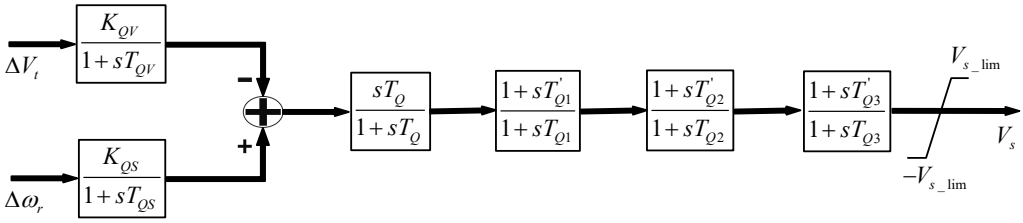


Fig. 4.5. Power system stabiliser block diagram

transient analysis. This can be achieved by utilizing dynamic test systems. However, the available IEEE test bed systems are only used for steady-state studies, since they consider only ideal sources and not dynamic parts. In [88] all the necessary modifications for reforming the most known IEEE test systems into dynamic test systems are presented. More specifically, a slight modification of the topology was applied to utilize sixth-order generator models along with their local controllers (all provided in [89]) instead of the ideal sources. Therefore, each generator is represented that way by its sixth-order model, while it is equipped with an exciter DC2A and a general-purpose governor as local controllers (Fig. 4.4). Furthermore, for obtaining a system with enhanced stability but for developing also novel methods which surpass the current situation, PSSs can be included as well (Fig. 4.5). Various simulation environments can be utilized for EMT simulations. These include PowerWorld, PowerFactory, and Matlab/Simulink. For the purpose of developing WAMC applications, Matlab/Simulink is shown in the literature to be the most desirable software.

4.6.2. Real-time simulation

Consideration of an offline software (e.g., Matlab/Simulink, PowerFactory, etc.), even with the development of dynamic test systems, is still not enough to ensure that the proposed WAC methodologies will be successfully implemented and that they will be able to operate effectively in actual conditions. This is mainly due to the fact that the WAMC is an application which needs to be executed in near real-time. The latter is actually a challenging task. To overcome this issue, advanced design and testing methods are necessary [90].

More specifically, real-time simulators are required to validate the performance of the wide area controller as well as other real-time methodologies, under highly accurate and real-time environments. The increased accuracy and real-time capability of these simulators combined with the realistic simulation environments results into an ideal testbed suitable for evaluating new methodologies under close to reality conditions. The industry considers the utilization of real-time simulators as a widely known and recognized validation procedure of engineering concepts, which is a necessary step before moving to real field tests since it reduces delays, risks,

and costs. This is mainly due to the real-time validation procedure, which provides realistic insights into practical design and implementation challenges [91]. Overall, the real-time simulators provide three simulation options for validating the proposed methodologies:

a. SIL: The first one is the software-in-the-loop (SIL) simulation, where the proposed method is integrated and tested into a real-time model of the system [92]. Figure 4.6 (a) presents an example of the SIL connection.

b. HIL: The second option is the hardware-in-the-loop (HIL) simulation, which provides the capability of connecting physical devices to the real-time simulation model. This kind of simulation is used for the development and testing of highly complex control, protection, and monitoring systems [93]. More specifically, HIL simulation is an alternative to traditional testing, where the real system is replaced by its equivalent real-time model and is used to interface with other physical equipment (Fig. 4.6 (b)).

c. PHIL: Finally, the advancement of the HIL option is the power hardware-

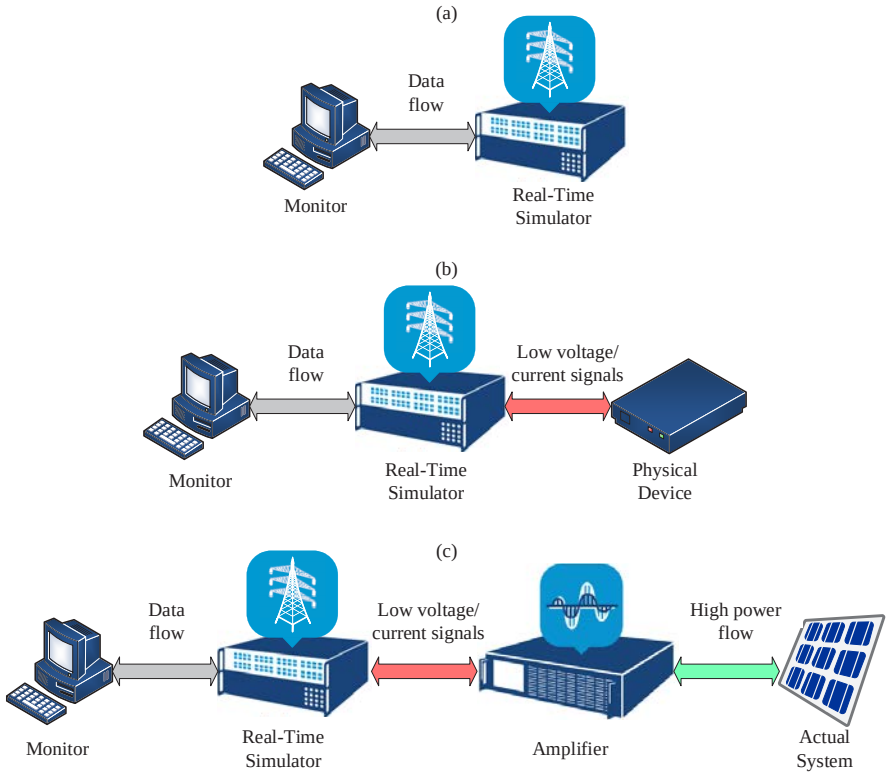


Fig. 4.6. Illustration of the three real-time simulation options: (a) – SIL; (b)– HIL; and (c) – PHIL

in-the-loop (PHIL). HIL is considered when low-voltage and low-current physical devices are connected to the real-time model, while PHIL provides the actual power flow required by higher power devices. As shown in Fig. 4.6 (c), this is achieved through the use of a power amplifier between the real-time simulator and the higher power equipment, which means that actual power system components (e.g., PVs, relays) can also be included into the loop [94].

4.6.3. Integration of wide area monitoring and wide area control in real-time conditions

d. The WAM-WAC integration is essential and beneficial in many ways. More specifically, the WAM-WAC combination will allow the optimal PMU placement according to the needs of the system and it will update in real-time the system topology that the WAC utilises (i.e., the bus admittance matrix). Therefore, this section will illustrate the evaluation of the WAMC performance in almost actual conditions by utilizing the real-time simulator and real equipment. For this reason, a conventional WAM method (linear state estimator) is considered along with an advanced wide area controller. Considering the real-time simulation and the development of the advanced wide area controller, further details are provided in [95].

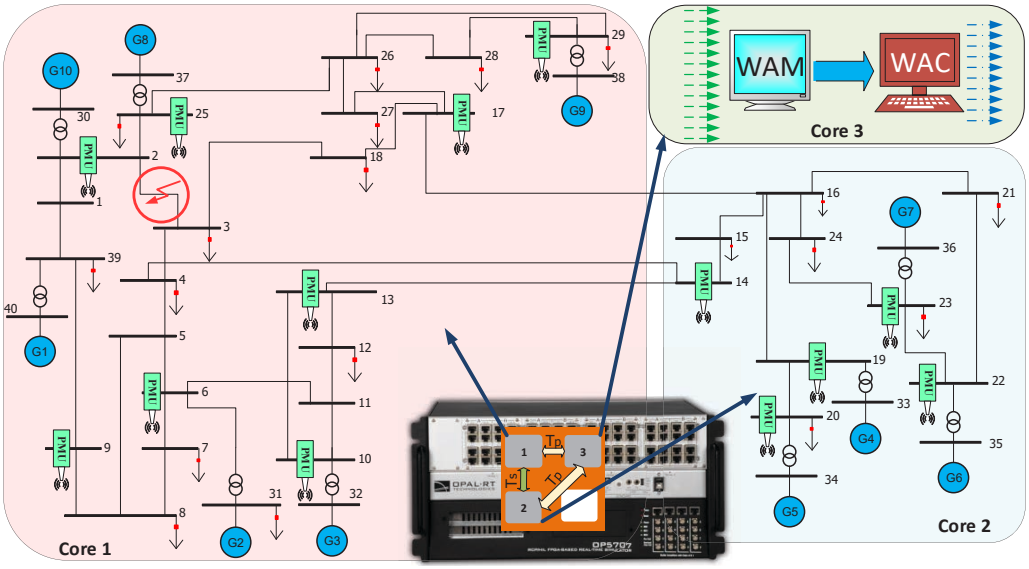


Fig. 4.7. IEEE 39-bus test system illustrating the separation of the system into 3 cores for real-time simulation and the optimal PMU placement in the system

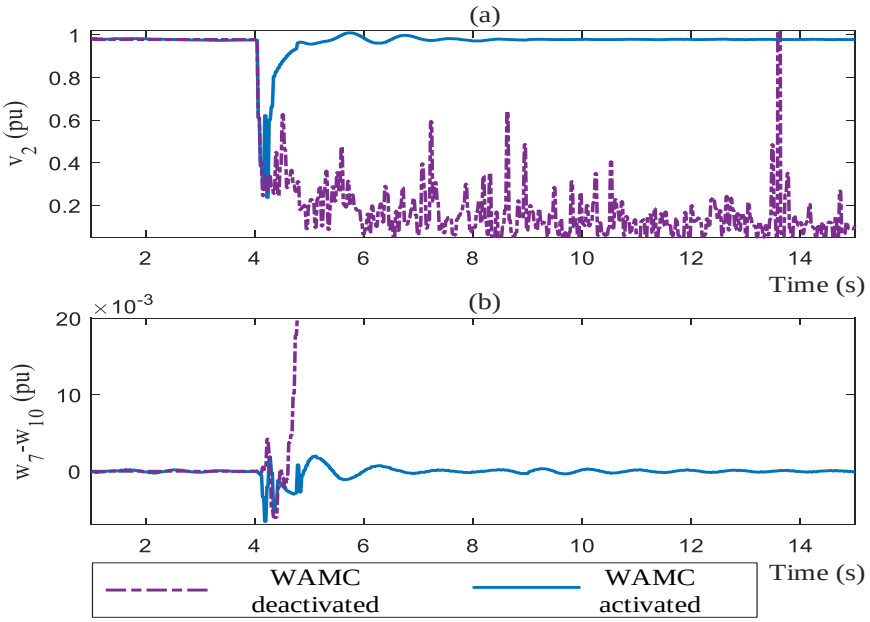


Fig. 4.8. Real-time simulation results show the system response when the WAMC scheme is activated and deactivated considering the compensation of (a) local and (b) inter-area modes

4.6.3.1. WAM-WAC integration in the real-time IEEE 39-bus dynamic test system

Here, the WAM-WAC integration is evaluated in real-time conditions using the real-time IEEE 39-bus dynamic test system. As aforementioned, the combination of WAM with WAC requires the optimal placement of PMUs in the system. Therefore, in case of the IEEE 39-bus dynamic test system, 13 PMUs in total are needed to be installed on specific buses in order to obtain full observability. The optimal placement of the PMUs for the case of the IEEE 39-bus dynamic test system is shown in Fig. 4.7. Based on the linear state estimator, the WAM will provide to the wide area controller at each time step all the required voltage and current phasors for all the generator buses. The WAM-WAC integration was tested in OPAL-RT real-time simulator where three cores of the OP5700 are used. Two cores are used for running in real-time the dynamic test system with a time step of T_s = 0.1 ms. Each core represents an area of the IEEE 39-bus dynamic test system. To acquire a realistic operation for the WAMC application, a separate third core is used, which holds the WAMC system and operates in a much higher time step (T_p = 40 ms).

The performance of the WAMC scheme is examined under the occurrence of a 5-cycle three-phase fault on the line, which connects bus 2 to bus 3, and is followed by a line tripping. The performance of the novel WAMC scheme is then evaluated by comparing the system's response when the WAMC is activated and when it is

deactivated. The graphical results of the real-time simulation (Fig. 4.8) illustrate clearly the performance enhancement offered by the WAMC system in damping effectively all the local and inter-area oscillations. In particular, the system remains stable only in case where the proposed WAMC scheme is activated.

4.6.3.2. WAMC implementation and validation in a real-time hardware-in-the-loop configuration

This subsection aims to utilize real equipment for developing the WAMC system in a real-time HIL configuration. This is a necessary and important procedure since it is one-step before the realization of the WAMC system and its application to a real field test. Figure 4.9 illustrates the laboratory setup considered for implementing the WAMC system in a real-time HIL configuration. The required components, which are described here, are the real-time simulator, the PMUs, the Global Positioning System (GPS) antennas, and the PDC.

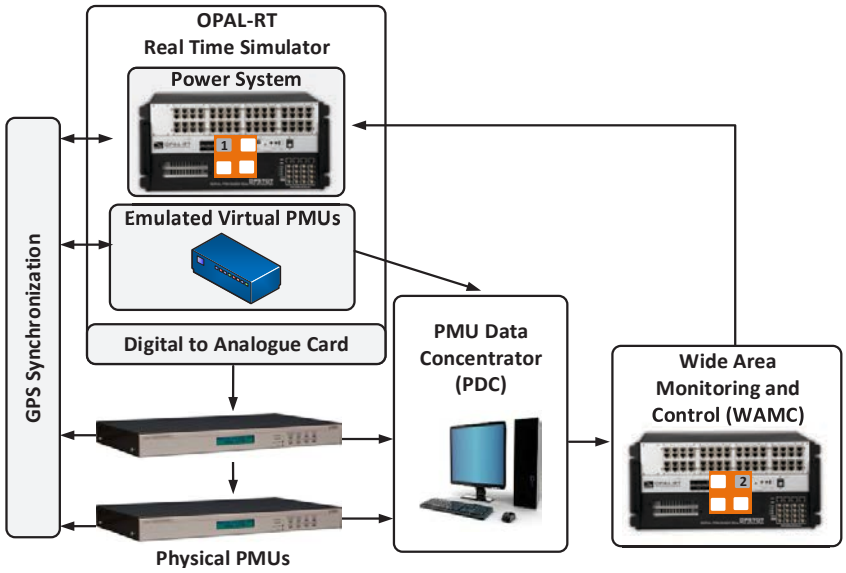


Fig. 4.9. Schematic diagram of the laboratory setup for testing the proposed WAMC scheme in actual conditions, which consists of the real-time simulator, actual and virtual PMUs, GPS antennas, and a PDC. Core 1 of the real-time simulator holds the IEEE 9-bus

a. Real-time simulator: The OPAL-RT OP5700 real-time simulator is considered here along with the eMEGASIM software package. The OP5700 holds the test system in one of its cores. More specifically, the IEEE 9-bus dynamic test system is used and modified accordingly in order to generate and provide, through its analogue outputs all the necessary three-phase voltages and

currents to the PMUs. The optimal PMU placement on the IEEE 9-bus dynamic test system, is shown in Fig. 4.9, where three PMUs are placed on buses 4, 7, and 9. Note that the WAMC scheme is developed in the second separate core of the real-time simulator, in order to utilize its high computational power for running this kind of algorithms in near real-time conditions.

b. Phasor measurement units: Two PMUs of Arbiter (model 1133A) are employed to take the analogue signals of the three-phase voltages and currents (provided by the real-time simulator) and derive their respective phasors. Due to the availability of only two PMUs and the limitation on the output analogue ports, the third PMU is implemented virtually into the real-time simulator. Note that all three PMUs satisfy both, measurement requirements and real-time data transfer requirements of the IEEE Std. C37.118.

c. GPS antennas: The provision of synchronized PMU measurements is a crucial procedure for the implementation of near real-time applications. For this reason, the GPS is utilized, which provides a precision signal for time synchronization with an accuracy of $\pm 0.2\mu s$. Furthermore, all the PMU measurements need to be synchronized to the Coordinated Universal Time (UTC). Therefore, GPS antennas are installed into the laboratory setup (Fig. 4.9) to operate as time sources for the synchronization of the actual and virtual PMUs.

d. Phasor Data Concentrator: A PDC is responsible for gathering and time aligning the synchrophasor data provided by more than one PMU. Therefore, the laboratory setup also considers a PC, where a python script is developed for executing the PDC functions (data gathering and time alignment) and for transferring the aligned measurements to the WAMC system.

To illustrate the enhanced damping capability achieved through the utilization of the proposed WAMC scheme, the system's response when the WAMC is activated is compared to the response when the WAMC is deactivated. Both scenarios consider the occurrence of a three-phase fault. Figure 4.10 presents the experimental results of the real-time HIL configuration. Based on these results, one can note the significantly better damping performance of the system when the WAMC scheme is activated. More specifically, as Fig. 4.10 denotes, by deactivating the WAMC scheme the system goes to instability during the disturbance. The outcomes derived here validate the performance of the proposed scheme in almost realistic and actual conditions, ensuring its effective operation.

4.6.4. Modelling and simulation tools for power systems

There are several software platforms that are commonly used for analysing and monitoring power systems. The software platforms offer a mathematical representation of the power system for analysing the performance of the power flow, for improving the transients between states, and for verifying the interaction between the substations with other elements of the system. Some of the mostly used tools are:

a. PSCAD/EMTDC[1] is used for studying the response of a complex system facing various sorts of disturbances or variations in the model parameters. This is the tool of choice to understand the behaviour of such complicated systems for which computer simulation is producing the system responses either based on models which could monitor instantaneous or RMS values in time domain, or could look into the decomposition of the frequency of the response signal. The EMTDC is mostly used for studying electromagnetic transients in electrical systems.

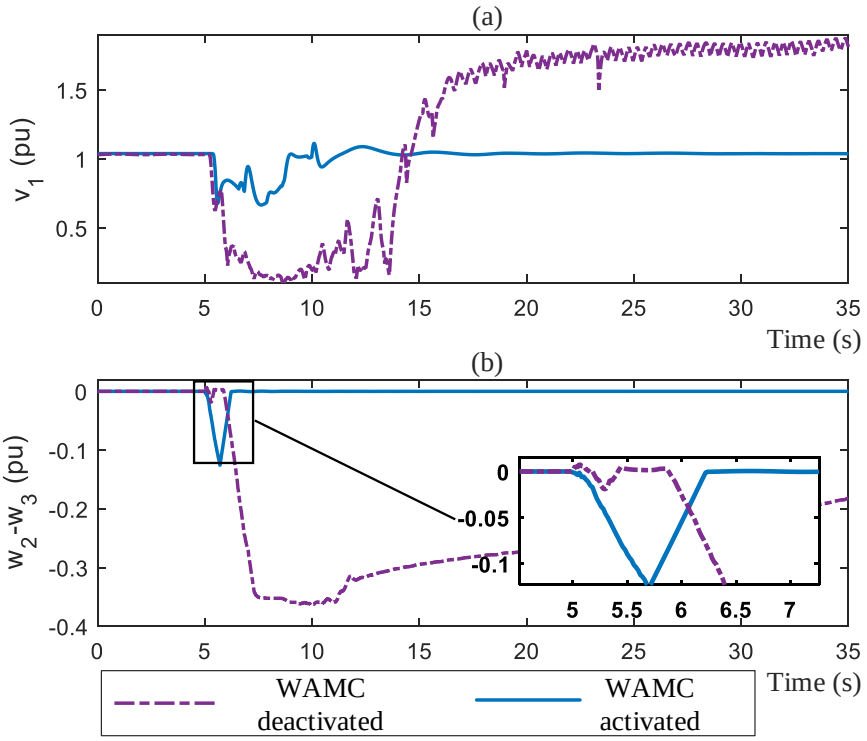


Fig. 4.10. Experimental real-time simulation results of HIL laboratory setup showing the system response when the WAMC scheme is activated and deactivated considering the compensation of (a) local and (b) inter-area modes

[1] https://www.pscad.com/

b. SIEMENS PSS®SINCAL[2] platform is used for 1) load flow (balanced and unbalanced systems) and standard-conforming short circuit analyses; 2) protection coordination with various range of capabilities; 3) short circuit including overcurrent time protection simulation and motor start-up for planning and optimising industry networks; 4) planning and optimising transmission networks; and 5) planning and optimising distribution networks.

c. SIEMENS PSS®E[3] platform is mainly used for optimising a workflow with Model Management module and providing an industry benchmark for simulation results.

d. Power System Simulation Tool (DIgSILENT)[4] is an industry state of the art power system analysis software. It can be used for the whole chain of electricity production vertical from generation, transmission, distribution to detail sections of industrial systems. It is highly versatile, covering a full range of functionalities, which lie from standard features to very advanced applications (e.g., renewable based distributed generation or real-time simulations). It can be also used to monitor the performance of the system in industry-based testing or supervision applications.

e. MATLAB® (and Simulink®)[5] is a very popular tool in both academic and industry engineering research. In the case of power system analysis it is used for a wide range of studies from design of power system components to the development of complex control algorithms. Without being exhaustive, within the scope of studying power systems as cyber-physical systems, MATLAB and Simulink could be used for 1) feasibility and grid integration studies for renewables for which pre-existing functions and applications are ready available; 2) parameter estimation to study automated control systems which need to meet specific design or regulatory requirements, especially in the case of renewables; 3) evaluating the grid code compliance and for simulating RES performance against production goals and grid compliance; 4) performing power quality analysis using EMT simulation and harmonics identification; 5) designing new power equipment (e.g., power electronics interfaces for distributed generation) and their associated control systems; or 6) code development which could be directly used for real-time and embedded systems.

f. GridLAB-D™ [6] is another popular research and development engineering simulation environment due to its high flexibility and interoperability, with a variety of third-party data management and analysis tools. The core of GridLAB-D™ uses an advanced algorithm, which is able to simultaneously operate with millions of independent devices, making it a state-of-the-art tool of choice especially for IoT applications for power systems. GridLAB implements each device to be coordinated as a set of multiple differential

equations. Its easy integration and vendor-agnostic properties, on top of the models accuracy and wide time-scales simulation options (e.g., from sub-seconds to several years) are among the major advantages of this simulation environment compared to the former enumerated tools. Furthermore, GridLAB does not require the use of reduced-order models in the case of, for example, studying the behaviour of many consumers on the operation of electrical systems. Thus, by avoiding unnecessary aggregation models, accuracy of traditional models might be improved (e.g., by avoiding erroneous or misapplied assumptions). This tool is recommended for industry and government planners in charge of the design of programs to improve the load growth management or for enhancing system reliability.

Chapter 5:

Transportation Systems:

Simulation, Modelling, Traffic Video Analysis

Volodymyr Sistuk
Kryvyi Rih National University

5.1. Introduction

In scientific literature, automation of traffic management processes takes the form of "cyber-physical system" (CPS) [96], [97]. A cyber-physical system involves a human-machine interface (HMI) that obtains feedback via sensors and impacts the physical process via actuators keeping and evaluating the collected data. Compared to the conventional transportation system, a cyber-physical system can accomplish more efficiency and reliability by strengthening feedback in a virtual and real-world interaction. For the cyber-physical system whose work is addressed towards road traffic management in real-time, physical components are traffic lights, road cameras, electronic control units, traffic management centre, and program components, which are specialized software. Information is a function of monitoring, projection, and prevention [96].

Road transport cyber-physical system (RTCPS) simulation includes the mechanism to support modern communication, intelligent and cybernetic technologies. The simulation results provide a set of parameters for sustainable control of road traffic system. The data on road user's behaviour could be also applied to the traffic control device adjustment or traffic management mode implementation [96]. To transfer the traffic management system into cyberspace, spatial planning is usually utilized. Geoinformation systems (GIS) constitute one of the main efforts of road transport cyber-physical systems simulation.

This chapter is devoted to the type of RTCPS, which is infrastructure-based. It is divisible into the physical elements such as camera, an embedded device or a server-based solution, and cyber elements, such as wired and wireless communication, and specialized software. This solution has various uses of traffic real-time control. This chapter also provides an overview of the technique of traffic survey by using platforms for automated traffic video analysis.

5.2. Infrastructure-based road transportation cyber-physical systems (In-RTCPS)

5.2.1. The architecture of In-RTCPS

An example of infrastructure-based road transportation cyber-physical system (In-RTCPS) is a visual programming language (FLOW), which was designed by RCE systems. The solution for road traffic analysis in real-time that contains an embedded system, a visual programming language (called FLOW), a neural network, and an Application Programming Interface (API) [98]. The main functions of this cyber-physical system are as follows:

- vehicle and pedestrian movement detection and tracking in real-time;

- extraction of the trajectories of the tracked objects;
- traffic analysis depending on the mission;
- visualization of traffic statistics in an interactive form.

The initial component of the FLOW solution is a trajectory of the tracked object. It fully describes its movement and visual characteristics. The main functions of RTCPS, such as investigation of the specific traffic situation, monitoring of traffic, detection of the road users, estimation of the forthcoming tendencies, and prevention of possible incidents are realized by a presentation of traffic flow as a set of finite trajectories. Therefore, the FLOW nodes create trajectories through different video feeds (road cameras) and GPS tracking system. The tasks of the FLOW cube are to receive trajectories collected by one or several nodes and to sort them by the same tracked objects. Consequently, continually obtaining the vehicle's trajectory by various video feeds can be explored in a holistic way that is a principle enshrined in the cyber-physical systems [98].

The FLOW blocks obtain the data from multiple cubes and evaluate merged trajectories. Finally, FLOW Insights accommodates all previous layers of the software. This client application is an upper layer of the FLOW solution adapted to analyse traffic in various ways. It helps to provide traffic control, adjusts adaptive traffic lights, finds conflicts between road users and realizes many more activities of traffic management. The statistics data can either be available via the API or accessible through the widget in the custom dashboard (see Fig. 5.1).

Fig. 5.1. Visualization of the traffic statistics in the FLOW Insights (*the picture from* [98])

FLOW as an open platform allows working together with third-party devices through HTTP and UDP protocols. In this case, the communication tool for input and output data is data sinks. FLOW supports the REST sinks and UDP sink with the JSON syntax for the messages.

With REST sinks, the third-party application uses HTTP protocol that sends queries to the server and gets responses [98].

5.2.2. Physical component of In-RTCPS

The physical component of In-RTCPS can be a stand-alone appliance as well as multiple devices linked in a network depending on certain tasks. To minimize costs, in some cases the simplest solution is a smart camera that involves microprocessor with AI with pre-installed software. The main functions of the microprocessor are traffic data, gaining, assortment, compression, analysing, and data storage. As for the software, it is responsible for a user interface (dashboard) wherein the tasks for In-RTCPS are created via the system of filters and adjustments. The software is also a tool either for traffic data export to third-party repository or for an external device, through the API. After providing all operations with filters, monitoring, and data export, the processed information is realised by using network connection [98], [99].

With the aim of analysis, multiple video streams in real-time another device are used instead. It has an integrated Poe switch or the technology of ETH, WiFi, and LTE that connects cameras to the server. The capabilities of the unit can be extended by adding three cameras to the system. It facilitates performance analysis of video streams from six cameras simultaneously. Similar to the previous case, traffic analysis is accessible due to the same kind of pre-installed software.

A large-scale In-RTCPS is an end-to-end solution for data centre or a centre of traffic control, which brings together smart cameras in a comparative grid analysing tens of video streams in real-time. The cameras connect to the centre via the global communication network of WAN that can transfer data over long distances.

Considering FLOW, the above-mentioned variants of physical components of the In-RTCPS and communications are realised in the form of TrafficCamera, TrafficEmbedded, and TrafficEnterprise solutions [98], [99]. The supplementary modules can be added to TrafficEmbedded and TrafficEnterprise. For example, with make and model recognition (MMR) module FLOW Insight can recognize up to 140 makes of vehicles and 1000 models of vehicles [98]. Therefore, FLOW is an integral solution for traffic surveillance in real-time and component of the In-RTCPS.

5.3. Intelligent video analysis of road traffic system

5.3.1. DataFromSky platform

The technology of video analysis of road traffic data consists of the following computing sequence. Artificial intelligence (AI) automatically detects vehicles and pedestrians from video footage in cloud service with Tracking Log creation. Tracking Log, initial data package of vehicle trajectories, is in Tracking Log files (suffix ". tlg") [99]. At this stage, Tracking Log is invalid for full traffic analysis because the video frame, scene does not contain any georeferenced points for the evolution of quantities or virtual objects for registering of vehicles displacement in studying areas. Video frame Geo-registration is used in DataFromSky Viewer software to annotate the vehicle trajectories. It is a type of software for road traffic data processing through formation, scrutinizing, editing, and analysing vehicle tracking data in video footages. The Viewer joins video data, geo-reference, scene annotation, and vehicles trajectories [99].

The tracked vehicle has the form of an object that has either a defined location in time or a specific trajectory in time-space. The identification number (ID), the values of speed, acceleration and deceleration, vehicle travel time, and other parameters are accessible for each tracked object. Also, speed and acceleration graphs are accessible for each tracked object. The trajectories of vehicles, the current scene annotation configuration, the OD-matrices (Origin-Destination) analysis results, the analysis of road safety, and time intervals between two following vehicles (headway time) are the subject for visualization. The OD-matrix shows the main parameters of a specific route from entrancee to exit the gate. There are vehicle counts, average travel time, minimal and maximal travel time, and standard deviation of travel time [100]. The application serves the functions for Tracking Log. There is file editing by geo-registration of the video sequence and hand annotation of vehicle trajectories; video sequence review; annotated and detected vehicle trajectories; and results of the automated traffic analysis in the form of various reports (CSV data). Several virtual objects annotate the scene. There are lanes, gates, analysis nodes, traffic regions, active regions, and anonymization regions.

The lanes predict vehicle arrival/departure in a specific area. They can be an entry, exit or neutral. The gates are virtual lines for determination of working space vehicle's crossing. It can be directional and selective, and entry, exit, and neutral.

The analysis nodes connect notations.

The core function of the traffic regions is to find a standing vehicle or standing time.

The active regions determine vehicle presence. In such a situation, the speed of the vehicle is available.

The anonymization regions are applicable for blanking of the specific scene area.

In Viewer, the adjustment of video frame notation (Manage Annotation Configuration) is responsible for lanes, gates, traffic regions, and action regions.

The main results of the intelligence traffic analysis are as follows [100]:

- determination of vehicle densities at the gates;
- classification of tracked objects, including bicyclists and pedestrians, calculation of origin-destination matrixes in defining directions with colour marking;
- measurement of speed and acceleration at the specific point of workspace showing coloured notation and creating heat maps of the scene;
- safety analysis using surrogate measures such as time-to-collision (TTC), post encroachment time (PET), and heavy braking.

5.3.2. GoodVision platform

The GoodVision platform is based on computer vision and artificial intelligence technologies to determine vehicle trajectories from traffic video sequences. The core analytical packages are the calculation of the quantity of the tracked objects, classification of the vehicles (car, truck, van, bicycle, OGV2, motorcycle, pedestrian), parameters of the car-following model (travel time, gap time) and traffic special aspects (jaywalkers, lane change behaviour, traffic lights), microsimulation model calibration parameters (saturation flows, automated detection of free-flow speed, regions of speed deceleration) [101].

The components of a workspace are lanes, movements, events, and scenarios. In GoodVision, lanes define the order of tracked objects calculation. A traffic movement is a flow of different vehicles (pedestrians) between two or more areas in the specific order. An event is a tool for detecting the vehicle that takes up a specific area of workspace within some time. Scenarios follow complex variants of traffic management.

As for the DataFromSky application, traffic analysis results are in the form of files in CSV format. In GoodVision the following traffic reports are accessible: traffic movement count (TMC), intrusion report, saturation flows, and OD-matrixes. In TMC the quantitative measures are determined for each tracked object throughout all periods of record or time lag per 1 second. The intrusion report shows time gaps

and travel times for every type of tracked objects of the corresponding movements. Saturation flows are determined by the value of passenger car equivalent per hour (PCU/h) for each movement and for the location taken as a whole. OD-matrixes define traffic between choosing movements [101].

5.3.3. Comparison of functional capabilities of the platforms

Table 5.1 shows an evaluation of the characteristics of the two platforms. The information is taken from open sources [99], [101].

The platforms are not differentiated by how many basic operations they realize (see Table 5.1). The traffic reports of both programs are powerful tools of analysis. However, the output data presentations differ more substantially. In GoodVision application, there is an emphasis on detailed reports on time gaps and saturation flows. The benefit of DataFromSky is the possibility to analyse traffic safety via conflict list for each tracked object. There is a significant difference in the classification of tracked objects. The use of one or another application depends on the accuracy of estimates of traffic frequency.

Table 5.1. Comparison of the Features of Platforms of Traffic Video Analysis

Measure	DataFromSky	GoodVision
Number of tracked objects	16	8
Traffic reports		
OD-matrices	✓	✓
Headway time	✓	–
Time-to-gate, time-to-follow	✓	–
Gap time	–	✓
Travel time	✓	✓
Saturation flow	–	✓
Safety analysis	✓	✓
Visualization		
Video visualization of tracking	✓	–
Traffic dynamics (graphs)		
Trajectory	✓	✓
Speed	✓	✓
Acceleration	✓	✓
Time gap	–	✓
OD-charts for the most loaded routes	✓	–
Speed-to-acceleration dependences for road vehicleunit	✓	–
Traffic heat map	✓	✓
Acceleration heat map	–	✓
Rush hour traffic load	–	✓

5.4. RTCPS case studies

5.4.1. Methodology

For a better view of the platform's capabilities, this subchapter sets out two case studies. The first study is for checking two traffic reports obtained by using DataFromSky and GoodVision; the video footage was captured in the worst conditions (IP camera). The principle underlying the technique is an assessment of the traffic flow OD-matrices that are found via the offered software. It follows the correspondence between virtual gates that determines the accuracy of tracked objects detection and traffic dynamics when passing through the appropriate gate in the specific direction.

The comparison of platforms consists of the following stages:

- analysis of general statistical parameters obtained for various tracked objects;
- traffic flows OD-matrices determination considering each type of vehicle counts;
- paired comparison of OD-matrices data;
- visual observation over the flow frequency in the laboratory to give actual values of traffic flows;
- finding the relative error of vehicle count for the studied software.

The second technique determines the accuracy of the object detection for an aerial video of intersection evaluating an actual number of recognized vehicles for each class.

5.4.2. Traffic video from a city camera

The video footage is captured for signalized cross-intersection of Taras Shevchenko Boulevard and Volodymyrivska Street situated in the centre of Kyiv and characterized by high frequency of traffic flow (see Fig. 5.2). Taras Shevchenko Boulevard is a major road. When entering the intersection from Taras Shevchenko Boulevard, near National Museum, a dash line separates the opposite lanes. There is also a dash line on Volodymyrivska Street at the Ministry of Education and Science. The stop line is marked downstream of Volodymyrivska Street. At the entrance to the intersection of Volodymyrivska Street from the left outer line, traffic is allowed only to the left; from the centre line it is allowed both straight and to the left and from the right lane it is permitted to the right.

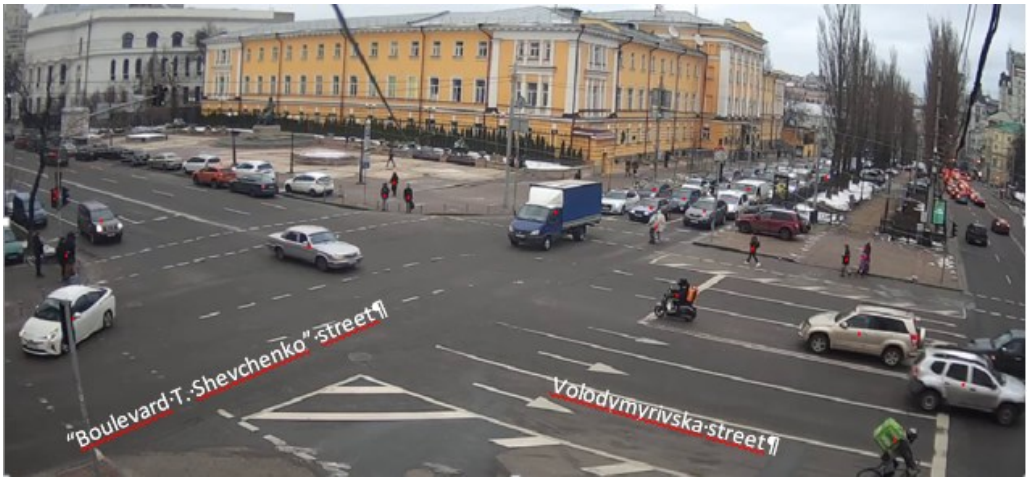


Fig. 5.2. The location view

Figure 5.3 shows a scene annotation configuration in DataFromSky and Good Vision software.

In DataFromSky Viewer the annotation configuration includes 9 lanes, 7 gates, 2 action regions, and 2 traffic regions. Active regions are created at the intersection entrance. The entire gates are marked by numbers 2, 3, 5 and the exit gates by numbers 1, 4, 6. In GoodVision the gates have the same location and numbering and correspond to traffic directions as in DataFromSky to unify the presentation of the results of analysis. The annotation configuration forms the basis for OD-matrix estimation considering the directions of traffic flows.

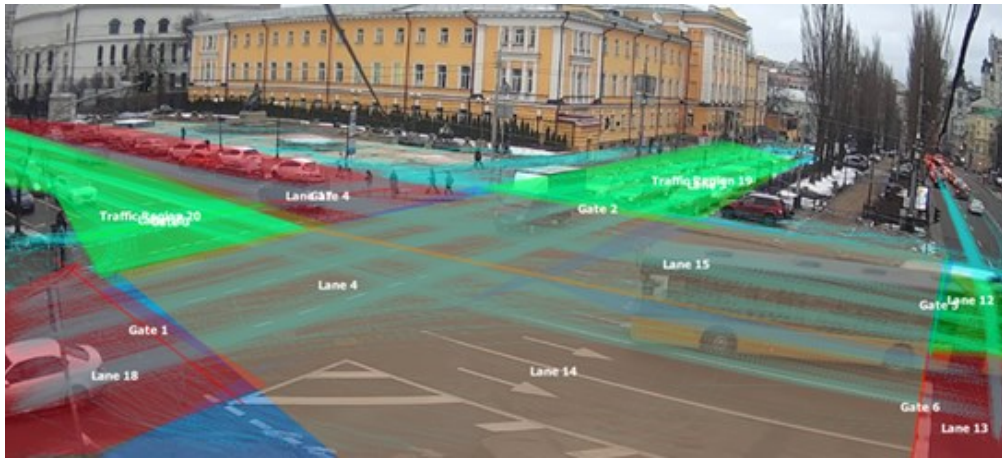
The platforms provide general statistics that include the flow's composition data. According to the results of the analysis in DataFromSky Viewer, the major contribution to the traffic frequency is the traffic of cars that account for 93.4 % of the total number of vehicles and 71.8 % of the total number of tracked objects. The number of cars in the system of GoodVision Insight has similar values. It constitutes 93.2 % of the overall number of vehicles and 78 % of the overall number of tracked objects. It is, therefore, possible to determine the capabilities of AI to identify the kind of objects by studying car recognition.

Summarizing the traffic objects count in the direction of gates in accordance with the software classifier, OD-matrices can be found for cars, buses, trucks, vans and all types of vehicles for the full video sequence. It was found that variances in program classifiers have no significant influence on the vehicle's recognition. We are especially interested in the OD-matrices of vehicles and cars count results (see Tables 5.2 and 5.3).

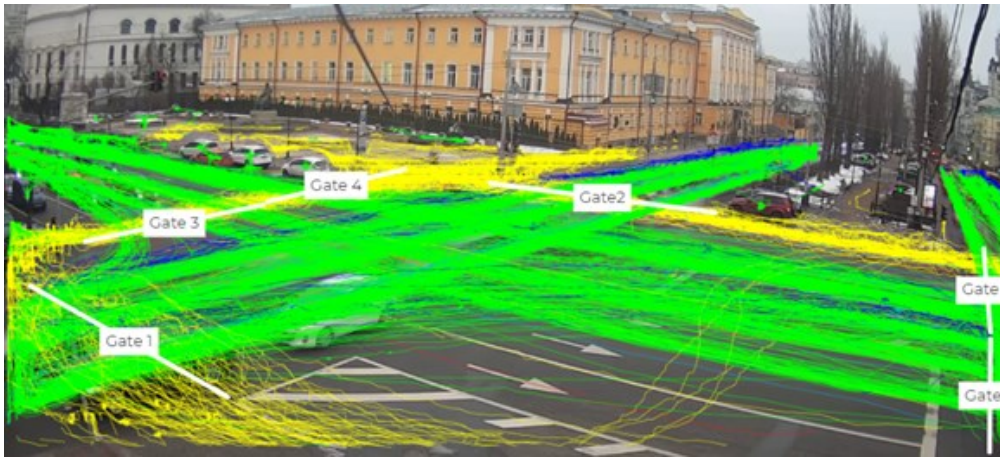
The platforms provide information about the most congested directions (sections)

of the intersection. In this case, there is straight traffic from Gate 2 to Gate 1 (on Taras Shevchenko Boulevard towards the Ministry of Science and Education). According to the results of the analysis, in 24 minutes the traffic volume is 350 and 382 vehicles in DataFromSky Viewer and GoodVision Insight, respectively. Besides, DataFromSky has an opportunity to visualize the most congested directions on the scenic view and in the form of a chart (see Fig. 5.4).

The number of vehicles in the direction from Gate 5 to Gate 1 (the left turn from Volodymyrivska Street to Taras Shevchenko Boulevard) is 97 according to DataFromSky analysis and 19 according to the report from GoodVision. The images of turning traffic are fused because of the side-view and low altitude of the video camera, which contributes to the loss of the accuracy of vehicle count. An indirect



a)



b)

Fig. 5.3. The annotation configuration: a) – in DataFromSky Viewer; b) – in GoodVision Insight

Table 5.2. OD-matrices of Vehicle Count Results for the Whole Intersection*

Gate Type	Exit Gate 1	Exit Gate 4	Exit Gate 6
Entry Gate 2	350/382/8 %	21/23/8 %	0/0/0 %
Entry Gate 3	54/61/11 %	0/0/0 %	123/54/127 %
Entry Gate 5	97/19/410 %	89/76/14 %	0/3/100 %

* The first value is from DataFromSky Viewer; the second value is from GoodVision Insight; the third number is the difference between the two values in absolute magnitude (%).

Table 5.3. OD-matrices of Car Count Results for the Whole Intersection*

Gate T ype	Exit Gate 1	Exit Gate 4	Exit Gate 6
Entry Gate 2	324/352/8 %	21/23/8 %	0/0/0 %
Entry Gate 3	53/58/8 %	0/0/0 %	119/53/55 %
Entry Gate 5	83/12/85 %	86/75/12 %	0/3/100 %

* The first value is from DataFromSky Viewer; the second value is from GoodVision Insight; and the third number is the difference between the two values in absolute magnitude (%).

indicator of vehicle count accuracy is the number of tracked objects passing through adjusting gates that is calculated as a percentage of the total number of all objects. The finding of the program's accuracy is in the value of measurement relative error (see Table 5.4). To give actual values of traffic flow, the visual observation over the traffic frequency was continued. The received data are added to OD-matrices for each vehicle type.

a) b)

Fig. 5.4. Visualization of the congested directions: a) – on the chart; b) – on the scene view

Table 5.4. Summary Report on the Accuracy of Road User's Detection

Vehicle type	The total number of objects from OD -matri ces		Actual value	Relative error , %	
	DataFromSky	GoodVision		DataFromS ky	GoodVision
Cars	686	576	778	13.4	26.0
Trucks	8	4	8	0.0	50.0
Buses	20	11	19	5.0	42.1
Vans	65	27	66	1.5	59.0

In this case, recognizing the function of cars strongly affects object detection. Therefore, the value of the relative error of car detection is 13.4 % for DataFromSky and 26.0 % for GoodVision, which indicates the advanced level of DataFromSky algorithm.

5.4.3. Traffic video from a drone

Aerial video of road traffic was captured at the X-shape intersection in Kryvyi Rih. This intersection of Lermontov Street and Starovokzaln Street has a high level of accident concentration and traffic congestion. The survey was provided at peak hours on Sunday, starting at 11 a.m. Video footage was obtained by drone DJI Mavic Air 2 from 50 m above the intersection. The frame rate is 59.9 fps, the frame number is 25336, and the video resolution is 1920 x 1080 pixels that fully meet the requirements of DataFromSky to a real video. The following information was obtained from the analysis results:

- identification and classification of vehicles;
- geometric parameters of the intersection (lane width and length);
- arrival and departure times;
- OD-matrices for traffic flows; headway statistics; acceleration/deceleration rate for each tracked object;
- road safety indicators.

Figure 5.5 shows the designed annotation of the scenic view. Annotation configuration lays the groundwork for vehicle dynamics analysis. By video post-

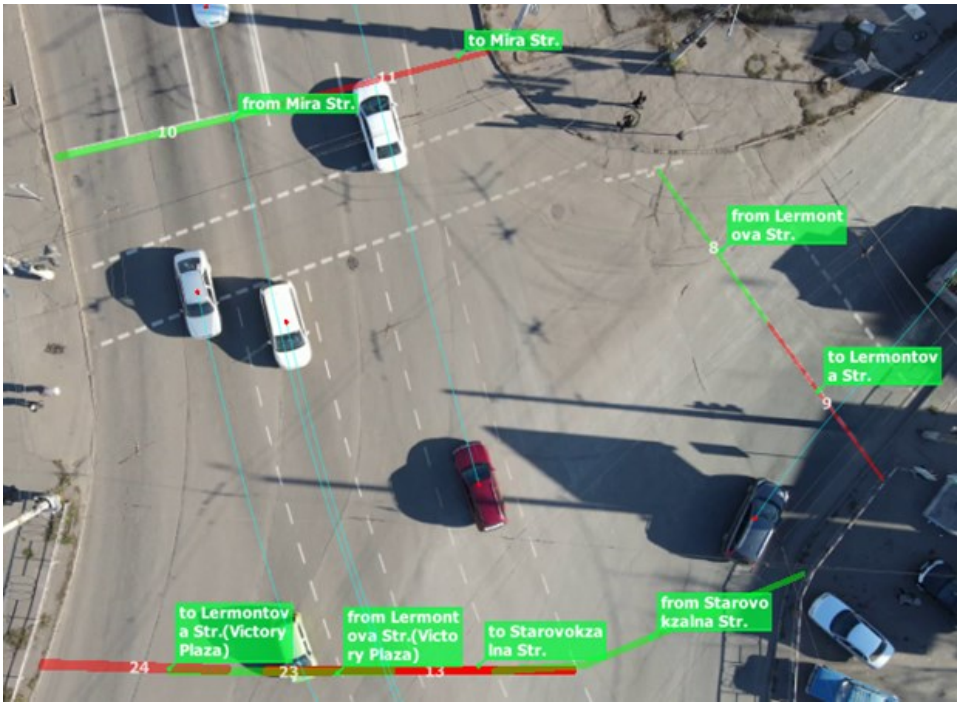


Fig. 5.5. Annotation of scenic view

process, the traffic consolidated figures, OD-matrices, reports on travel times, and headways for each flow direction were found. The average traffic frequency was calculated with emphasis placed on the congested movements via OD-matrices.

Table 5.5 is an OD-matrix with calculated and observed values of the number of vehicles for a specific direction.

The maximum flow frequency is from entry Gate 23 to exit Gate 11 (straight traffic from Lermontova Street to Mira Street). The traffic volume from entry Gate 10 to exit Gates 13 and 24 is also significant compared to other directions. The overlap of Gates 13, 14, 23, and 24 was utilized at scene annotation (see Fig. 5.5) to consider more vehicles that are precise count results.

Table 5.5. OD-matrix with Calculated and Observed Parameters

Gate Type	Exit Gate 9	Exit Gate 11	Exit Gate 13	Exit Gate 24
Entry Gate 8	0/0	0/0	19/22	16/18
Entry Gate 10	2/2	0/0	45/49	39/41
Entry Gate 14	25/22	38/42	0/0	0/0
Entry Gate 23	0/3	73/75	2/0	0/0

As expected, the maximum discrepancy between calculated and observed values of the number of vehicles at the appropriate route was obtained from turning flow. It is the traffic from Lermontova Street to Starovokzalna Street. The relative error of vehicle count for this route is estimated to be 15.7 %. In other cases, the relative error ranges between 5.0 % and 10.5 %. The average relative error is 9.2 %, the corresponding performance standard that determines the permissible variation of frequency estimation instrumental methods [102].

In average travel time, the route from Gate 10 to Gate 9 attracts attention that is perceived as the turning traffic flow from Mira Street to Lermontova Street. This

Table 5.6. Gate Statistics

Measure	Gate 8	Gate 10	Gate 14	Gate 23	Gate 9	Gate 11	Gate 13	Gate 24
Min speed [km/h]	18.55	4.88	5.20	2.27	13.67	8.73	2.08	6.91
Max speed [km/h]	57.53	24.02	49.64	24.41	75.15	27.10	24.87	16.45
Average speed [km/h]	38.60	14.15	15.91	13.24	29.78	18.05	12.91	11.69
Car count	15	80	77	114	41	108	99	49
Medium vehicle count	2	2	3	10	3	5	7	0
Heavy vehicle count	1	2	1	2	1	1	1	1
Bus count	1	5	6	5	7	3	5	5
Motorcycle count	0	0	0	0	0	0	0	0
Bicycle count	0	0	0	0	0	0	0	0
Pedestrian count	0	0	0	0	0	0	0	0
Number of all vehicles	19	89	87	131	52	117	112	55

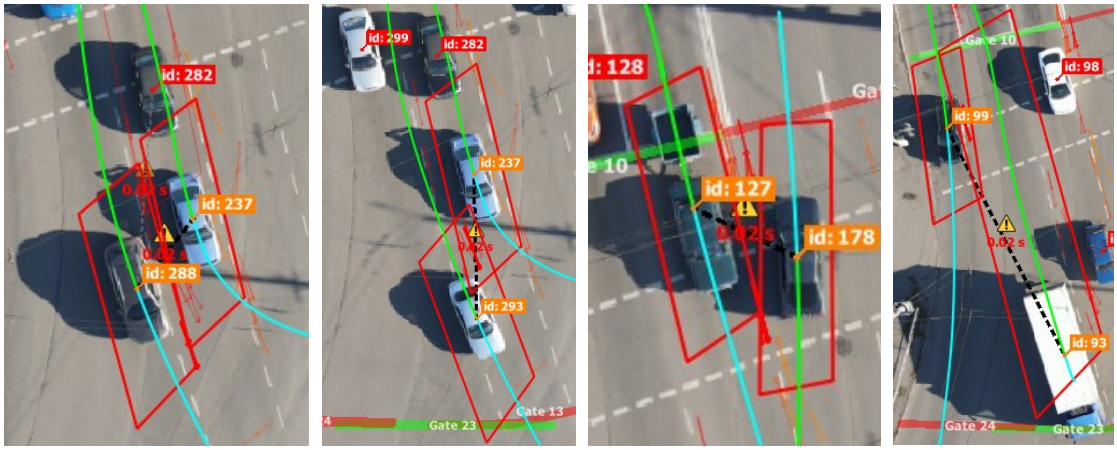


Fig. 5.6. Examples of the worst TTCs

information is also obtained from OD-matrices. However, in the last-mentioned direction, only two tracked objects were observed over a period of the video sequence. Table 5.6 shows aggregate statistics for the number of tracked objects that passed a specific gate. This information can be applied further for calibration of such measures as vehicle inputs, flow composition, and average speed for each route of individual and public transport in microsimulation software.

The measures of the vehicle's dynamic are as follows: average acceleration and deceleration, average traffic speed and headway time. For example, subsequent to the results of 329 trajectories processing it was found that the average longitude deceleration is –0.08 m/s^2 and the maximum longitude deceleration is 0.60 m/s^2. The average flow speed is 16.4 km/h with a standard deviation of 9.1 km/h. The speed distribution is close to normal law.

Headway statistics are also accessible for a Tracking Log created for the video sequence from the drone. Headway time is an aggregate parameter of traffic density. The maximum value of the parameter is 4.00 s. The average headway time at the studied intersection is 1.22 s with a standard deviation of 0.89 and the number of gates passed – 371. This suggests that the traffic flow is saturated.

DataFromSky traffic safety analysis is a powerful tool for conflicts study at various intersections, roundabouts, and highways. It was clear from the safety analysis that the most severe conflicts associated with time-to-collision (TTC) surrogate measure are typical in cases of a lane change in the same or opposite direction (see Fig. 5.6).

Post Encroachment Time (PET) is also determined in DataFromSky Viewer. It defines the time difference between instances of the conflicting vehicle leaving a zone and another vehicle arriving in similar area. For the studied intersection, the

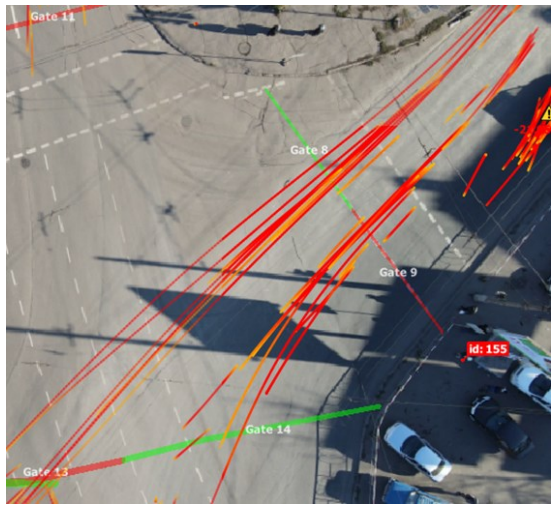


Fig. 5.7. The worst deceleration rates

software determined two PETs with the values of 0.83 and 0.93 s.

Heavy braking plays a significant role in road safety. With the threshold value of the deceleration rate of –2.50 ms^{-2}, the clarified values of heavy breaking are in the critical range of 13.75 to 21.95 ms^{-2}. Most cases occur for turning flows from Lermontova Street to Victory Plaza direction (see Fig. 5.7).

Consequently, additional modules of the software allow carrying out safety analysis with detection of possible conflicts of vehicles, which interact with each other.

The presented traffic data analysis capabilities within the DataFromSky platform give complete meaning to the traffic field data. Despite some omissions in recognizing the tracked objects, the computer vision has a significant perspective on transportation analytics. First of all, the method reduces a great deal of time to predict and analyse the traffic frequency.

5.4.4. The advantages and limitations of the technology

A smart traffic video analytics technique has both positive and negative impact. Let us summarize some of them.

The advantages are set out in the following arguments:

1. An automated traffic survey provides the data set on the travel density and flow composition for the whole location simultaneously; it significantly increases the quality of field data compared to the visual observations.

2. Intelligent video analysis makes possible estimation of traffic management by checking the statistics of prior periods.

3. The technique can facilitate the calibration of driver behaviour model in microsimulation software towards its adequacy to real conditions considering traffic dynamics for every tracked object and generate a statistics report on flow composition and distribution of the general properties of the traffic flow (density, speed, gap time) that are close to the actual values.

4. The calibration method of simulation model enhances the quality of predictive state assessment as well as alternative comparison through microsimulation and intelligent video analysis tools.

Limitations of the technique include:

- accuracy of vehicles and pedestrian trajectory detection depends largely on

video footage parameters (resolution, frame rate, angle and height of the survey) that promotes the use of expensive drones for video capturing;

- object detection and identification depends on the classifier's advanced level;
- use of all the possibilities of intelligent video traffic analysis to calibrate the microsimulation model is limited by the permanence of the parameters of the driver behaviour model in the microsimulation model.

In the highway construction national standard of Ukraine [102] the threshold observational error of hourly traffic instrumental measurement is established. In comparison to visual observation, the measurement error should not exceed 10 % for the roads of categories IA, IB, II, and III, and 20 % for roads of categories IV and V. The case studies with post-processing of traffic video from the municipal camera and drones have shown that computer vision can achieve results that are within the normative range. The angle of camera view has one of the greatest influences on the accuracy of tracked objects recognition. The complete implementation of the technology and functionality is conceivable by involving the software in In-RTCPS as it is realized in the FLOW solution.

5.5. Conclusions

With increasing complexity of transport networks and traffic control systems, a central assessment tool of road traffic is a computer vision technology as part of infrastructure-based road transportation cyber-physical system. An example of such a cyber-physical system is the FLOW solution used during the development of the smart city in the process of traffic surveillance, parking control, and road safety analysis. As technology developers have pointed out, it can be argued that the municipal camera becomes an intelligent sensor. The platform is a one-size-fits-all solution for various devices from the camera with the microprocessor to the server. According to the developers, it should be sufficient for the job to derive a camera and assign the appropriate task given the kind of available data. The software automatically recognizes and ranks objects for traffic monitoring and control and road safety evaluation.

In view of this, traffic frequency assessment is one of the major concerns of In-RTCPS. With Ukraine challenges that are met by using traffic, frequency data are presented in State standard "Determination of traffic frequency and traffic stream mix" [102]. It should be indicated that the standard covers automated traffic count via fixed or mobile centres of control of traffic frequency and flow composition as well as visual observation method. This is based on the assumption that computer-aided observation stations are operating day and night on national roads. Considering the high costs of pursuing research at computer-aided observation stations, the traffic frequency data have been determined manually usig the correction factors. In such a situation, road vehicle units normalized to the car unit unify the value of traffic frequency. Instrumental control of traffic frequency in forwarding and opposite

directions simultaneously provides insight into road conditions on a going basis. The standard discourages the use of instrumental control at the sites of road services, public transport stops, pedestrian crosswalks, and upstream traffic lights.

The solution of deep analysis of traffic video by means of AI is an automated system for carrying out field data processing and traffic reports. The advantages of this technology are the possibility to recognize each target, to obtain trajectory statistics, and to export trajectory data revealing the route covered, the speed, the acceleration, and the deceleration for each object followed. OD-matrices determine traffic volumes on routes, including quantitative and temporal indicators. The results visualization unit is developed to a high level. Safety analysis is based on surrogate measures of safety estimation, such as time-to-collision, post-encroachment time, and heavy braking.

By comparing the computer vision technologies of two different suppliers, one can argue that their accuracy in determining the frequency of traffic is in most cases at the level of a human eye.

Chapter 6:
Communication Systems and Tools for Cyber-Physical Systems

Nadezhda Kunicina
Riga Technical University

Co-authors:

Rasa Bruzgiene
Kaunas University of Technology

Andrejs Romanovs
Riga Technical University

Igors Utesevs
Riga Technical University

Antons Patlins
Riga Technical University

Anatolijs Zabasta
Riga Technical University

6.1. Introduction

In its broarder sense, a cyber-physical system (CPS) is perceived as a system of systems, which could be both monitored and controlled by specialized computer algorithms and for which integration with Internet and its users is also a common feature. More specifically, this term is defined in NIST SP 1500-201 framework, which states that such systems are "smart systems in which physical and computing components interact using communication networks, technologies, methods and tools" [103]. This means that sensing, computation, communication, and control capabilities are internetworked to manage and control both physical and cyber entities and processes (Fig. 6.1).

In cyber-physical systems, the components related to physical layer and software layer are convoluted in such a way that each could operate at different spatial and temporal scales, demonstrating various and different behaviours. Further, these layers may interact between each other in multiple ways. CPS incorporates interdisciplinary approaches combining the scientific theory of cybernetics, mechatronics, design,

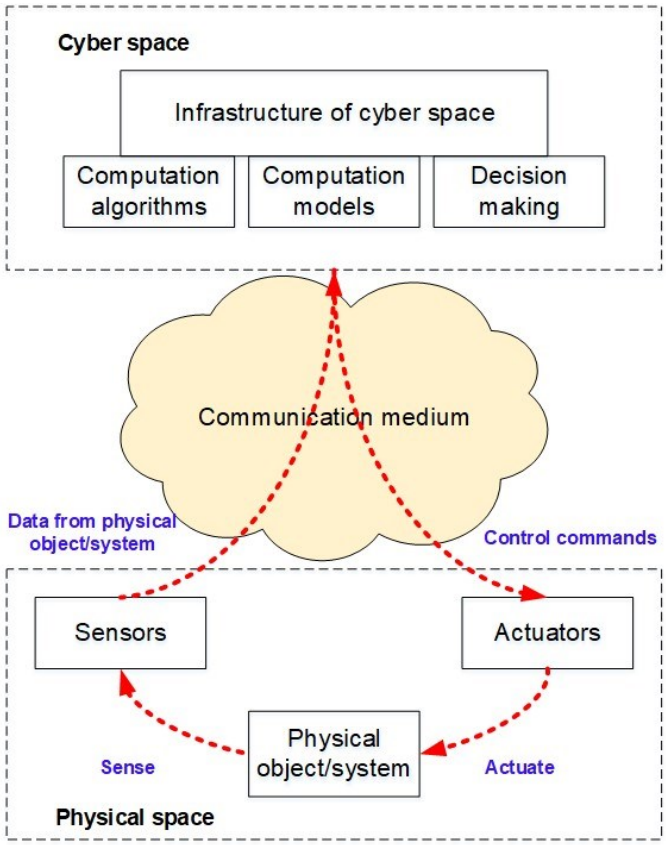


Fig. 6.1. Interaction between cyber and physical spaces in cyber-physical systems

automation, electronics, informatics, communication, and other engineering processes. The application of CPS covers critical service areas, which are important for the functioning of society, such as health system, energy, transport, buildings, supply systems, IT and networks, security of society and defence of infrastructure systems, industry, and household. Examples of such systems include smart grids, autonomous automotive systems, monitoring applications for medical system, robotics, automated avionics, process control systems, and others. Interestingly, the precursors of modern cyber-physical systems can be found in multiple diverse fields, such as transportation (especially aviation or automotive), civil infrastructure", which includes energy, telecommunication, healthcare, chemical and manufacturing processes to entertainment, and complex portable or stationary consumer devices.

It is important to mention that the concept of cyber-physical system is often closely associated with Industrial Internet, Internet of Things (IoT), Machine-to-Machine (M2M), concepts which share several features, and thus, one can notice overlaps in their meaning. CPS and IoT are often identified as the same system. This is both right and wrong. It is correct to use these two terms interchangeably. However, there might be some differences between the two, depending on the physical and logical architecture and implementation of the system itself. Figure 6.2 shows some of the basic differences between cyber-physical systems and Internet of Things systems. Physical objects (PH1, PH2, PH3, ... PHn) are connected by physical interfaces in a physical environment. The communication of these objects takes place using Internet, therefore its data processing is also in the digital environment (PHc1, PHc2, PHc3, ..., PHnc). In this case, the interface between cyber and physical environments, through a way of communication creates the concept of CPS. However, the Internet of Things is on the level of physical objects. If, for example, IoT is used in case of a car traffic monitoring, only the connection between the digital environment is used, but not the object in it. In this case, the Internet of Things creates a ubiquitous connection in the physical environment.

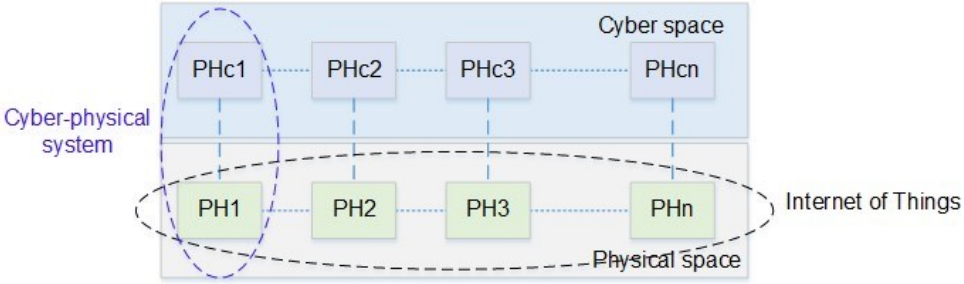


Fig. 6.2. Cyber-physical systems vs Internet of Things

Based on the general definition of CPS, communication between cyber and physical environments plays a key role in the reliable operation of such systems. CPSs are applied in areas where systems can vary from stand-alone to highly complex in

an architectural point of view. Communication can also be implemented in different ways, from client-server communication to communication solutions based on intelligent technologies. In this case, cyber-physical systems must support different types of technologies, networks, and protocols.

6.2. Communication protocols

The reliability of interaction between physical and cyber components is directly dependent on the communication technologies and protocols used at different levels of the CPS architecture. The general CPS architecture can consist of 5 main components [104], which describe the basic process of system operation:

- Connection – stage of CPS operation intended to receive external information and data via sensors or via a network of sensors.
- Conversion – stage of CPS operation intended to manage the conversion from data to information, data processing and analysis.
- Computation – stage of CPS operation intended to collect and analyse information from overall nodes of a CPS and create a communication way for components of the system.
- Cognition – stage of CPS operation intended for decision-making under collected and analysed information.
- Configuration – stage of CPS operation intended to achieve response to external factors and changes in system's environment together with control and adjustment of CPS components.

The communication protocols and technologies can be classified on the basis of general CPS architecture (see Fig. 6.3) at the same time maintaining the interface with a general TCP/IP (Transmission Control Protocol/Internet Protocol) network model [105].

In a cyber-physical system, a large amount of data is generated by sensors located in the physical CPS infrastructure at the physical level. This sensed data is sent to access and core networks using the appropriate infrastructure (switches, routers, etc.). In the cyber part, flows of big data are processed, analysed, decisions are made according to appropriate computational algorithms, and then control commands are sent to equipment at the physical level (actuators, controllers, microcontrollers, etc.). These control commands are sent through the same communication medium.

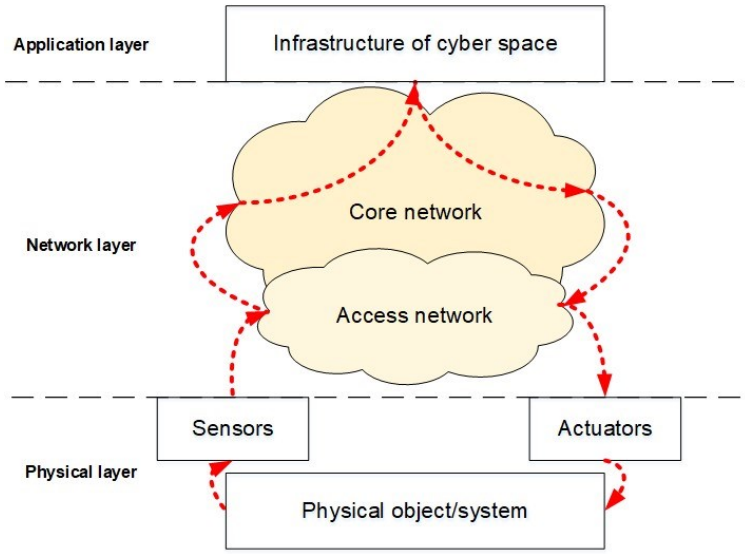


Fig. 6.3. CPS architecture on top of TCP/IP model

The general communication protocols in TCP/IP stack serves more in casual office work applications; however, CPS systems also use field bus protocols, more for industry processes. In this case, the CPS can be classified into categories according to the covered geographical area [106]:

- cyber-physical systems that cover small or personal geographic area (C1);
- CPSs that cover relatively limited geographic area (C2);
- CPSs that cover from large to very large geographic area (C3);
- CPSs that are mobile (C4).

Cyber-physical systems that cover a small or a personal geographic area are typically located on the personal body or are embedded in a particular device. Examples of this CPS category could be cyber-physical systems for human healthcare [107], [108], self-driving vehicles [109], unmanned aerial vehicles (UAVs) [110], and autonomous underwater vehicles (AUVs)[111].

CPSs that cover relatively limited geographic area are typically located in a house, a building or a small manufacturing area. Examples of such CPS category include smart buildings [112], smart manufacturing control, smart manufacturing monitoring, control of a greenhouse [113], wind power plant, sun power plant, and hydropower plant [114].

Table 6.1. Communication Protocols in CPS Wireless Networks

Communication protocol	Category of CPS	Description
IEEE 802.15.4 (Zigbee)	C1	Data rate: 20 Kbps – 250 Kbps Range: 10–20 m
IEEE 802.15.1 (Bluetooth v.5.2)	C1	Data rate: up to 1400 Kbps Range: 10–245 m
IEEE 802.11a	C1, C2, C4	Data rate: 6, 9, 12, 18, 24, 36, 48, 54 Mbps Range: 120 m outdoors
IEEE 802.11b	C1, C2, C4	Data rate: 1, 2, 5.5, 11 Mbps Range: 140 m outdoors
IEEE 802.11g	C1, C2, C4	Data rate: 6, 9, 12, 18, 24, 36, 48, 54 Mbps Range: 140 m outdoors
IEEE 802.11n	C1, C2, C4	Data rate: 15, 30, 45, 60, 90, 120, 135, 150 Mbps Range: 250 m outdoors
IEEE 802.16 (WiMAX)	C3	Data rate: 2–75 Mbps Range: up to 56 km
LPWA (LoRaWAN)	C2, C3	Data rate: 250 bps – 37.5 Kbps Range: more than 17 km
4G cellular	C3	Data rate: 300 Mbps – 1 Gbps Range: depends on cell radius
5G cellular	C3	Data rate: 1 Gbps – 10 Gbps Range: depends on cell radius

CPSs that cover from large to very large geographic area are located in an area ranging from a city size to the whole world. Examples of these cyber-physical systems cover all critical areas – energy, smart cities, water supply systems, fuel and gas supply systems, and even large power plants.

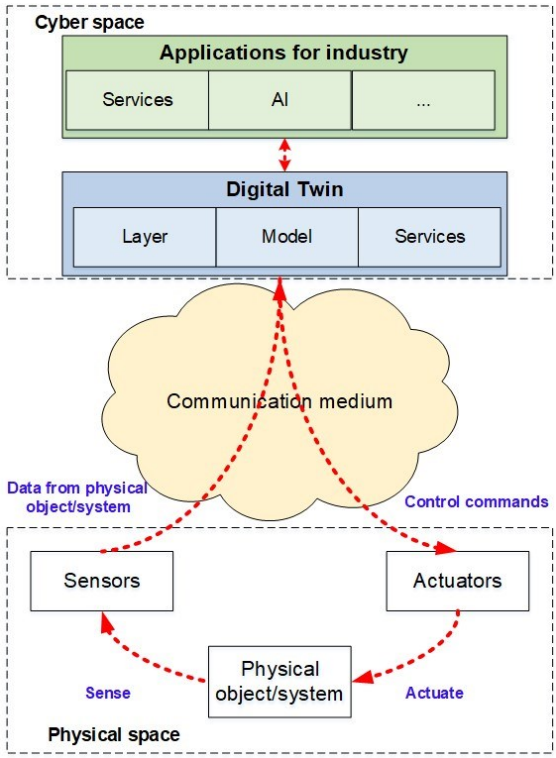


Fig. 6.4. Concept of an industrial CPS

CPSs that are mobile create mobile ad hoc networks (MANETs) [115]. The communication is based on wireless links and a multihop approach in this network, as the nodes are moving in and out of range of each other. Examples of these cyber-physical systems can be found in vehicular safety solutions as well as in UAVs.

In relation to CPS classification, appropriate communication protocols and technologies can be used. Table 6.1 presents some of the communication protocols most commonly used in CPS wireless networks. IEEE 802.15.4 and IEEE 802.15.1 protocols are used only for cyber-physical systems that cover small or personal geographic area. This is due to the physical coverage of the operation of these protocols and the relatively low data rate. IEEE 802.11 series protocols are most commonly used in C1, C2, and C4 categories of cyber-physical systems. 4G and 5G communication covers urban, national and cross-border geographical areas, making it the most suitable way for communication in the CPS of C3 category. Depending on the purpose and area of application of the CPS, IEEE 802.16 and LPWA protocols can also be used for communication of C3 systems. Communication technologies based on these protocols cover a fairly large geographical area – WiMAX covers up to 56 km, while LoRaWAN covers more than 17 km.

Table 6.2. Communication Protocols for Industrial CPSs [104], [116], [117]

Communication protocol	Description
Modbus	Serial communication protocol; used for Programmable Logic Controllers (PLC) and various control devices; functioning in full-duplex RS232 mode, half-duplex RS485 mode; compatible with TCP/IP stack; communication over master-slave or server-client architecture
Profibus	Broadcast protocol; communication over master-slave or server-client architecture – master devices are active stations in a linear bus architecture, while slave devices, which are passive stations, are usually peripherals (I/O devices, drives, etc.)
OPC	Object Linking and Embedding for Process Control protocol; used for process control, manufacturing automation applications; enables access to data from a source and communication to any user application
Profinet	Process Field Net protocol; used for data collection from and communication over Ethernet in industrial CPS
P-Net	Used for data communication in industrial CPSs, especially in automation and process control systems
WorldFIP,	World Factory Implementation protocol; used for process automation CPSs
INTERBUS	Serial protocol; used for data transmission between control devices and I/O modules, which are connected to sensors, actuators in industrial CPSs
SwiftNet	Used to perform application to application communication in industrial CPSs, especially in finance area CPSs
CC-Link (CLPA)	Used for communication between controllers and field devices as well as backbone communication along a manufacturing line system
HART	Highway Addressable Remote Transducer Protocol; operates in analog-digital industrial automation CPSs; communication over network topologies like point-to-point, multipoint and wireless mesh
VNET.IP	Protocol dedicated to a plant network system for process automation; based on Ethernet; enables communication over 1 Gbps Ethernet

TCnet	Used for connection and interchange of data between control systems; based on Ethernet; compatible with TCP/IP and UDP/IP stack
EtherCAT	Ethernet-based fieldbus protocol; enables real-time computing in automation CPSs
Ethernet Powerlink	Ethernet-based real time protocol; used for transmission of time-critical data and less time-critical data as well as for time synchronization of system nodes
EPA	Ethernet for Plant Automation protocol; used in monitoring and controlling of CPS components according to the analysed data; mainly is used in CPSs for environmental and gas monitoring
SERCOS	Serial protocol; used for real-time communication between industrial controls, drives, I/O devices
RapieNet	Real-time Automation protocol for Industrial Ethernet; enables real-time data transmission in industrial CPSs; communication over ring or line network topology
SafetyNet p	Used for fieldbus communication over Ethernet in industrial automation CPSs; application examples can be found in industrial robots, factory automation, transport technology areas
OpenSAFETY	Serial control protocol; communication in publish-subscribe and client-server architectures; enables hardware-based safety functions; used for safety relevant data transmission over industrial CPSs
MECHATROLINK-II	Used for connection of a drive to a field network in industrial automation CPSs; enables operation of a drive, monitoring of system status; communication over master-slave architecture
Wireless HART	Used for wireless communication in order to process data on wireless link; based on IEEE 802.15.4 protocol; low power protocol
WIA-PA	Wireless Networks for Industrial Automation/Process Automation protocol; used for communication in Chinese industrial automation CPSs
ISA100.11a	Used for wireless communication in industrial cyber-physical systems; based on IEEE 802.15.4 protocol; enables wireless operation of monitoring and control in CPS applications.

In cyber-physical systems, physical-level peripherals, such as sensors and controllers must be able to perform multidisciplinary tasks, especially when it comes to the application of CPSs in industry. In general, physical layer, network layer, and application layer components (peripherals and network equipment, digital twins, industrial applications, communication, and security solutions) are developed independently of each other, as it

was used in industry beforehand (Fig. 6.4). Moreover, it is important to emphasize that it is the seamless interaction of these components that creates added value for the use of cyber-physical systems in modern industry. In this case, communication protocols primarily used for data routing and networking in access and backbone networks will not be suitable. Special communication protocols are implemented for both wired and wireless connections. Interaction between physical-level hardware and software, in an industrial environment, is based on it.

Table 6.2 presents the most common communication protocols for industrial CPSs. In industrial cyber-physical systems, data collected by equipment at the physical level, using specialized communication protocols and technologies, are transmitted to CPS nodes, which store, process, and send data for further processing into system components.

6.3. Wireless communication technologies

As shown in Table 6.1, Zigbee (IEEE 802.15.4), Bluetooth (IEEE 802.15.1) and Wifi (IEEE 802.11) wireless technologies are most commonly used for communication in cyber-physical systems. WirelessHart and ISA100.11a are the most popular wireless technologies in industrial CPS.

Zigbee (IEEE 802.15.4) is a high-level wireless data transmission technology based on the IEEE 802.15.4 protocol for small networks. It provides efficient energy saving algorithms and smart data routing. This technology is used for automation of residential and industrial premises, automation of production processes and remote monitoring of parameters. The main advantage of Zigbee-based devices is the simplicity of designing new systems and integration of new solutions into the existing system, mainly because it is not necessary to build new or change existing data transmission lines and does not require a permanent power supply. The main features of this technology are:

- available for short-range (10–20 metres) communication;
- operates in beacon mode and thus allows for the implementation of energy-efficient communication in CPSs;
- does not guarantee quality of services;
- low data rate (20 Kbps – 250 Kbps);
- supports different network topologies: peer-to-peer topology, star topology and mesh topology;
- for operation, two types of devices are needed: a full-function device (FFD) and a reduced-function device (RFD) [118];

- wireless channel access through CSMA-CA and slotted CSMA-CA;
- latency down to 15 ms and cannot comply with reliability requirements;
- uses 128-bit AES data encoding algorithm.

With the evolution of cyber-physical systems, Zigbee has become a special technology used in wireless personal area networks (WPAN). This is due to its feature to support low-cost wireless communication. Topologies that Zigbee supports include peer-to-peer networking, star networking and mesh networking (Fig. 6.5). Zigbee technology is not suitable for cyber-physical systems that require higher rates for data transfer. The reason is that the technology itself supports low data rates and does not use Request to Send (RTS) and Clear to Send (CTS) notion in order to avoid data collisions.

Sensors play a key role in the process of industry digitalization. Obviously, wireless sensors and actuators demand for wireless sensor and actuator network (WSAN). It is suitable to use a wireless sensor network (WSN) and its functionality as well as the possibilities in order to create such WSAN network. A wireless sensor network (WSN) is a wireless communication network composed of spatially located autonomous devices that use sensors to monitor physical signals or environmental conditions. Arranged measurement nodes wirelessly transmit signals to a network node that acts as a network coordinator to verify the authenticity of the nodes, form a message queue, and match the wireless technology-based network to a wired Ethernet, where data can be transmitted for further processes. In this case, WSN

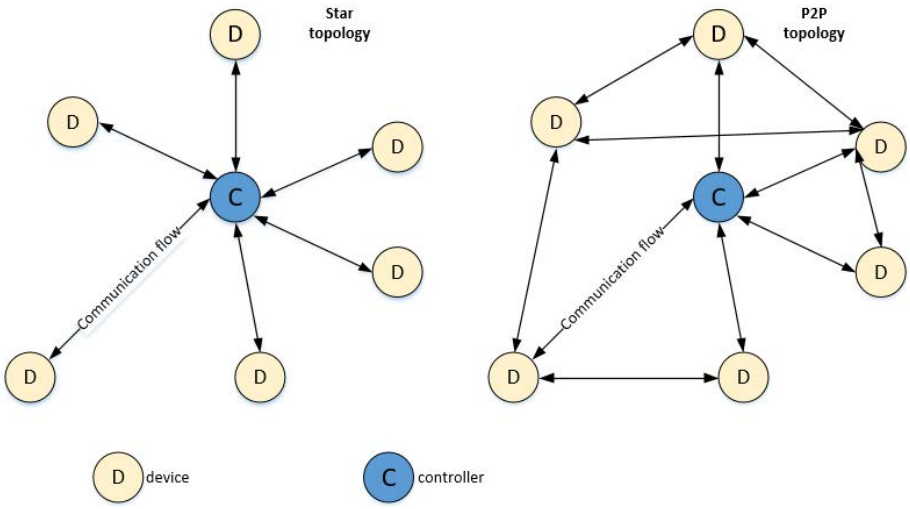


Fig. 6.5. Network topologies for Zigbee based communication in cyber-physical systems

based on IEEE 802.15.4 can be a suitable solution for communication in a cyber-physical system, for example, in categories C1, C2 and even C3.

Bluetooth (IEEE 802.15.1) is another available wireless technology for low cost communication in cyber-physical systems. Bluetooth version 5.2 was introduced in 2020 with new low energy (LE) power control feature [119]. The new version of this technology provides faster data rate – up to 1400 Kbps, from 15 % up to 50 % lower power consumption due to shorter Tx/Rx (transmit/receive) and a range multiplier for 0.8x. The operation of Bluetooth v.5.2 is based on Enhanced Attribute protocol (EATT), which supports new Logical Link Control and Adaptation protocol (L2CAP) mode. Bluetooth technology features include:

- LE power control by dynamically changing the Tx power level based on Rx signal strength indicator (RSSI);
- low power consumption by dynamic power management implemented between connected devices;
- reliability improvement by using active maintenance of receiver signal strength;
- improvement of interference with other wireless devices operating in the 2.4 GHz frequency band over a specific system environment.

WiFi (IEEE 802.11) is another wireless technology widely used in cyber-physical systems. This technology has a higher data transfer rate and coverage than Zigbee or Bluetooth, but it uses more energy than the latter two. In 2019, the WiFi Alliance introduced the new WiFi 6 standard (IEEE 802.11ax), and the first devices to support it appeared that measure data transfer speeds in gigabits per seconds. It was approved as IEEE 802.11ax-2021 standard in 2021. The maximum theoretical WiFi 6 data transfer speed can be as high as 10 Gbps. Against this background, the use of this technology for communication in cyber-physical systems is becoming increasingly popular. WiFi 6 is designed to operate in the frequency range from 1 to 7.125 GHz, so today its two designations of it marking are accepted:

- WiFi 6 – operates in 2.4 GHz and 5 GHz bands [120];
- WiFi 6E – operates in 6 GHz [121].

The advantage of this technology lies in the higher bandwidth allowed by the higher spectral efficiency. IEEE 802.11ax uses orthogonal frequency division multiple access (OFDMA), which can be equivalent to cellular technology. Along this line, more efficient use of spectrum is also driven by energy control mechanisms

that avoid interference with other networks. To further increase the throughput, up-link direction was added with down-link of MIMO (Multiple-Input and Multiple-Output) and MU-MIMO (Multi-User, Multiple Input, Multiple Output), as well as higher order 1024-QAM (Quadrature amplitude modulation), added security TWT (Target Wake Time), and WPA3 (Wi-Fi Protected Access) protocols. The use of WiFi technology makes it possible to enable standard security schemes that would ensure the confidentiality, authenticity, and availability of data in cyber-physical systems.

WiFi HaLow (IEEE 802.11ah) is a wireless technology, that WiFi Alliance specially introduced for Internet of Things systems [122]. It can be applied in a frequency band lower than 1 GHz, but at the same time it can facilitate additional user accesses and low power consumption. The predecessor of WiFi Halow was the IEEE 802.11ac standard. In this case, the modification of WiFi Halow physical and Link layers is related to the reduced clock rate of 802.11ac. The physical layer of WiFi Halow is divided into a) 2 MHz, 4MHz, 8MHz and 16 MHz transmission modes and b) 1 MHz transmission mode. This wireless technology uses MCS10 modulation encoding mechanism for a long-distance data transmission. Communication in WiFi Halow can be organized using peer-to-peer and star network topologies. The main features of WiFi Halow that make it attractive for communication over CPS are:

- large geographical coverage;
- less-energy consumption;
- support of native IP (Internet protocol) addressing;
- support of large amount of system devices for connection establishment.

ISA 100.11a is a prominent wireless technology, created especially for industrial systems. In industry, sensing of data and the communication process must be extremely accurate. The update frequencies can vary from 100 ms up to less than 1 ms. ISA 100.11a is based on IEEE 802.15.4 standard, and it enables wireless operation of monitoring of non-critical areas and control in CPS applications.

WirelessHART was also based on IEEE 802.15.4 wireless standard only for the purpose of operating in industrial systems. Its features provide simple and flexible installation, access to data ensuring robustness and reliability in IT-based communication. WirelessHART enables wireless communication in order to process data on wireless link. It can also be used for mission critical applications, assurance of end-to-end delay in CPS. This technology has low energy consumption, which affects the longer life of wireless devices (such as sensors). The security of WirelessHART is based on AES-128 (Advanced Encryption Standard) encryption.

6.4. Data transmission methods over several CPSs

When it comes to the wide range of applications of such systems – industry, transport, and health – cyber-physical systems must be able to interact with each other. Interoperability takes place through network interfaces of a particular communication technology, using network protocols and data transmission methods. Data exchange between different cyber-physical systems takes place using Application Program Interface (API). Depending on the type of API, several different models for data transfer are distinguished – request/response, publish/subscribe, and event-driven models.

A model that describes request/replay for data transmission/exchange between several cyber-physical systems is presented in Fig. 6.6. The request/response data transmission model is typically used in service-oriented architectures. During this process, the query parameters are presented to the end-node of a system. The end-node of a system determines the type of service; the descriptive attributes of the service and the signature are provided according to the query parameters.

The sensing of the usage of smart meters in the field of energy can be analysed as an example for this:

Read_Meter_Usage_Data(string meterID, date startDate, date endDate|)

The purpose of such a request is to read data from a meter, such as its ID (identity description), start time of its usage, as well as end time of its usage. Such a data transmission model is quite easy to implement, as the structure of requests and responses is typical. However, in such service-oriented architectures, the structure of the data transmitted itself must be either known in advance or identified during a system operation. Implementation of a service-oriented API is based on Service Oriented Application Protocol (SOAP) [123]. The disadvantage of query/response model in a service-oriented architecture is that APIs are less flexible.

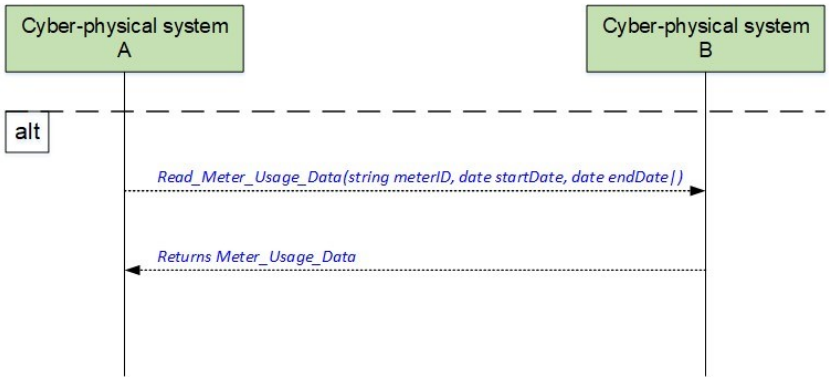


Fig. 6.6. Data exchange in service-oriented API [103]

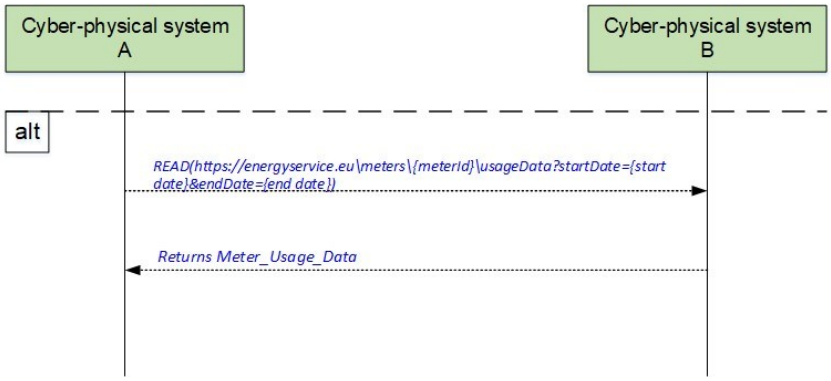


Fig. 6.7. Data exchange in data-oriented API [103]

In a data-oriented architecture, data exchange is performed using the references to data (Fig. 6.7). The parameters of these queries are like filters for the results returned as a data set in the form of a response. The use of such a model allows for a relatively flexible use of APIs for different purposes but requires a prior understanding of the complex data structure. Implementation of a service-oriented API is based on Representational State Transfer (REST) [124]. The disadvantage of such a model in a data-oriented architecture is that limits of data service signature can impact the access to it, especially if the data is highly structured and bundled into data sets. For example, the same case for sensing the usage of smart meters can be analyzed:

READ(https://energyservice.eu\meters\{meterId}\usageData?startDate={start date}&endDate={end date})

READ service is used for the link to the meter usage data, while query parameters include a start time and an end time of meter usage.

Unidirectional signals can also be used in exchange of data between multiple cyber-physical systems (Fig. 6.8). In this case, the data or its flow is simply transmitted in a line from the source to the destination node of the system. Obviously, neither requests nor responses are sent. Data transfer is performed as a default, uninterruptible bond.

In addition to the models mentioned above, there is a fourth one – publish/subscribe model for data transfer between multiple cyber-physical systems (Fig. 6.9). Besides the source and destination nodes, a broker node must be installed in this model. The role of the broker is to record the data received from the source node (publisher) and send it to the destination node (subscriber) as soon as the data is subscribed.

Identification of data is based on identification tags, which is known to broker. According to these tags the data is subscribed. When the subscribed data reaches the

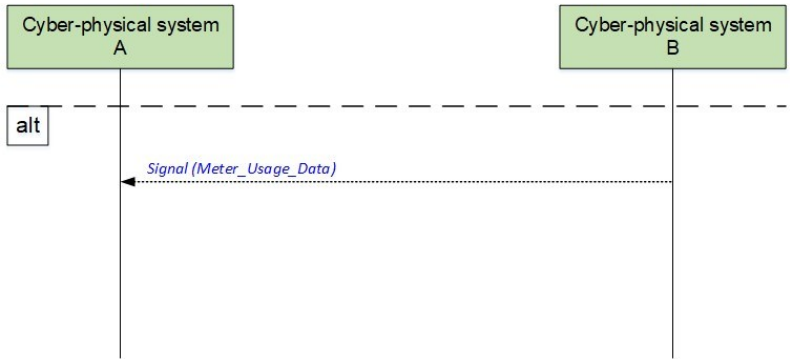


Fig. 6.8. Data exchange in signal-oriented API [103]

broker node, it is transferred to the subscriber of that data. The broker node can be implemented eithers as a middle node in the system or as an additional node of each node which publishes data. The disadvantage of this data exchange model is the trust and security issues related to broker node.

6.5. Software-defined communication in cyber-physical systems

Software-defined communication allows the development of such cyber-physical systems in which demand for placing control methods from the hardware at the physical level, to the software-defined control level is defined. Software-defined communication is applicable in real-time cyber-physical systems, which must be characterized by reliability, cost-effectiveness, *etc*.

Looking beyond the design challenges of CPS, sensing of data in real-time cyber-physical systems results in really large amount of data, called big data. In this case,

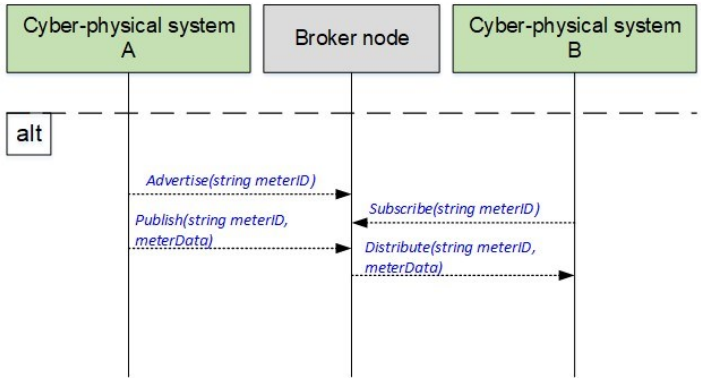


Fig. 6.9. Publish/subscribe model for data exchange [103]

a task for filtering and analysing of those data should be solved by appropriate algorithms. In the context of CPS scalability, administration of heterogeneous devices (hardware, software) as well as control and configuration of right APIs is another important process for suitable operation of CPS. In line with it, allocation of resources and making of decisions should also be performed in real-time by so called "online" computing algorithms. In brief, distributed, decentralized, and software-defined communication can help with the proper abstraction of the operation and control processes between physical and cyber environment.

Wireless technologies are the most widely used for communication in cyber-physical systems. However, in wireless communication, where many peripherals are connected, efficient use of energy resources and radio channel for the wireless signal transmission must be ensured. In this case, Software-Defined Radio is one of the technologies that allow efficient, software-based real-time management of radio channel utilization. For example, the use of cyber-physical systems based on SDR communication can be found in agriculture [125], as control of radio channel in basic – hardware – way is quite challenging there. The reason is behind the signal attenuation to the proximity to the surface of the earth.

Cognitive radio (CR) is another optional technology for communication of CPS. Basically, a cognitive radio evolved from SDR, which has the ability to sense the environment in which it is implemented, to track changes and to react to any findings on the changes. Cognitive radio exchanges data frequently in the communication network and can offer access to other units of CR. The use of cognitive radio in cyber-physical systems has arisen due to CR's ability to make efficient use of the radio electromagnetic spectrum. More specifically, it makes efficient use of the still underused radio frequency bands by sharing and allocating the radio spectrum in relation with the state of the radio channel. For example, a cognitive radio technology can be applied to cyber-physical systems, where communication is performed through

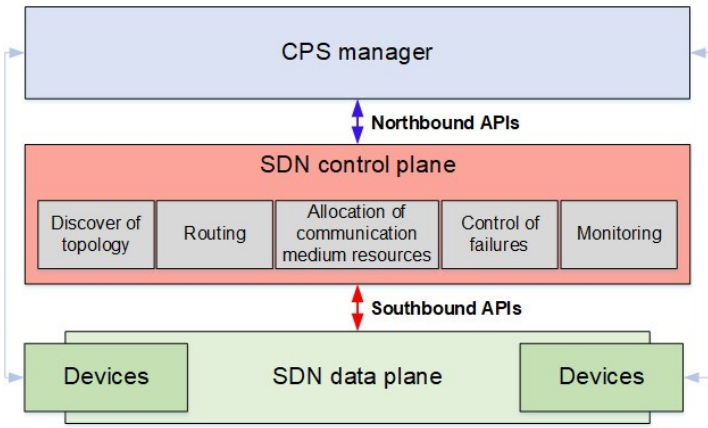


Fig. 6.10. SDN in cyber-physical system

a wireless sensor network. In this case, sensors sense the environment nearby and adapt to any changes caused by the environment by modifying the parameters of data transmission in real-time. Modification of the parameters are done according to the information that comes from the environment. The examples of CR application in cyber-physical systems can be found in smart transport systems [126], where vehicular can be connected with different applications in relation to green transport, smart assisted, automated or self-driving, road safety, etc.

Software-Defined Network (SDN) presents an appropriate solution upon the separation of the processes of communication and control in the architecture of cyber-physical systems. Allocation of the resources or the most reliable paths without global data from the network, especially if it is a wireless network, is a particularly crucial challenge. In this case, the SDN controller is able to find the most reliable paths on known parameters, from the wireless link – bandwidth, latency, etc. Separation of application layer from the infrastructure layer by SDN controller allows the implementation of multiple network functions, for example, routing of data, data traffic filtering or monitoring, at the same time exchanging the specific information between cyber-physical spaces (Fig. 6.10).

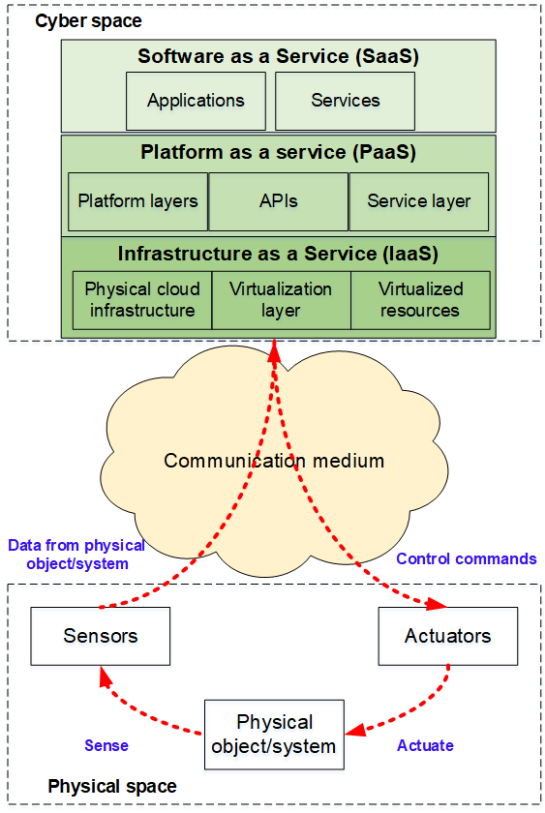


Fig. 6.11. Concept of a cloud-based cyber-physical system

There are two major types of APIs – northbound APIs and southbound APIs – in such CPS architecture. Northbound APIs are responsible for interaction of modular applications with control logic of the network. Southbound APIs translate network policies to control instructions and lock up the complexity of the network. In this case, a SDN controller serves as a gateway for conversion of northbound APIs to changes in data paths.

6.6. Cloud-based cyber-physical systems

Cyber-physical systems stand out for their interaction between physical and cyber environments. The information collected with the help of peripheral equipment is sent to the cyber environment via the communication network for further processing and decision making, and for sending commands back to the equipment in a physical environment. Therefore, it is very important to properly select and implement models and technologies for performing such actions in a cyber-environment.

Digital twins are one of the computational modules used in today's cyber environment. It processes the received data and, as a result of decision-making on its computations, sends control commands for the necessary changes, i.e., changing of certain parameters and reconfiguration, to the physical equipment. Digital twins are like an image or virtual copy of a physical system in a cyber environment [127].

Cloud technology delivers computing power and "space" in cyber-physical systems. It means that the performance of infrastructure resources and capabilities of data analytics can be provided within cloud technology at a high level. Moreover, the improvements on control of a feedback in CPS can be achieved as well. Integration of cloud technology into CPS can be done by Software-as-a-Service (SaaS), Infrastructure-as-a-Service (IaaS) or Platform-as-a-Service (PaaS) models. Figure 6.11 presents a basic concept of a cloud-based cyber-physical system.

IaaS covers the layers responsible for the physical infrastructure in the cloud, virtualization, and virtualized resources. PaaS covers the layers of the platform, APIs as well as the layer of the service. Obviously, PaaS relies on the virtualized resources from IaaS. The SaaS layer covers all applications as well as services provided to the users as a service. Cloud-based cyber-physical systems can be applied in areas such as the manufacturing industry, robotics, automated systems, energy plants, and commercial logistics chains.

Chapter 7:

Embedded Systems Used for Cyber-Physical Systems

Volodymyr Kasymyr
Chernihiv Polytechnic National University

Co-authors:

Oleksandr Khropatyi
Chernihiv Polytechnic National University

Oleh Lohinov
Chernihiv Polytechnic National University

7.1. Role and importance of embedded systems in CPS

The creation of an embedded system that closely interacts with physical processes requires a technically complex design at a low level. Embedded system designers struggle with interrupt controllers, memory architectures, and assembly-level programming to use specialised instructions or fine-tune timing, device driver design, network interface design, and process scheduling, rather than focusing on defining the desired system behaviour. The mass of deviations from the main course of solving the final problem and the complexity of these technologies force us to look for new approaches to design.

Experts from the University of California at Berkeley believe that such approaches should be based on modelling the entire system and technologies for joint design of hardware-software, networks and physical processes [128]. This approach is called cyber-physical – from the term "cyber-physical systems" (CPS), which was introduced by Helen Gill in NSF, USA, around 2006, to indicate a view of systems from the standpoint integration of computing and physical processes.

Cyber-physical systems (CPS) are systems that enable the integration of computing and physical processes. Embedded computers and networks in the form of defined models in this respect are able to control the necessary processes, usually with feedback loops, where physical processes affect the computations performed and vice versa. The design and mechanisms of these types of systems require an understanding of the collaborative dynamics of computers, software, networks, and physical processes.

In CPS, the time required to complete a task can be critical to the proper functioning of the system. Moreover, in CPS, multiple processes are performed at the same time. Physical processes are at the heart of the work of cyber-physical systems. They denote the totality of many tasks happening at the same time as opposed to programming processes that are deeply rooted in successive steps. In addition, concurrency is an integral part of CPS. Many technical issues in the design and analysis of embedded models are related to ensuring the need to link sequential semantics to the internally parallel physical world. Working with embedded models has a big advantage. Models in this respect can have formal properties with any possible physical implementation of the system. The model instils confidence in the physical implementation, and the study of embedded systems allows us to understand how these systems will behave in the physical world [129].

Cyber-physical systems define a basic methodology for representing the transformations of various types of technologies and environments. This ensures the exchange of various information between these components, as well as the adequate operation of all embedded systems in different control conditions. CPS can only be developed with preliminary simulations. This is often referred to as creating a

"counterpart" for a given CPS which must interact with other digital counterparts. From this, it becomes clear that a modular approach will be used to combine these views. In addition, standardized communication is required between the digital counterpart and the given CPS, which requires the use of specific interoperability standards [130].

7.2. Main features of embedded systems

Before moving on to the issue of embedded systems, some basic concepts need to be explicated. A system refers to the many elements that are in relation to and connection with each other. It can also be defined as a way of working, organising, or performing one or more tasks according to a fixed plan. An embedded system in simple terms means something that is related to something else. An embedded system can be thought of as a computer hardware system in which software is embedded. These types of systems can be independent systems or part of a larger system. An embedded system can be presented on the basis of a microcontroller or a microprocessor, IoT or cyber-physical systems. The system must be designed to perform a specific task [131].

An embedded system consists of three components:

- application software,
- hardware subsystem, and
- real-time operating system (RTOS).

RTOC monitors the application software and provides a mechanism for the processor to start a process on a schedule, following a latency control plan, defining how the system works. It sets up rules at runtime. A small embedded system may not have an RTOS.

Thus, the embedded system is based on a microcontroller, software controlled, reliable real time control system. To build application systems and embedded models, certain characteristics of the embedded system must be observed [132].

Real time work

When we say that an embedded system must operate in real time, we mean that the system must perform certain computations in strictly defined time intervals. If the system cannot perform the necessary calculations in the allotted time interval, then at best the object of its control will operate with low technical characteristics, and in the worst case an emergency situation will be created. By using the term "real-time computation", we mean that the time interval provided for these computations is limited. Moreover, its numerical value is determined by a specific problem and

can differ significantly for different systems. For example, the anti-lock system of the wheels of a car must interrogate the status sensors of each of the four wheels (the wheel is sliding or rolling) and generate the necessary signals for the brake actuators within a few milliseconds. The system must be designed in such a way that the required computation cycle falls within the allotted time interval. To achieve this, first, it is necessary to select the appropriate computing performance of the microcontroller, develop an algorithm that is efficient in terms of speed, and develop interface circuits with the lowest possible delays in signal transmission. Second, the embedded system must be robust against external data. For example, to generate the result, the system must receive data from the outside. And if these data do not arrive on time, the system cannot produce the required result at the required time, however, it should not "hang". It should continue to deliver results in real time, but in a different, possibly abbreviated form.

Miniaturisation of sizes

Many modern systems need to be built into rather miniature devices such as a mobile phone, TV remote control, and water flow sensor. The geometry of the printed circuit board of a system is often determined by the case of the device for which it is intended. Therefore, miniaturisation of execution is one of the problems of the developer of modern embedded systems.

Minimising energy consumption

General-purpose computer designers (excluding laptops) pay significantly less attention to device power consumption than embedded system designers. The fact is that, first, personal computers are powered from a centralised network, which does not impose significant restrictions on energy consumption, and, second, the volume of a personal computer case is large enough to accommodate a forced cooling device. In contrast to general-purpose computers, today's embedded systems must operate under drastic power constraints, as the number of self-powered embedded systems continues to grow. In addition, users are increasingly demanding the miniaturisation of systems. Think of the modern mobile phone, the pocket electronic organiser, and CD player.

To limit energy consumption, developers use different solutions. One of them is to reduce the clock frequency of the MK. However, such a measure has a limitation, because for any real-time task, there is a lower limit on computational performance. Another solution (or additional to the first) is to temporarily turn off the power to those peripheral modules of the MC that are not being used at the moment of program execution. A modern MCU hardware provides this capability. The latter method requires special attention from the developer, because the disconnection of any module in the system can lead to a change in the electrical characteristics of its inputs and outputs, which should not affect the performance of the system as a whole.

User interface and object interface

Any embedded system must interact with the user or the environment. For example, a robot moving in space must use infrared sensors to detect obstacles and avoid them. The microwave oven must interact with a person using the mode buttons installed in the front panel of the device. And the security alarm system must interact with both the sensors of the safety of the room and the human control bodies. Examples like these can be continued, and on their basis, we can conclude that for a developer of embedded systems, the issues of making decisions on interaction with a person and with a control object are an extremely important task. Moreover, possible solutions lie at the junction of the choice of the type of sensors (including the principle of operation of the sensor), the design project, the design, the hardware solution of the electronic units, and finally, the information processing algorithms.

Multitasking

Most embedded systems need to serve multiple external devices in real time. Moreover, the repetition periods of the real-time computation algorithms for each of the devices are different. When developing such systems, the developer faces a dilemma whether to use one high-speed MC to solve the problem, or to make a multiprocessor system in which for each task its own microprocessor or microcontroller will be used.

Cost minimisation

Each embedded system has many possible solutions, both at the level of the implementation method (microcontroller or programmable logic matrix, variations of interface circuits for both solutions) and at the level of choosing a specific element base. Therefore, choosing the right design strategy in order to minimise cost is one of the main problems in the design of an embedded system. It is also worth paying special attention to the important terms used in the embedded system.

Reliability

It is a measure of the probability of the system surviving when a function is critical at runtime.

Fault tolerance

Fault tolerance is the ability of a computer system to survive in the presence of faults.

Flexibility

It is an assembly of systems with built-in debugging capabilities that allows

remote maintenance.

Portability

Portability is a measure for the ease of use of the same firmware in different environments. This requires generalised abstractions between the very logic of the application program and the low-level system interfaces.

Single functionality

An embedded system usually performs a specialised operation and does the same repeatedly.

Hard restrictions

All computing systems and models in this case have limitations on design metrics, but in an embedded system they can be especially stringent. Design metrics are a measure of implementation features such as cost, size, power, and performance.

Base

Embedded models must be based on a microprocessor or microcontroller.

Memory

Embedded systems must have memory, because its software is usually embedded in ROM.

Connection

Embedded systems must have peripheral devices connected in order to connect input and output devices.

7.3. Basic concepts of embedded systems building

7.3.1. Industry 4.0

Industry 4.0 (I4.0) can be perceived as a set of concepts, methods and tools for creating intelligent technologies that can be composed of various network resources. The principle of Industry 4.0 is that by connecting machines, parts and systems, as well as smart grids, chains are created, the elements of which can independently control each other. Industry 4.0 represents a paradigm shift from “centralized” to “decentralized” manufacturing, which is facilitated by technological advances that represent the abolition of the traditional logic of the manufacturing process. The novelty introduced by I4.0 manifests itself in the adoption of the concept of cyber-

physical systems (CPSs) as a system with embedded software and electronics. They are connected to the outside world through sensors and actuators and can communicate with other CPSs through standard networking technologies.

Cyber-physical systems are the integration of computing into physical processes. Embedded computers and networks inspect and control physical processes, usually with feedback loops where physical processes affect computation and vice versa. Cyber-physical systems provide technologies that combine virtual and physical worlds to create a networked world in which intelligent objects communicate and interact with each other. In the manufacturing industry, for example, CPS includes all sorts of sensors, devices, and mechanisms that become interactive through embedded software and network connectivity to monitor and control physical processes using feedback loops. CPS collects, stores and analyses sensor data through its local business logic to provide and use data and services. Such decentralized intelligence creates an intelligent network of objects and independent process control, interacting with the real and virtual worlds. This represents a completely new aspect of industrial and manufacturing processes [133].

7.3.2. IoT

Cyber-physical systems must be connected to the network to provide the intended flow of information. This network is not limited to a factory – it can also be connected to the Internet, which makes CPS part of the Internet of Things (IoT). It is a network of physical objects or "things" embedded in an electronic system, software, and sensors, with the ability to connect to facilitate the exchange of data among people (manufacturer, operator, service technician) and/or other connected devices in order to achieve greater value and instant support. Each "thing" is identified on the network through an embedded computing system and is capable of interacting across the entire Internet infrastructure. With its connectivity to all things on the Internet, IoT offers more than established machine-to-machine communication. The interconnection of these embedded devices, including smart objects, is expected to support automation across multiple domains through universal applications. Moreover, cyber-physical systems are required to exhibit a high degree of autonomy and adaptation ideal for Industry 4.0 applications.

7.3.3. IoT architecture

The architecture of the Internet of Things for each IoT system is unique, but the underlying and overall data processing flow is very similar. The first stage is a set of objects that are connected to the Internet. These objects have built-in sensors and mechanisms with which they can collect data about the environment, as well as various information, which is then transmitted to the gateways of the Internet of Things. At the next stage, a huge amount of raw data is collected and converted into digital streams, filtered and processed in order to prepare this data for analysis. This process is achieved by using IoT data collection systems and gateways. Further

processing and advanced analysis of the data takes place using edge devices. The next stage is represented by data centres (cloud and on-premises). Here the data are analysed, stored, and processed to obtain information [134].

It is necessary to describe in more detail the layers of the architecture of the Internet of Things.

Physical layer

Connected devices that need to provide data are the backbone of any IoT system. It is necessary to use sensors in order to collect parameters from the external environment or within a given object. These sensors can either be part of a device or be independent measuring and data acquisition devices, for example, agricultural sensors, the purpose of which is to measure various parameters such as air and soil moisture, temperature, soil pH, and the effect of sunlight on plants.

Actuators are also a very important element of this stage. Using sensors, these mechanisms can change the received data into physical actions. These data are generated by special objects called intelligent, for example, an intelligent irrigation system with built-in sensors. The system analyses the situation constantly in real time using data from built-in sensors. If the system, in the course of its analysis, finds that the soil moisture in a conditional place is below the required value, then a command is given to the actuators to open the selected water valves in the required places. After the information is received from the sensors that the soil moisture has reached the required value, the valves are closed. This all happens automatically.

It should also be noted that the connected objects must be able to recognize each other as well as collect and exchange information with each other, and be able to work together in real time. It is not enough just to keep in touch with the data collection systems and gateways; however, there is a problem if the devices have a limited resource and battery power, because such tasks require large computing power and consume a large amount of energy and bandwidth.

Data collection level

This layer is very important because it is about collecting, transmitting, and filtering data. Data transfer is carried out to the cloud and to the edge infrastructure. Therefore, despite the fact that this layer is closely related to actuators and sensors, it is necessary to designate it as another stage of the IoT architecture. The process of aggregating, collecting and transferring data must be a priority because deploying a huge number of devices generates a large amount of input and output. This stage is an intermediary between the first and the rest of the levels, which, in fact, binds everyone together.

Gateways have the ability to filter, control, and determine which data need

to be sent to the cloud and which not. This minimizes the amount of information transmitted, which, in turn, reduces the cost of sending data over the network and the response time. Thus, gateways make the exchange of data between the system and the sensors less complicated. It is also possible to convert the sensor data into any format that is necessary for the rest of the system components. In essence, gateways are places where local pre-processing of data received from sensors is performed, which is converted into packets ready for sending and processing.

It should be noted that gateways also maintain security because they are responsible for information that travels in both directions. For this purpose, the necessary encryption and security tools are used. This makes it possible to prevent theft of data from the IoT cloud and provides protection against external attacks on various IoT devices.

Target device level

Peripherals are not a necessity for the architecture of the IoT, but the use of these devices has advantages, in particular in huge IoT projects. If there is limited access to data as well as a limited speed of their transfer to cloud IoT platforms, the use of edge systems can contribute to greater efficiency in the processing and analysis of IoT data, as well as a possible faster response time. How quickly and efficiently the data are analysed is very important in certain industrial IoT applications. Based on this, one can observe the increased popularity of edge computing in the industrial Internet of Things.

Peripheral devices can interact faster and more efficiently with IoT materials in real time as well as provide input in the form of analytical data, since this infrastructure can be located closer from a physical point of view. In this case, only huge chunks of data are sent to the cloud, which require the enormous power that the cloud provides.

Data centre and cloud resource tier

The cloud plays a very important role in the body of the Internet of Things. Compared to peripheral solutions, this type is used to store, analyse, and process a huge amount of information. The cloud system has a lot of power, so it can efficiently process huge amounts of data in contrast to peripheral systems that cannot cope with this task.

Cloud computing can improve performance, reduce downtime and energy consumption. There are a number of other business benefits as well. Therefore, over the past few years, this technology is gaining more widespread use.

With the necessary application solutions, the cloud can be used for business intelligence as well as to help people monitor and interact with the system, and obtain the necessary readings and data in real time.

7.3.4. High-level architecture: concept and main advantages of use

Distributed high-level architecture (HLA) modelling technology combines systems that are usually built with different structures, platforms, and products for interaction in a defined environment. The ability to include individual elements, personnel-controlled simulators and entire equipment systems (machines, controllers, boards, etc.) in the simulation cycle not only eliminates the need to simulate complex technical devices and human behaviour, but also reduces possible errors in modelling physical processes, which significantly increase the reliability of the results.

The HLA is made up of extremely important components, such as the run-time infrastructure (RTI) [135] and the federation (which has a set of modelling components called federates) that work together to address various specific issues. In different experiments, the same federations may be part of different federations. HLA technology does not impose restrictions on the internal structure of federates but defines a standard for describing information about simulation objects – the object model template (OMT). This standard ensures the interaction of federates and allows their multiple use regardless of the internal structure in different federations for modelling. Therefore, all federation objects must be described on the basis of OMT. RTI provides interaction between federations, includes services for the interaction of simulation participants, and supports different ways of synchronisation within the federation units.

When it comes to HLA, the information that federates exchange can be of two types: object attribute and interaction. Each object is characterised at any time by its state, which is determined by a set of current values of its attributes. A federation that controls an object (an object attribute) can change its state by changing the value of the attribute. With the help of RTI services, the federation transmits a new attribute value to other federations – the federation updates the attribute; the federate that takes on the new value displays the attribute.

Interactions, unlike the state of objects, are not maintained in the system permanently, they are instantaneous in nature. They are an action performed by the object of the federation and cause possible changes in the state of the object of another federation, for example, a shot (instant action) from a gun (object that performs an action) at a target (an object that can be affected by an action).

As mentioned above, all data exchange between federates occurs via RTI. The exchange mechanism is implemented in the form of a "subscription", i.e., the federation in this respect obtains specific attributes using RTI services. However, RTI only tracks data changes. Interactions are always transmitted and obtain attributes only in cases of anticipated changes in their values.

Thus, we can conclude that the simulation of HLA plays the same role as the

one played by well-known technologies such as CORBA, COM+, DCOM, and others. A detailed comparison of HLA with other distributed modelling technologies was made in [136], based on which it can be concluded that the use of distributed HLA architecture has the following advantages:

- integration of heterogeneous simulation systems, models for which have already been developed and tested;
- integration of modelling programs implemented by different developers;
- simultaneous joint development of geographically remote users with one simulation model;
- selection of various time management schemes in the simulation process, in particular, real-time mode;
- use of geographically distributed components of one model;
- use of different data for one model depending on the purpose of modelling (due to separation of data from architecture in HLA);
- selective distribution of information between interacting models, which, in turn, allows for effective scaling of systems;
- modification of distributed modelling mechanisms to ensure their compliance with specific tasks (through the use of open standards);
- interaction of all types of models due to the universality of object identification and object connections.

7.3.5. Controling E-networks

It is important to choose an aggregate theory for the development of embedded models, which will be the system providing the description of the structure of the CPS. Controlling E-networks (CENs) can help define the dynamics of aggregates. Then, the formal CPS model can be denoted as [137]

$$CPS = (\Sigma, E, \gamma),$$

where

Σ – a program structure, which is essentially a graph that consists of arcs $\Gamma \in A \times A$ and aggregates A; with the help of these programs control it is carried out;

E – a set of processes performing a controlling role, which are implemented by

aggregates A;

$\gamma: \Sigma \to E$ – an aggregated display that summarizes the definitions Σ and E.

There are some differences between Petri nets and those presented by E-net and CEN. The bottom line is that positions in these networks can have a label with many parameters that can be changed if the transition is triggered. The difference between CEN and the usual E-net lies in the additional capabilities that are manifested in the use of a set of network variables, which are presented in both digital and analog form. There are also a large number of functions for managing network transitions, as well as for determining the desired number given the state of the object. If the conditions that are associated with marking the positions and values of network variables are met, then the transitions are triggered. In the event that the transition is triggered, the values of the variables and attributes of the labels change, and the time can be delayed.

Thus, it is possible to use CEN as a control model with a specific control algorithm, the marking of the transition changes when the transition is triggered. This determines the state of the algorithm. Label and network variable attributes provide data with which the necessary calculations can be made.

Primarily,

$$CEN = (P, T, F, V, U, M_0),$$

where

$P = \{P_S, P_R\}$ – a set of positions in which there must be at least one position, and it cannot be infinite;

P_S – ordinary positions, P_R – crucial positions, $P_S \cap P_R = \emptyset$; a set of the entry positions can be a part of a set of ordinary positions $P_{in} \in P_S$ and positions, which are called boundary, is allowed $P_{in} = \emptyset$ and $P_{out} = \emptyset$, $P_{in} \cap P_{out} = \emptyset$;

T – a set of transitions in which there must be at least one transition, and it cannot be infinitive; it consists of five kinds of transitions $\{"T_T", "T_F", "T_j", "T_X", "T_Y"\}$;
$F: P \times TT \times P \to \{0,1\}$ – incidence function;

$V = V_i \cup V_o$ – a set of network variables in which there must be at least one variable, and it cannot be infinite; V_i – entry and V_o – output signals, $V_i \cap V_o = \emptyset$;

$U = \{r, \sigma, \alpha, \tau, z\}$ – a set of control functions that are responsible for triggering transitions;

$M_0: P \to \{0,1\}$ – a function that is essentially an initial marking; its task is to check

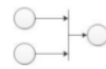
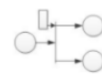
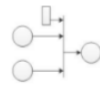
for the presence of marks in the given places.

Vector $M = (M(p_1), M(p_2), \dots, M(p_n))$ is responsible for CEN labelling, where $n = |P_s|$. Position $p_i \in P_S$ is free, if $M(p_1) = 0$, otherwise $M(p_1) = 1$ when the position is in a form of taken. For marking M the set of occupied positions is denoted as $P_M = \{p \in P_S | M(p) > 0\}$.

In order to configure the interaction of the elements of the embedded model and CPS objects during the operation of the algorithms that are responsible for control, the methodology of the HLA architecture is used, as well as RTI to ensure the connection between the objects. The aggregate receives certain attributes in the form of tags with the necessary information in XML format using the WEB protocol. Data and events that are passed through network variables are tracked using RTI. Synchronization of objects occurs in real time for all aggregates in the CPS.

The implemented five main types of parallel-sequential processes that are defined for networks are represented in the basic set of CEN transitions. Using these transitions, it is possible to create a conceptual model of a network graph, as shown in Table 7.1.

Table 7.1. Main Transition Set

Transition	Work Plan Node	E-network node
T-transition. The transition works when there is a label in the input position and no label in the original position.		
F-transition. It works when there is a label in the input position and no labels in the original position.		
J-transition. The transition is triggered by the presence of labels in both input positions and absence in the output position.		
X-transition ("switch between alternate paths"). The direction of process development is determined by the value of the decisive procedure R.		
Y-transition ("choice"). Y-junction is able to reflect the conditions of continuation of the process in case of multiple mergers. Useful for displaying feedbacks		

7.4. Embedded system architecture

The basic architecture of a typical embedded system is shown in Fig. 7.1.

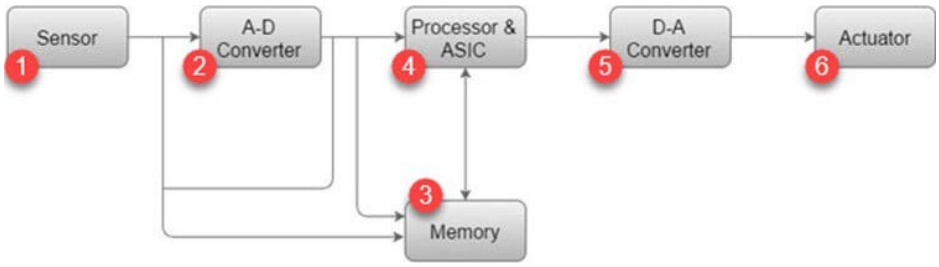


Fig. 7.1. Embedded system architecture

As shown in the figure, an embedded system architecture consists of the following building blocks.

1. Sensor: The sensor helps to carry out the process of measuring a physical quantity and then converting it into an electrical signal. It is also suitable for storing measurements in memory. The signal can be prepared by any electronic device, for example, the A2D converter.

2. AD converter: ADC (analog to digital converter) converts the analogue signal transmitted by the sensor into a digital signal.

3. Memory: It is logical that memory is usually used to store information. In general, the embedded system contains two memory cells: volatile and non-volatile.

4. Processor and ASIC: This component is responsible for data processing. This is necessary in order to measure the result and then store it in memory.

5. DA converter: DAC (digital to analogue converter) helps convert digital data transmitted by the processor into analogue data.

6. Drive: The drive compares the output provided by the DA converter with the actual output stored there and stores the validated output in memory.

Three types of embedded systems [137] are shown in Fig. 7.2.

Small-scale embedded systems

This type of embedded system must be associated with a 16- or 8-bit

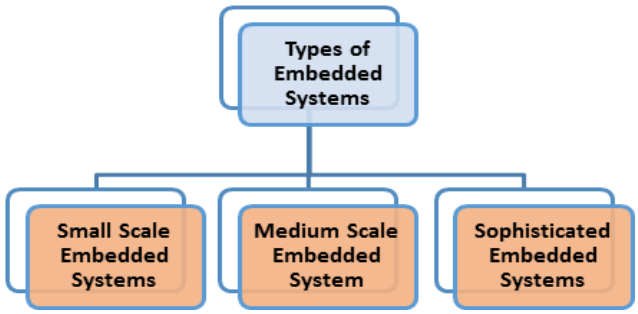


Fig. 7.2. Types of embedded systems

microcontroller. To develop this kind of small embedded system, certain programming tools such as editor, cross-assembler, and assembler (IDE) are important.

Medium-scale embedded systems

The considered types of embedded systems in this case can be developed using 16- or 32-bit microcontrollers. These systems are presumed to have some complexities with regard to hardware and software. Typically, C, C++, Java programming languages as well as source code development tools are used to develop this type of system.

Complex embedded systems

This type of embedded systems can have several software and hardware complexities. IPS, PLA, ASIPS configuration or scaling processor can help in this regard. To develop this system, co-design is required, as well as the hardware and software components to be integrated into the final system.

7.5. Principles of embedded system realisation

The principles of implementing embedded systems can be considered in relation to two levels of system interpretation. Here, a macromodel is distinguished (in the form of a structure from separate parts) and a micromodel (as a process of work and interaction of the parts under consideration).

The formal system in this case is the theory of aggregates. It is responsible for describing the structure of the CPS, CEN, and for defining the dynamic behaviour of the aggregates. The formal CPS model can be defined by the well-known Equation (7.1).

The HLA construction diagram is shown in Fig. 7.3.

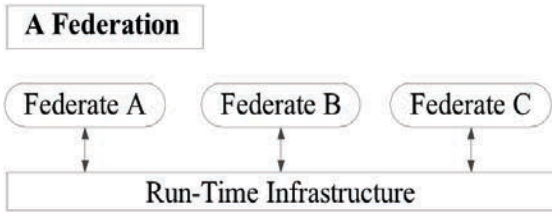


Fig. 7.3. The scheme of HLA formation

Considering it, the following explanations need to be introduced:

- a federation includes controlled physical objects of the cyber-physical system (in other words, aggregates);
- management processes are defined by CEN models (in other words, implementation models);
- RTI is an aggregation mapping required for device communication across IoT components.

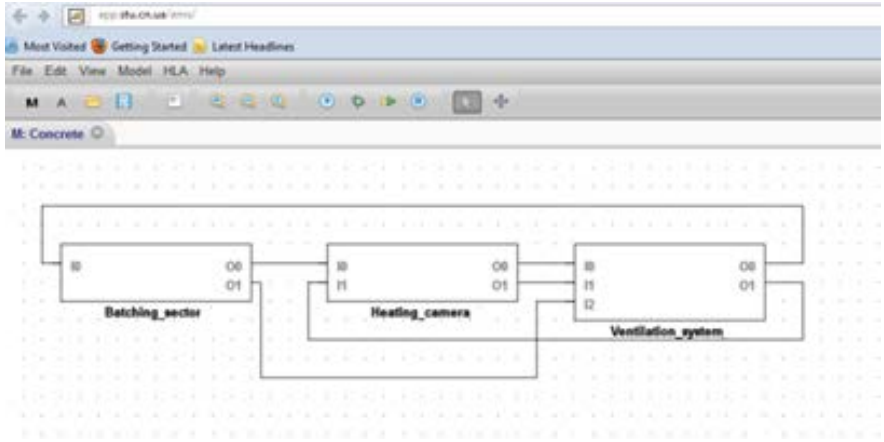


Fig. 7.4. Aggregate formalized representation of a cyber-physical system

Examples of the formalized presentation of the CPS structure and a separate element of this structure are shown in Figs. 7.4 and 7.5, respectively.

The interaction between the entities of the cyber-physical system should take place within the framework of the HLA architecture. For this purpose, units for the implementation of connection functions via RTI, during the execution of various control algorithms should be developed. In this way, data transfer is carried out by transferring XML tags with attributes via the WEB protocol.

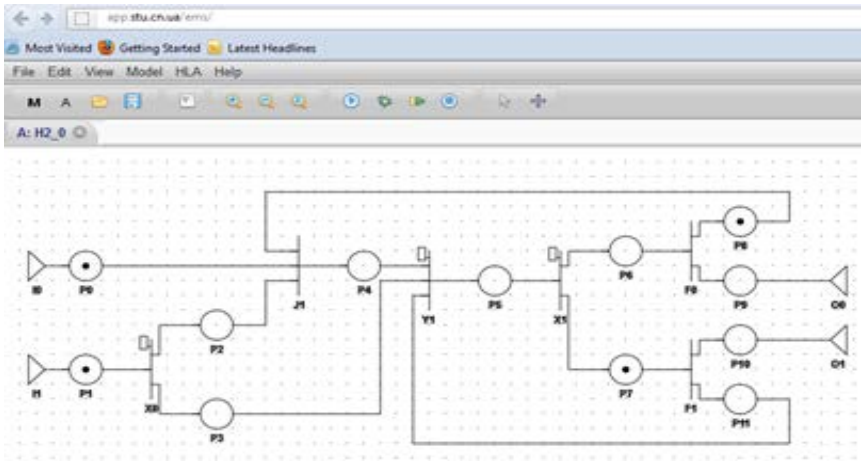


Fig. 7.5. A network management algorithm called control E-networks

The exchange mechanism sounds like "subscription". This means that the aggregate must complete the subscription process using the RTI service. RTI monitors changes in data parameters, so events are always transmitted instantly. In turn, attributes are transmitted when their values change and at a specified time.

The developed CM must be able to accept co-variables and attributes based on the actions of the hierarchical model. After that, the transfer of attributes to the aggregates should be organized through the so-called subscription mechanism. Synchronization of entities in this case is performed in real time for all units included in the cyber-physical system.

Interaction of HLA components with RTI services is possible based on the presented interfaces (RTIInput and RTIOutput) of the CM (Fig. 7.6).

The process is explained by simple conclusions. The I/O interface is subscribed to receive or transfer attributes from aggregates. Thanks to these interfaces, the data exchange mechanism is implemented directly between the units. It is worth clarifying that the internal structure of CM is represented by the CEN model, because CM is, in fact, an aggregate.

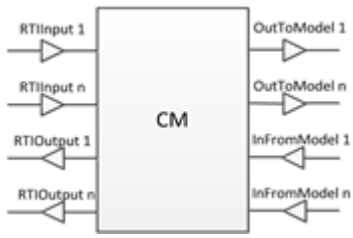


Fig. 7.6. CM Interface

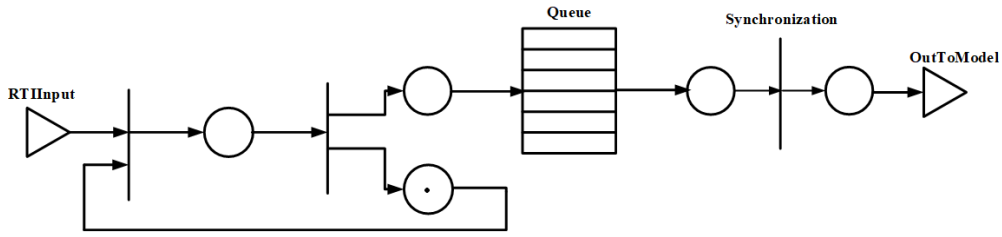


Fig. 7.7. CEN model as a component of the CM module

Recall that only attribute values are transmitted during data exchange. Therefore, it is necessary that the label constantly functions in the structure of the connection module. Attributes from various other federations are entered into the label (Fig. 7.7). The attribute, when RTInput occurs, starts the execution of the method. The process sets a new attribute in the tag's attribute list. After that, the tag with the attribute goes to the priority queue. There it is ordered and passed to the model. It should be noted that the values of the input parameters of the network can be changed if necessary.

CM can generate message queues. This unique feature is required for successful unit synchronization. The formation of message lists is carried out due to inconsistencies in the transmission time of attributes. The queues store the time when the message has been received and the time when the execution started as components. Messages are generated according to well-known HLA standards and criteria. We can say that during the transmission of messages, the timestamp of the values of the label attributes and network variables is transmitted. This is needed to identify when the attributes have been updated. Thus, the synchronization of the units is achieved.

7.6. Implementation models as an embedded control program

The section considers the distributed simulation system – E-net model system (EMS) [139]], which allows the creation of distributed models based on the formal apparatus of hierarchical E-networks within the framework of the HLA architecture.

7.6.1. Architectural features of the EMS system

The developed EMS system has a client-server architecture. In addition to the distributed mode, the system also supports a single-processor simulation mode. This mode can be used for debugging models and in the process of modelling non-resource-intensive tasks.

The system is accessed through a browser. Therefore, on the client side, the user

does not need to install or run any additional software. Using the subsystem of the graphical interface, the user creates a model in the graphical editor of the system, configures the experiment, and starts the simulation process itself. In this mode, the entire modelling process is performed on the web server. Based on the simulation results, it is possible to view the results in the form of tables, graphs, and histograms. The results obtained, as well as the developed models, are stored in the database.

In the distributed modelling mode, the models are executed on the modelling servers. Beforehand, the EMS system must be installed on the modelling servers and the RTI must run on the web server.

As in the uniprocessor simulation mode, the user connects through a browser to the system's graphical interface on the web server and creates models in the graphical editor. To work in distributed mode, first, a federation on a web server must be created and run, then, a federation must be created based on the developed aggregates. During the creation of a federation, the user creates a connector module (CM) for each aggregate and specifies the IP address of the modelling server from the list of available ones on which this federation will be performed. The experiment parameters are configured on the web server using the experiment subsystem.

The communication between the web server and simulation servers is provided by a communication module. Its tasks include sending the configuration of federations created on the web server, receiving simulation results, and managing the work of federations, including starting, stopping, and reloading federates on the simulation servers. The work of federates is controlled by transmitting a sequence of commands (set, get, call). Thus, the simulation server receives all the necessary settings to run the simulation, namely: the federation name, the federation name and the aggregate name from which the federation was created, the configuration of the input/output CM interfaces of this federation, and the parameters for running the experiment.

The accumulation of statistical data takes place on the modelling servers; at the end of the modelling, the results are transmitted to the web server to the statistics collection subsystem, where the final report is generated.

Interaction and synchronisation of federates during distributed modelling takes place using RTI services and the developed CM module. RTI is launched on a web server, federates interact with RTI using specially designed CM interfaces.

7.6.2. EMS core structure

The main object of the kernel is a model that stores the current model time, the end time of the simulation, and a time list of events. An aggregate is used to store the structure of the model. It can contain transitions, positions, queues, inputs, outputs, and nested aggregates inside. The model contains a reference only to the root aggregate; they are the objects of the root aggregate that are taken into account

when the model is run. Thus, in essence, a model is also an aggregate and can be used precisely as an aggregate in other structurally more complex models. The main difference between the model and the aggregate is that the latter cannot be launched to carry out experiments and collect statistical data.

Variables defined by name, type, and value can be defined at both the root and nested aggregate levels. The scope of a variable is determined by the aggregate in which it is created and the aggregates of its children.

The label is determined by the position at which it is set, which allows creation of several labels in the aggregate with a different set of attributes. Similar to a variable, each attribute is characterised by a name, type, and value.

7.6.3. Language for describing models in EMS

The EMS system uses a graphical mode of building models [140], which simplifies the process of their development for subject matter experts, eliminating the need to learn universal programming languages. The model is created on the basis of a specially developed set of components in the graphical editor of the modelling system. However, to define the functions of delay, transformation, as well as the decisive function of transitions of the hierarchical E-network, it is necessary to use a specially developed language.

IEL (IE-net language) is a developed interpreted language that is used in the EMS system and allows the setting of functions on IE transitions, label priority functions in IE queues, supports all basic data types, control structures, basic mathematical functions and comparison operations, and functions for generating random variables according to distribution laws. Let us take a closer look at IEL capabilities.

7.6.4. Language grammar and identifiers

The developed language supports the following data types [141]:

1) INTEGER – integers in the range from –263 to 263–1 (0, 128, –144);

2) REAL – real numbers, the exponent can take values from –2147483648 to 2147483647 (–0.05, 32E+2, 2E–3);

3) BOOL – Boolean data type (TRUE, FALSE);

4) STRING – text data type, the number of characters in a line is not limited (“LINE”, “LINE”);

5) UNDEFINED is a type that represents undefined values, for example, the result of division by 0.

Language identifiers are used to specify variable names and consist of any sequence of letters, numbers, or underscores. A digit cannot be the first character. The names must not match the values of the keywords. To declare variables, use the VAR keyword followed by the variable name. Repeated announcements are prohibited. At the moment of declaration, the variable can be initialised, otherwise it will take an undefined value. For instance:

VAR UNINIT;

VAR INIT = 4.5.

Using EL, the user has the opportunity to organise a full interaction with the model. To refer to aggregates, use the key character A, to the root unit (model) – ROOT, to labels – T, to positions – P, to variables – V.

To access an aggregate variable, reference the required aggregate using the key character A and the aggregate identifier or ROOT to access the root aggregate and specify the variable name. For instance:

VAR X = V [1];

A [8] .V [12] = TRUE.

If no aggregate reference is specified, the current aggregate is accessed. The result of reading a non-existent variable is an undefined value. Enabling/disabling dynamic creation of unit variables during interpretation is controlled by the system and can be set separately for each unit.

To access the attributes of a label, one must refer to a specific position in the required aggregate. For instance:

A [2] .A [nested_agg1] .P [11] .T [145] = 12.5;

P [2] .T ['markAttr1'] = TRUE.

As in the case of the network variable, if a reference to an aggregate is not specified, then the current aggregate is accessed, and the result of reading a non-existent label is an undefined value. Enabling/disabling the dynamic creation of label attributes during interpretation is controlled by the system and can be set separately for each unit or position.

To check for the presence of a label, use the T key character. The type of the calculated value is BOOL. For example: VAR isPlace1Marked = P ['place1']. T.

To address the current label, use the T key character without specifying a position.

For example: VAR attrFromCurrentMark = T [1].

7.6.5. Operations

Mathematical operations are defined for types INTEGER and REAL. If an operand of a different type is used, the result of the operation will be an undefined value (unless the operator used for this type, for example, '+', can be used to concatenate strings). The resulting type is INTEGER or REAL. Allowed operators: '+' – addition; '–' – subtraction; '*' – multiplication; '^' – exponentiation; '-' – sign change; '!' – sign change; '/' – division. Division by zero is an undefined value.

Comparison operations are defined for all data types. Values of different types during comparison are treated as carriers of a type common to both operands (for example, REAL for INTEGER and REAL, STRING for INTEGER and BOOL). An exception is the UNDEFINED type, the operands of this type are comparable only with each other. Comparing an operand of undefined type with an operand of any other type results in an undefined value. The resulting type is BOOL. Allowed operators: >,> =, <, <=, ==,! =.

Logical operations are defined for the BOOL type. If an operand of a different type is used, the result of the operation will be an undefined value. Allowed operators: '&&' – logical AND; '||' – logical OR; '!' – negation; '-' – negation. For instance:

! (2> –2 || 10 <= 100) && 2> –100 // = FALSE

–(2> –2 && 10 <= 100) || 2> –100 // = TRUE

The concatenation operation is defined for any data type. It uses the '+' operator.

7.6.6. Functions (mathematical)

The following functions are defined for the values of type INTEGER and REAL in the range from 4.9E–324 to 1.7976931348623157E308 (hereinafter "range Double"):

- SIN (X) – sine, result – REAL;
- OS (X) – cosine, result – REAL;
- TAN (X) – tangent, result – REAL;
- COT (X) – cotangent, result – REAL;
- ATAN (X) – arctangent, result – REAL;
- LN (X) – natural logarithm, result is REAL. The function is defined for argument values >0;

- SQRT (X) – square root, the resulting type corresponds to the type of the argument, the function is defined for the argument values > = 0.

The following functions are defined for any value of type INTEGER and REAL:

- ABS (X) – module, the absolute value of the argument. The resulting type matches the type of the argument.

- SIGN (X) – 1 for X > 0, 0 for X = 0, – 1 for X < 0

- ENTIER (X) – the largest integer not exceeding X. The computed value type is INTEGER.

Functions for obtaining random variables.

The internal generator of the EL interpreter is responsible for generating sequences of random numbers.

Each time before starting the interpretation, the generator will be initialised with the current time value in ms.

SEED (X) – initialises the random number generator with the X parameter. The parameter type is INTEGER. For example, SEED (60,000).

POISSON (X) – generates integers distributed, according to Poisson's law with parameter X. Type of parameter X is INTEGER or REAL, range Double. The computed value type is INTEGER. The function is defined for argument values > 0. For example, POISSON (2.5).

UNIFORM (A, B) – generates numbers evenly distributed over the interval [A, B] (B > A). Parameter type A, B is INTEGER or REAL, range is Double. The computed value type is INTEGER. The function is defined for values B > A. For example, UNIFORM (2, 10.9).

EXPONENTIAL (X) – generates exponential numbers with the X parameter. The type of the X parameter is INTEGER or REAL, the range is Double. The computed value type is REAL. For example, EXPONENTIAL (10.9).

NORMAL (A, B) – generates numbers distributed according to the normal law with mathematical expectation A and standard deviation B. Type of parameters A, B – INTEGER or REAL, range Double. The computed value type is REAL. The function is defined for values B > 0. For example, NORMAL (2, 10.9).

BINOMIAL (A, B) – generates numbers distributed according to the binomial law with the number of tests A and the probability of success B. Parameter type A is

INTEGER, range from 0 to 231−1 inclusive, B – INTEGER or REAL, range from 0 to 1. The type of the computed value is INTEGER. For example, BINOMIAL (2, 0.4).

7.6.7. Operators

Conditional statements are used to skip or execute certain statements depending on the computed values of the given constructs. Operators can be grouped into blocks using curly braces. Defining constructs are of two types:

- A given logical expression. Example:
  VAR X = 0; IF (2 == 2) X = −92; ELSE X = 92; RETURN X; // = −92
  VAR X = 0; IF (2! = 2) X = −92; ELSE X = 92; RETURN X; // = 92
  IF (2 == 2) {IF (2! = 2) RETURN −1; RETURN −2;} ELSE RETURN −3;
  RETURN −100; // = −2

- Checking for the existence of a value. Example:
  IF ('aa') RETURN 2; ELSE RETURN −2; // = 2
  IF (2/0) RETURN 2; ELSE RETURN −2; // = −2

The return statement from the RETURN function can return the specified value. The TIME operator returns the current simulation time, value type REAL. Operator E returns the value of constant E, value type REAL.

7.6.8. Organisation of the experiment

With the help of the subsystem of the experiment in EMS, a single or multiple run of the model is carried out during a given simulation time.

According to the terminology accepted in the theory of experiments [157], the parameter that changes during the experiment is called the factor, the values of the parameter – the levels of the factor, the obtained values of the investigated quantity corresponding to the levels of the factor – the responses of the system. The number of factor levels is not limited and is set indirectly by determining the initial and final values of the factor, as well as the interval of its change. In a particular case, only one point of the selected parameter can be specified, which corresponds to a single-level experiment.

All parameters of the experiment are set on the server before starting the federations and are sent to each federation participating in the simulation in this federation by the communication module using special commands.

EMS provides for the solution of problems of strategic and tactical planning of the experiment. With regard to the strategic planning of the experiment, this version of the system provides for only one-factor experiments. In this case, the parameters of the model are assumed to be constant, and one of them is changed over the entire

range of values. If necessary, an experiment for each parameter separately can be conducted sequentially . The external factor in relation to the federation can be only that variable on the values of which the input of the federation is subscribed.

Regarding the tactical planning of the experiment, EMS provides two options for carrying out the simulation: first, with a predetermined number of runs to obtain each response point at fixed values of the factor; second, with the determination of the runs required number with the rule of "automatic stop".

The "automatic stop" rule is based on the confidence interval method. In this case, it is assumed that the accuracy of representation d of the mathematical expectation E of e response y and the level of significance a, which guarantees that E falls inside intervals (Y + d, Y − d) with probability P = (1 − a), is assumed. In this case, Y is the mean value calculated over a sample of volume N and is an estimate of E. The confidence interval in which the arithmetic mean of the desired value y obtained as a result of N realisations is located, depends on the selected confidence probability P.

7.6.9. Representation of models in PNML format

Today there are many formats for serialising data. The most popular of these are XML [158] and JSON [159]. It is convenient to store application data in a database for the operation of which the database server does not need to start. An example of such a database is SQLite [145].

JSON is a format for storing and transmitting data. However, JSON libraries used for working with data are underdeveloped, which makes it difficult to use it as a data storage format.

SQLite is a library for working with relational data stored in a separate file. Ease of use and configuration, as well as high speed, make it convenient for the developer. Some of its disadvantages include the inability to read the contents of the file without specialised software as well as the need to use additional libraries.

XML is a common format for storing application data. Many technologies and libraries working with XML, which support different programming languages, make this format convenient for the developer. XML is the language most commonly used in simulation and simulation software to create standardised data formats for basic input and output files, and is also the basis for files that store the simulation model. The use of XML files in simulation software packages provides the advantages of standardised data access, facilitates the addition of new features and plug-ins, and creates conditions for software integration. There are implementations of using XML to link the simulation model to other software and between simulation models. XML is also used in multi-modelling. The term "multi-model" [146] means that the simulation scheme supports:

- plurality – using more than one model;
- heterogeneity – the use of models of different types;
- hierarchy – hierarchy of models;
- customisability – alternative views for the same model type.

There are several systems that support certain aspects of multi-modelling, for example, the Rube system, the OpenSML web system. EMS also supports all four aspects of multi-simulation. The model is a tree, where the root is the root unit and the leaves are its constituent aggregates. Considering all the above advantages of XML, EMS uses this very format for storing data.

Petri Net Markup Language (PNML) is an international standard [147] that defines the format for storing the structure of a Petri net in XML. PNML allows the storage of information of positions, transitions, connections between them, as well as data about the coordinates of each object, its colour, and signature, when displaying a Petri net. However, the most significant advantage of using the PNML standard is the ability to automatically build a program model from its graphical representation.

PNML defines a standard representation of only a small number of Petri net extensions. However, the PNML standard provides two methods for extending it:

1. The ability to create your own PNML attributes and annotations (annotations, unlike PNML attributes, describe elements that affect the image of an object – colour, shape, etc.). PNML attributes are represented as XML elements. In the PNML schema, this extension is represented as follows:
<define name="arc.content">
<interleave>
…
<ref name="arc.labels"/>
…
</interleave>
</element>
</define>
<define name="arc.labels">
…
<empty/>
</define>
To define PNML attributes, create a new XML file with the following content:
<?xml version="1.0" encoding="UTF-8"?>
<grammar ns="http://www.pnml.org/version-2020/grammar/pnml"

```
        xmlns="http://relaxng.org/ns/structure/1.0"
        xmlns:a="http://relaxng.org/ns/compatibility/annotations/1.0">
        <define name="arc.labels" combine="interleave">
        <optional><ref name="PTArcAnnotation"/></optional>
        </define>
</grammar>
```

2. Inheritance of the ToolInfo class of the PNML object model. This method can be applied if it is necessary to add new data for an existing type of Petri net with the aim of their further application in certain programs.

To work with hierarchical E-network models, the international PNML standard was expanded by adding new PNML attributes to transition and place objects by considering certain features of E-networks such as:

- the presence of transitions and queues of different types;
- the presence of nested units;
- the ability to set functions at each network transition;
- availability of various input/output interfaces;
- the presence of a position label by the attribute.

This approach makes it possible to automate the process of constructing software models based on their XML descriptions, as well as to set standard graphic data defined by the standard for all objects of the hierarchical E-network.

Different models may include aggregates of the same structure but with different parameters. The EMS introduced the concept of the type of unit as a set of all units with a single structure. This approach makes it possible to reuse the developed aggregate in different models but at the same time to save the description of its structure only once. If necessary, it is possible to change all units of the same type in different models if during the development of the model the user has made adjustments to its structure.

This is the definition of Tx transition in the XML file:

```
<transition id="ValidT">
        <name>
        <text>XTest</text>
        <graphics>
                <offset x="22" y="-14"/>
        </graphics>
        </name>
```

```
        <definition type="transition" subType="T">
        <delayFunction type="el"><![CDATA[
                RETURN 123;]]>
        </delayFunction>
        <transformationFunction type="el"><![CDATA[
                T['attr'] = 123;]]>
        </transformationFunction>
        <permitingFunction type="el"><![CDATA[
                T['attr'] = 123;]]>
</permitingFunction>
        </definition>
        <graphics>
                <position x="950" y="484"/>
        </graphics>
        </transition>
```

An example of the description of the initial position marking:

```
<place id="validPlace">
        <name>
        <text>validPlace</text>
        <graphics>
        <offset x="22" y="-10"/>
        </graphics>
        </name>
        <token>
        <attribute name="val1" value="12.3" />
        <attribute name="val2" value="14.3" />
        </token>
        <graphics>
        <position x="334" y="404"/>
        </graphics>
        </place>
```

An example of describing the definition of variables for each type of unit:

```
<?xml version="1.0" encoding="UTF-8"?>
<pnml xmlns="http://www.pnml.org/version-2020/grammar/pnml">
        <net                                                    id="rootAgregate"
type="http://www.cs.stu.cn.ua/jess/enets">
            <variables>
                <variable name="var1" value="-12.34" />
                <variable name="var2" value="+12.34" />
            </variables>
        …
        </net>
```

```
</pnml>
```

The file with the description of the model must define the type of the root aggregate as well as the difference between the aggregates used in the model and their definitions. The following is an example file describing the model:

```
<?xml version=”1.0” encoding=”UTF-8”?>
<model xmlns=”http://www.cs.stu.cn.ua/jess/enetsdefinitions”>
        <rootAggregate type=”Root”>
        <aggregate name=”A1”>
            <transition name=”T1”>
                <transformationFunction type=”el”>
                    RETURN 9 + 1;
                </transformationFunction>
            </transition>
            <aggregate name=”child”>
                <place name=”P1”>
                    <initialMarking>
                        <attribute name=”attr” value=”12.34” />
                    </initialMarking>
                </place>
            </aggregate>
        </aggregate>
        <aggregate name=”A2”>
            <variables>
                <variable name=”varName” value=”34.56” />
            </variables>
            <transition name=”T1”>
                <transformationFunction type=”el”>
                        RETURN rand;
                </transformationFunction>
            </transition>
                 <queue name=”queueName”>
                <priorityFunction>
                        RETURN 1;
                </priorityFunction>
            </queue>
        </aggregate>
        </rootAggregate>
</model>
```

The structure of the XML file can be checked by means of the RNG schema; It is imperative to separately make sure that there are no following semantic errors:

- **violation of the rules of the E-network;**

- **cyclical definition of the unit;**
- **absence of an element with the desired name.**

As these errors cannot be validated using the XML schema, they must be validated when the model is loaded.

7.7. Verification and testing of embedded models

Let us have a closer look at an example of determination of ventilation parameters in the production of aerated concrete.

The aerated concrete production process is a complex multi-stage technological process accompanied by an increased release of hydrogen at the initial stages of production. Therefore, the task of ensuring the proper functioning of the ventilation system comes to the fore among the safety issues of this production.

Increased hydrogen evolution is observed at two stages of production: at the pouring site and in the holding chamber. Let us consider them in more detail. At the pouring site, the starting materials: sand, lime, cement, gypsum and water are weighed and fed into the mixer, where they are mixed for 3 minutes. After that, aluminium suspension is added to the prepared mixture, which provokes the reaction of aluminium with calcium hydroxide, and as a result hydrogen begins to evolve. The chemical reaction of this process is:

$$3Ca(OH)_2 + 2Al + 6H_2O = 3CaO \times Al_2O_3 \times 6H_2O + 3H_2 \quad (7.3)$$

Mixing is finished, the mixer outlet is opened and the finished mixture flows through the pouring pipe into the mould, which is located on a mobile platform. After filling the mould, the vibration frame is immersed in the mixture and by means of vibration, removes large air inclusions from the mixture. In total, there are 2 mixers in the pouring area, each filling in 6 forms of the mixture. Each mould is filled with 3m3 of the mixture, considering the fact that during the preparation of aerated concrete, the volume of the initial mixture increases.

The next stage is a holding chamber in which a constant temperature of 650 °C is maintained and the initial mixture is heated. A chamber with a volume of 2079.88 m^3 is located in a general workshop with a volume of 84480 m^3, and contains 8 tracks, where on each 5 forms can be located simultaneously. Thus, 40 moulds can be simultaneously in the chamber. When heated, the intensity of the reaction increases, which provokes an increased release of hydrogen into the atmosphere of the chamber.

The results of calculations, carried out on the basis of the chemical reaction (7.3) and data on the consumption of raw materials per 1 m^3, show that one form releases an average of 0.622 m^3 for the entire process of aerated concrete production. The maximum permissible concentration of hydrogen in air is 0.8 %. Whereas one form

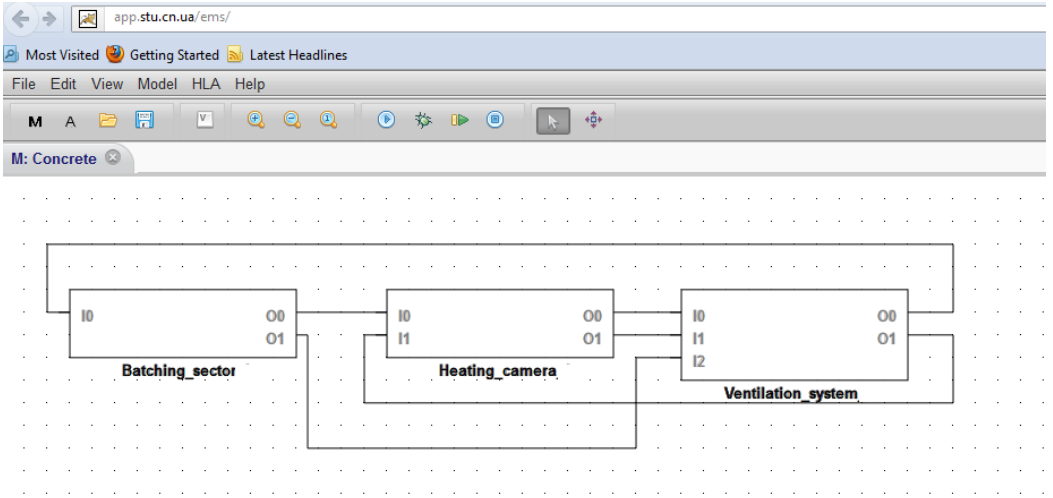


Fig. 7.8. Model of the technological process for the preparation of aerated concrete

during the entire technological process releases a volume of hydrogen equal to 0.1 % concentration. Therefore, it is extremely important to determine the level of hydrogen concentration released by all forms at each time section of the technological process.

Mathematical calculations do not allow tracing the dynamics of the increase in the concentration of hydrogen in the air, while it is precisely such information that will facilitate the effective use of the ventilation system in the production process. Therefore, it is better to solve the problem with the help of simulation models.

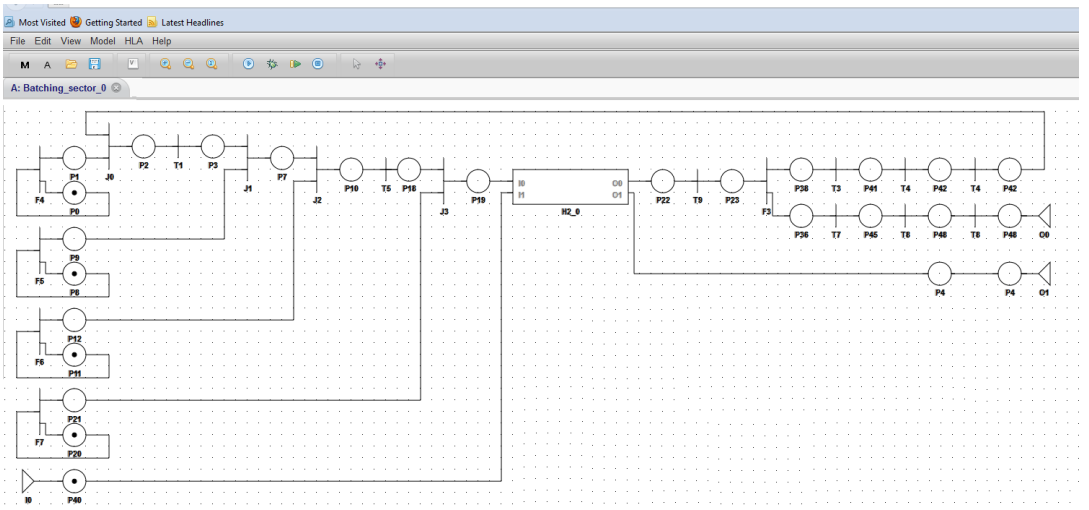


Fig. 7.9. Aggregate that simulates the pouring area

The simulation model developed in EMS (Fig. 7.8) consists of three units. The

“Batching_sector” unit simulates the technological processes in the filling area, “Heating_camera” – in the heating chamber, and “Ventilation_system” simulates the operation of the ventilation system. Let us consider in more detail the operation of each unit.

The transition functions of the “Batching_sector” aggregate (Fig. 7.9) have a definition that is given further:

Transitions F4, F5, F6, F7 simulate the arrival and subsequent weighing of raw materials.

Transition T1 simulates the process of filling a mixer with sand, gypsum and water and mixing them. Transition delay function: RETURN UNIFORM (25.35).

Transition J1 determines the delay time required to add lime to the prepared mixture. Transition delay function: RETURN UNIFORM (15,25); // adding lime .

Transition T5 simulates the main mix. Transition delay function: RETURN UNIFORM (140,150); // main batch .

Transitions J3 simulates the process of adding aluminium and mixing the mixture. At the transition, the reaction start time is recorded: V [‘Tstart’] = TIME.

The presence of the label in position P22 semantically means a mixture ready for pouring into moulds in which hydrogen continues to evolve during the reaction of aluminium with calcium hydroxide.

Transition T7 simulates the dwell time required for the vibrating frame to operate. Transition delay function: RETURN UNIFORM (35.45).

Transition T8 simulates the delay time required to fill the form. Transition delay function: RETURN UNIFORM (230,240).

Transition T9 simulates the dwell time required to transport the mould to the heating chamber. Transition delay function: RETURN UNIFORM (85.95).

Transitions T3, T4, T6 simulate the delay time required to unload the mixer, prepare for rinsing and rinsing. Delay function on transitions: RETURN UNIFORM (15.25).

The nested unit H2 (Fig. 7.10) simulates a continuous process of increasing hydrogen concentration in the air, and operates in parallel with other processes of

the “Batching_sector” unit and interacts with the “Ventilation_system” unit, which simulates the ventilation system, as well as the “Heating_camera” unit, which simulates the chamber heating. The process of adding aluminium begins when three conditions are met: the presence of a free form, the presence of a prepared mixture, and the arrival of permit signal from the ventilation system. The aluminium addition process is interrupted when an inhibit signal is received from the ventilation system.

With the help of tag attributes, data on the hydrogen concentration at each time point are transmitted to the ventilation system.

The transition functions of the “Heating_camera” aggregate (Fig. 7.11) have the definition given further:

Transition T0 simulates the opening time of the heating chamber shutter. Transition delay function: RETURN UNIFORM (10.15).

The priority queue QFP0 simulates the set of shapes emerging from the fill area, which must be distributed along the camera paths.

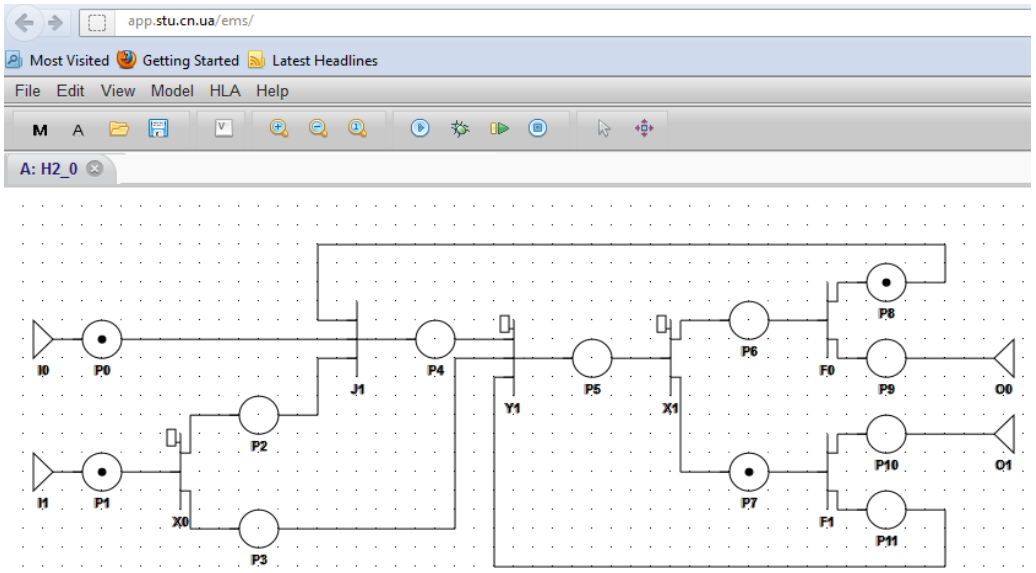


Fig. 7.10. Nested unit simulating the process of hydrogen evolution in the filling sector

X2 transition distributes shapes to paths based on line filling. The paths in the chamber are filled in sequence:
VAR isPlace4Marked = P [‘P4’]. T;
VAR isPlace5Marked = P [‘P5’]. T;
VAR isPlace6Marked = P [‘P6’]. T;
VAR isPlace7Marked = P [‘P7’]. T;

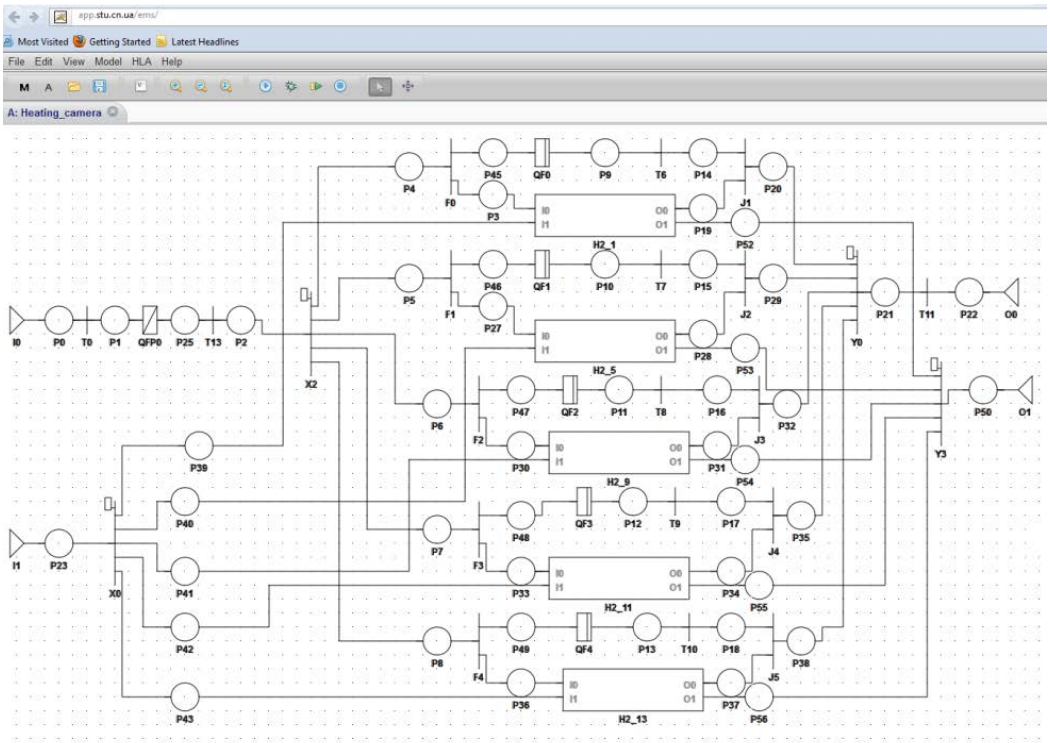


Fig. 7.11. Unit that simulates the heating chamber

VAR isPlace8Marked = P ['P8']. T;
IF (isPlace4Marked == TRUE) RETURN 0;
IF (isPlace5Marked == TRUE) RETURN 1;
IF (isPlace6Marked == TRUE) RETURN 2;
IF (isPlace7Marked == TRUE) RETURN 1;
IF (isPlace8Marked == TRUE) RETURN 2;

Transitions T6, T7, T8, T9, T10 simulate the heating time of the mould in the chamber in parallel on each of the paths. Transition delay function: RETURN UNIFORM (3550.3600);

During the heating process, the evolution of hydrogen begins to increase. It is the continuous process of hydrogen evolution that simulates the H2 unit embedded in each branch of the main unit path. Unit H2, in turn, interacts with the ventilation system, signals from which come through input I1 and are distributed at transition X1. Thus, in the heating chamber it is also possible to plot the dependence of the change in the hydrogen concentration on the heating time.

The structure of the nested unit H2, which simulates a continuous process of increasing hydrogen concentration in the heating chamber, is similar to the unit H2

of the filling sector. The difference is that this process cannot be stopped, the heating time is 60 minutes and the hydrogen concentration is calculated in 1-minute steps.

The transition functions of the "Ventilation_system" aggregate (Fig. 7.12) have a definition as shown further:

Transition X1 receives data on the level of hydrogen concentration from the filling sector and determines whether to activate the ventilation process.

Transition F1 (initiates the ventilation process).

Transition T4:

activation function

{RP1 == 1 & RP2 == 0 & RP3 == 1;} – the transition is activated when the concentration level exceeds the maximum allowable norm;

conversion function

{VE1 = 1; VE4 = 1} – valves are opened through which gas is pumped out of the chamber.

Transition T5:

conversion function

{RP1 == 1 & RP2 == 0 & RP3 == 1;} – the pumps are starting up.

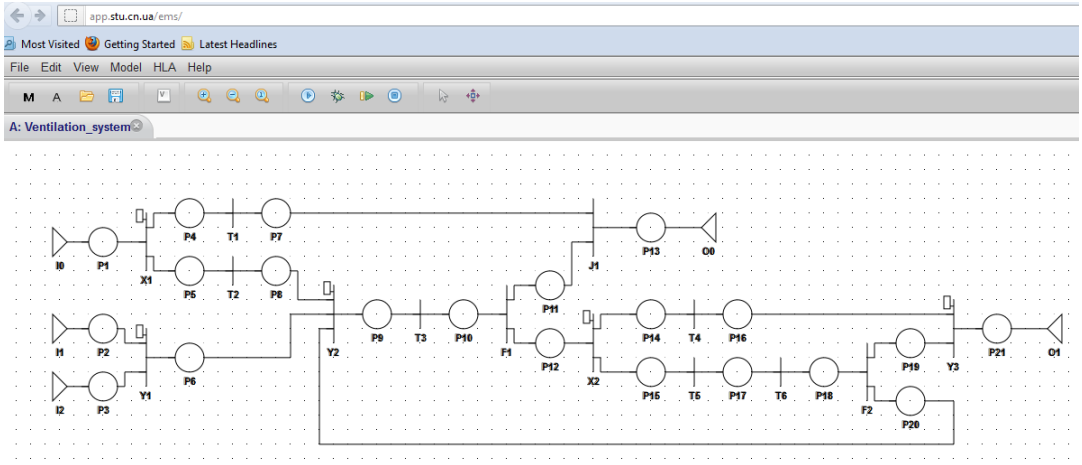


Fig. 7.12. The unit simulating the ventilation system

Transition T6:

activation function

{RP1 == 1 & RP2 == 0 & RP3 == 1 & VE1 == 1;} – the transition is activated when all pumps are running and the valve for pumping hydrogen is open;

conversion function

{VE2 = 1; VE3 = 1;} – air pumping starts.

All aggregates, including nested ones, run in a distributed environment based on the HLA architecture. The simulation results show that when one form of the mixture is heated, the rate of hydrogen evolution is characterised by a burst at the 7th minute and reaching a maximum at the 16th minute.

Thus, the developed models make it possible to determine the ventilation parameters required to prevent an emergency in production and to issue recommendations on the modes of its operation.

7.8. Hardware realisation of embedded models in IoT

7.8.1. Microcontrollers & microprocessors

The Internet of Things allows our world to become interconnected. Microcontrollers help the Internet of Things in many ways. Typically, microcontrollers are tiny stand-alone computers that sit on a microchip, allowing multiple components to be plugged into them for global control. Microcontrollers are designed to perform specific functions and integrate into industrial equipment, household appliances, etc.

Intelligent IoT devices are likely to become the building blocks for a new global economy based on data and services in the near future. Today, workstation solutions are already at the heart of this evolution, allowing various industries to bring high-quality and secure IoT devices to international market quickly and at minimal cost. Advanced RISC machine's (ARM's) powerful suite of IoT solutions offers proven and secure hardware, software, and services designed to help organisations take full advantage of the IoT transformation of their business models, such as [148]:

- Comprehensive coverage from chip to cloud. ARM helps build any IoT system at any scale and with seamless, secure connectivity, whether a particular system is connected to the local cloud or fully connected to the cloud.

- Extensive partner ecosystem. ARM is at the epicentre of the world's largest computing ecosystem comprising a vast community of software, tool and service providers that support the spread of IoT technologies to markets,

industries and applications.

- IoT device security. With the support of its ecosystem, ARM offers a full range of security solutions to help mitigate all types of device attacks. Regardless of connectivity requirements and time-to-market constraints, at ARM, security is never optional and remains a key consideration from the earliest design stages to delivery.

7.8.2. Advanced RISC machine (ARM)

ARM (Advanced RISC machine) is essentially a reduced instruction set 32-bit computer microcontroller (RISC). It was introduced by the computer organization Akron in 1987. ARM is a family of microcontrollers with different architecture versions, each of which stands out with its own characteristics, advantages and disadvantages.

The implementation of the ARM Cortex microcontroller can be seen in Fig. 7.13 [149].

ARM9

ARM9 chips can reach 400MHz clock speed. Although these microcircuits seem a little outdated, they are still in demand on the market. A set of simple instructions for such a chip makes it easy to run many Java applications, which simplifies the operation and implementation of many systems such as wireless routers.

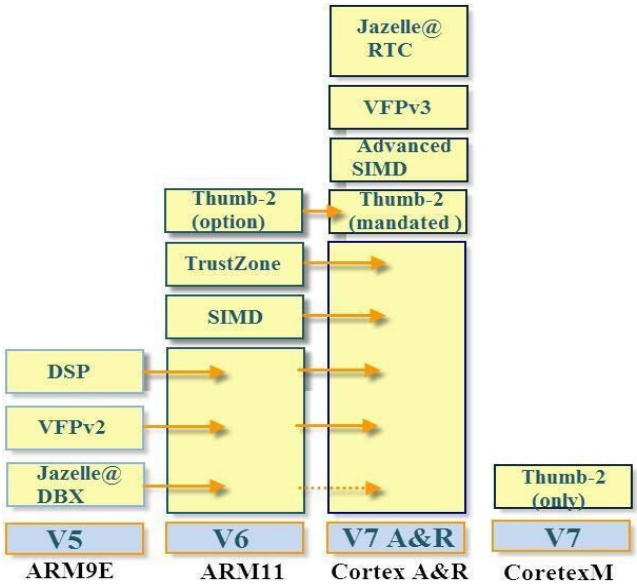


Fig. 7.13. ARM Cortex microcontroller implementation

ARM11

ARM11 processors have a more complete set of simple and understandable instructions expanding their functionality and high clock speeds (up to 1 GHz). Due to their low power consumption and low cost, ARM11 chips can be used in different devices, such as certain types of smartphones.

ARMv7

ARM architecture chips belong to the ARMv7 family. This means that the quantitative characteristics of eight cores and a clock frequency of more than 2 GHz have already been achieved. The processor cores belong to the Cortex line and are used by most SoCs without any changes.

ARM Cortex-A8

Cortex-A8 can be considered the first processor core of the ARMv7 family. It is at the heart of the well-known SoCs of its time (for example, Samsung Hummingbird and Apple A4). This microcircuit demonstrates about twice the performance compared to the previous ARM. However, its higher power consumption makes this chip less popular now.

ARM Cortex-A9

Following the structure changes, ARM Limited has introduced a new generation of the well-known Cortex-A9 chips. The performance of Cortex-A9 cores has nearly tripled over the Cortex-A8. In this respect, it is worth noting that it is possible to combine them by two or even four on one chip.

ARM Cortex-A5 & Cortex-A7

The development of the Cortex-A5 and Cortex-A7 processor cores has a clear goal. In this regard, ARM Limited aimed to reach a compromise between the acceptable Cortex-A8 performance and to minimize power consumption. It is worth noting the likelihood of combining two or four cores in practice.

ARM Cortex-A15

The Cortex-A15 processor cores can be considered a continuation of the Cortex-A9. Thus, chips with ARM architecture were able to roughly equal the speed of Intel Atom. This is an indicator of success and high productivity. There is an underlying reason that Canonical has included a dual-core ARM Cortex-A15 processor in the system requirements for a full multi-tasking version of Ubuntu Touch [150].

The pace of development of microprocessor devices has recently slowed down somewhat and is no longer subject to Moore's law [151]. The reason for this phenomenon is that the limit on the energy density has been reached, which does not facilitate any increase in the clock frequency above 1.5 ... 2 GHz, since the faster the processor runs, the more it heats up.

The advent of multi-core processors eased the situation, however, with the increase in the number of cores, other problems emerged. In particular, the question of the optimal use of resources in parallel computing has become more acute, and there is also a need to develop additional tools for automatically dividing tasks into parallel threads.

Microprocessors are convenient to use in applications where a large set of interfaces is not required. In addition, they are effective in performing floating point calculations. With the increase in the number of cores, the question of dividing the task into parallel becomes more acute. The more cores, the more difficult it is to perform dynamic parallel computing, because it is not always possible to efficiently distribute tasks between cores.

7.8.3. Field-programmable gate array (FPGA)

The logical capacity and performance of FPGAs has grown substantially due to a number of factors: an increase in the degree of integration on the chip, the appearance of faster serial interfaces and communication protocols, a higher technological level, and the use of specialised computing cores and advanced logic circuits.

With the development of FPGA technologies, they are widely used in embedded computing systems for military and aerospace purposes, which are characterised by severe restrictions on power consumption, size and weight of elements. FPGAs are well suited for devices such as radar systems, electronic intelligence systems, image processing systems, and signal processing devices, i.e., they are intended primarily for those devices in which signal processing and vector or matrix calculations are performed. In such applications, the main criterion is not the cost, but the characteristics of the device, especially speed.

Due to the ability to perform cumbersome calculations, FPGAs have become widely used in complex applications. They can perform tens of thousands of calculations in one clock cycle, operating at relatively low clock frequencies of the order of hundreds of MHz, while consuming much less energy than microprocessors with the same performance. If we recalculate the performance per watt of power consumption, it turns out that the matrices are superior to microprocessors by about 50–100 times.

Matrices are not suitable for any project. Their key weaknesses lie in the ability to perform only periodic tasks, the inability to perform floating point calculations, and the complexity of developing program code.

Firstly, FPGAs are not suitable for all algorithms. Matrix resources are fully used when performing the tasks that are easily divided into parallel or repeating subtasks.

Secondly, in the case of working with floating point numbers, it is theoretically possible to implement such applications on FPGAs; however, this will require unreasonably many logic elements, which ultimately will limit the computational density of the matrix, negating all its advantages.

Thirdly, the FPGA-based device designs is much more complicated than a microprocessor-based design. This is due to the underdevelopment of code writing tools. As a result, writing applications for matrices requires more time and developer skills than creating similar applications for the processor. Development costs may also be too high. In addition, microprocessor programming uses a higher level of abstraction than FPGA programming.

Other less important factors influence the choice of a hardware platform for embedded models project implementation, such as the interfaces primarily used. If the designed device needs to support non-standard or outdated interfaces that are not provided in modern processors, then it is better to use FPGA. Matrices allow developers to configure a wide range of standard or specialised serial and parallel (bus) interfaces [152].

7.8.4. Hybrid platforms

The advantages of FPGAs and microprocessors precisely empower developers to the highest level. Therefore, in the world of embedded models, hybrid platforms are increasingly used, containing both types of devices. This solution is notable for the combination of FPGA resources in conjunction with microprocessor performance, which provide support for multi-threaded applications and additional flexibility. IoT support also plays an important role.

Hybrid platforms are best suited for systems that require middleware and a complete protocol stack, as well as synchronisation circuits.

Another solution to this may be to embed the computing core in the matrix. The largest manufacturers of FPGAs (Xilinx, Altera or Lattice Semiconductor) offer arrays containing synthesised logic elements or hardware processor units. The processor architecture can be any: standard, custom or in-house development. On hybrid platforms, it is better to build machines and devices that do not require high speed, for example, to connect to an external processor or to provide a user interface. In addition, hybrid circuits can be used as an alternative to microcontroller devices.

FPGA processor cores are different from classic microprocessors [153]. Due to the peculiarities of the matrices, not all possible nuclear structures can be realised.

Unlike microprocessors, matrices do not facilitate the use of circuitry methods to improve signal transmission quality. As time has shown, matrices with integrated processor units are most widely used in devices such as system controllers, secondary processors in multiprocessor clusters, signal processors and automatic machines.

This approach is beneficial for several reasons. Firstly, it allows one to save on licenses. Since the processor core is located inside the matrix, in fact, when buying a license to use FPGA, a processor license is also acquired. When using an external processor, it would have to be licensed separately. Secondly, it is sometimes impossible to connect an external processor to the FPGA. In this case, the developers resort to software implementation, however, then there are many unused elements on the matrix.

One of the attractive features of using cores in FPGAs is the ability to implement hardware accelerators. They can be synthesised on RTL blocks or can be defined as a separate function. In the latter case, the decoding and executive logic circuits automatically make the changes necessary to embed this new function in the FPGA pipeline. This approach is used only in cases when one operation needs to be replaced.

Matrices with integrated computing cores have several disadvantages, such as the need to use additional schemes and the complexity of project verification. For the computational core to work, additional circuits are needed – memory modules, peripherals, a graphical user interface that helps to connect all the blocks of the system, create an address structure, etc. When connecting RAM to the kernel, difficulties may arise; it all depends on what type of memory access the kernel uses.

The process of verifying an FPGA with an integrated processor core is more complicated than the one used for a conventional matrix, and it is carried out in several stages: verification of the kernel, verification of the microcomputer as a subsystem, and debugging of program code.

Tools offered by the manufacturer generate the correct RTL models, but no manufacturer can guarantee the correct interconnection of the modules. Many projects can no longer be used when adding a processor core, bus, and peripherals. This is because the system becomes too complex to be verified by a simple run. More advanced tools, high-speed models, and methods that analyse the system at the VLSI level are needed. It is necessary to carry out not only the modelling of the operation of all functional blocks, but also the verification of third-party IP blocks.

The situation is slightly simpler if the manufacturer provides debugged and tested code examples, as Lattice does. Software debugging in most cases is carried out either by using internal cores, such as CoreSight in ARM processors,

or by using JTAG tools that allow developers to track step by step the work and set breakpoints.

Let us consider Xilinx Zynq-7000 [154] as the optimal architectural solution for a hybrid platform for building and developing embedded models (Fig. 7.14). At present, Xilinx Zynq-7000 SoC is Xilinx's most advanced development in the field of crystal systems. It is implemented on 28 nm technology and contains a 2-core ARM Cortex-A9 processor, programmable logic, as well as a large selection of peripherals. The combination of function blocks provides a high level of programmability, flexibility and performance when working on a single crystal. This family is positioned as a new subclass of FPGA – Extensible Processor Platform (EPP).

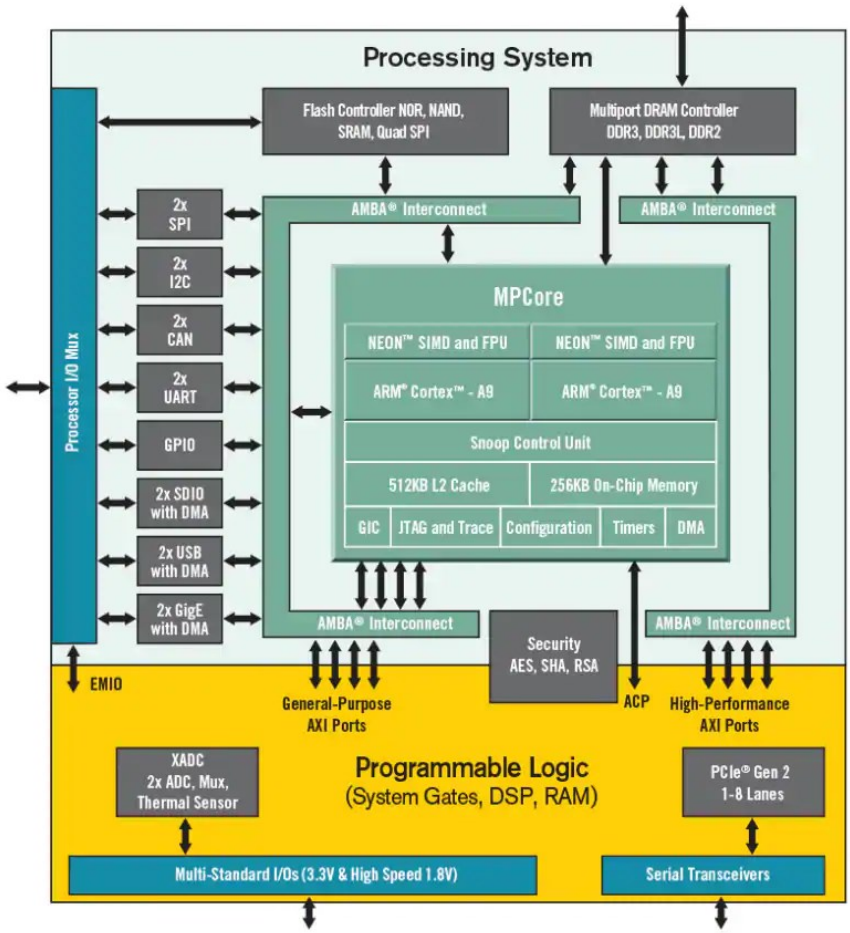


Fig. 7.14. Xilinx Zynq-7000 architecture

The use of the AMBA AXI (Advanced Extensible Interface) broadband interface for data transmission is also considered one of the key architectural decisions. This solution was developed by Xilinx together with ARM to develop an ARM architecture

that would best fit FPGA applications. There are two versions of the AXI4 bus:

- **version with distribution;**
- **streaming version.**

The second key aspect of the architecture is that Xilinx has implemented a large number of standard interface IP blocks in the Zynq-7000 crystal. Zynq-7000 contains up to 512 KB of Layer 2 cache used by both processors. Zynq-7000 EPP has 256 KB of temporary memory, which is used by both the processor and FPGA.

The Zynq-7000 SoC platform has an ARM Cortex-A9 processor with memory and a number of peripheral interfaces. Based on the classification of ARM processor cores, it follows that Cortex-A9 kernels are of the application type. They support high performance in tasks related to the organisation of the user interface and the display of information, as well as support for developed interfaces and storage devices. The great advantage of this type of cores is their rather low power consumption compared to x86-based processors.

The processor complex (meaning both ARM cores with corresponding buses) interacts with the memory of the crystalline periphery without involving the resources of programmable logic, which acts as a hardware platform for the implementation of additional peripherals and computational accelerators, both standard and special, different and special project.

7.8.5. Hardware description language (VHDL)

Hardware description languages began to appear in the early 60s, mainly for reasons and requirements such as working with projects at an abstract formal level (without dropping to the level of small details and physical implementation), the rapid increase in the complexity of digital systems, and the formation of new development requirements for computer-aided design (CAD) systems. The VHDL language appeared, which began to spread very quickly, because it is excellent for the design of embedded models.

Currently, the VHDL language is designed for describing projects of varying degrees of various levels of complexity, from the simplest parts to entire systems consisting of hardware and software parts. It facilitates the building of models at various levels of abstraction, performs simulation and generates time diagrams, maintains strict documentation of the project, synthesise the structure by behavioural description, verifies the project with formal methods, and automatically generate tests.

VHDL [155] inherits many of the properties of high-level programming languages. In addition, the concepts of model time, signal, event, component, and others have

been introduced into the language. It is noteworthy that the use of VHDL does not facilitate the connection of the development projects of embedded models in advance to a specific physical implementation method, because the same logical VHDL implementation is a source of generation of different physical ones.

The VHDL language is based on the following construction principles:

- support for functional decomposition (functional hierarchy and recursion);
- support for structural decomposition (structural hierarchy);
- presentation of the system in the form of parallel functioning interacting processes;
- use of abstract data types;
- use of event-driven modelling;
- support of various levels of abstraction and detail of the project presentation.

Embedded model systems designed to perform specific transformations should generally:

- get some input from their environment;
- perform transformations;
- get some output.

From the analysis of the above, it follows that embedded systems must have connections with their environment, which are called an interface. The system interface is usually described in VHDL by its Entity, which is the main project unit for any system. Transformations are performed by the internal part of the system or by the body (Body), which is called the architecture (Architecture). To obtain additional system capabilities, packages (Package) and libraries (Library) are used.

The recommended structure of the project, which can be a project of a system of embedded models, considering the applied main parallel and sequential proposals and their location, is shown in Fig. 7.15.

Sequential statements are executed in the order that they appear in VHDL code. The order of execution of parallel clauses is not related to the order of their appearance within the architectural body. Parallel sentences are activated by signals that are used to link parallel sentences.

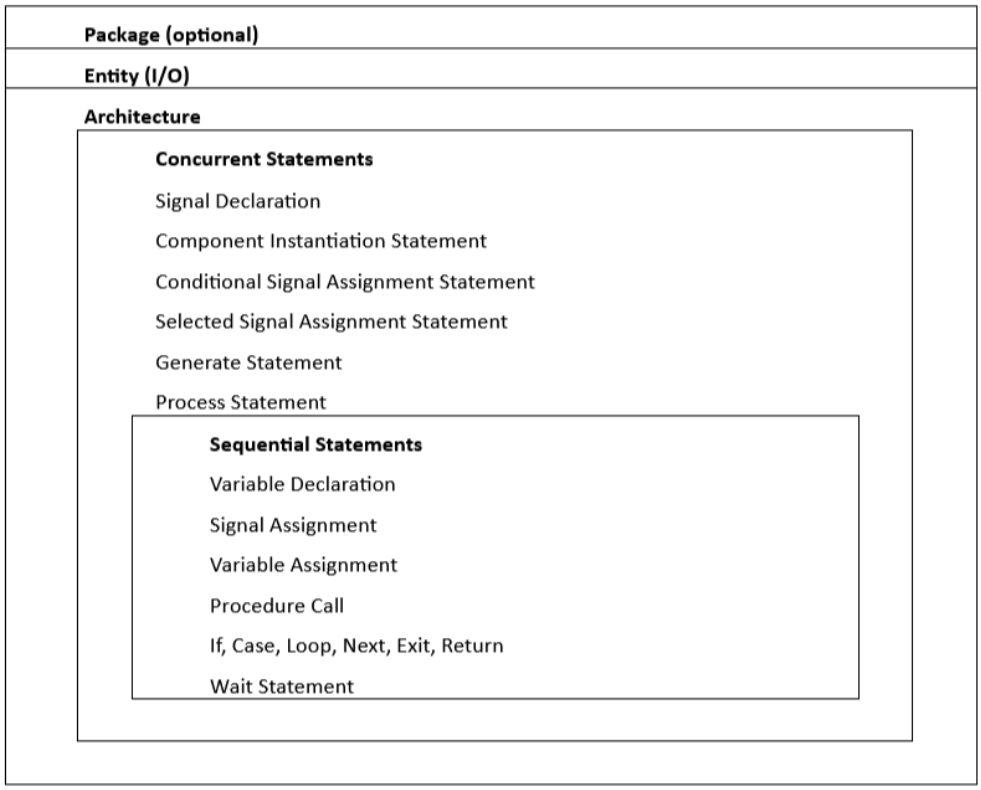


Fig. 7.15. General structure of the VHDL-description of the embedded model system

7.8.6. Example of embedded model realisation

When developing embedded models, as well as projects related to ARM and FPGA, certain rules and engineering practices must be followed. However, with the rapid development of embedded devices, embedded systems are faced with increasingly complex use cases and new challenges.

As an example, consider an embedded system model that implements CEN and consists of a set of aggregated nodes connected to an RTI bus as shown in Fig. 7.16. It is noteworthy that the architecture is implemented in the reverse hierarchy FPGA>ARM>RTI instead of the recognizable ARM>FPGA>BUS.

FPGA implements a transition of a certain type as an object consisting of a transition and associated with its positions. ARM architecture performs the function of communication between transitions and RTIs. In this case, all the nodes of the presented architecture are the same in their composition.

Let us consider a special case of transitions implementation in the VHDL language.

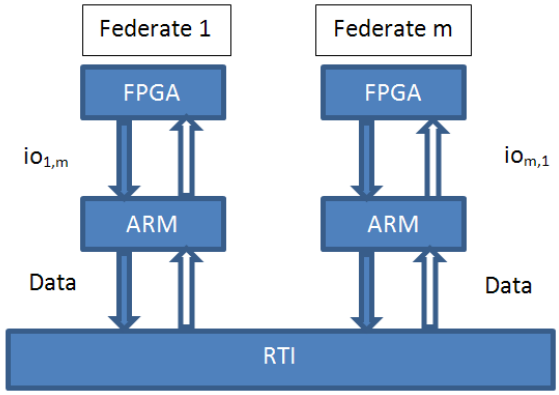


Fig. 7.16. System architecture

Figure 7.17 shows the various formations of the E-net transitions.

Figure 7.17 shows 5 variants of transitions, 5 disparate cases (3 simplest T-transitions), 1 J-transition and 1 Y-transition). Various transition explanations have been defined (user-defined). E-network transitions have their own CONPAR

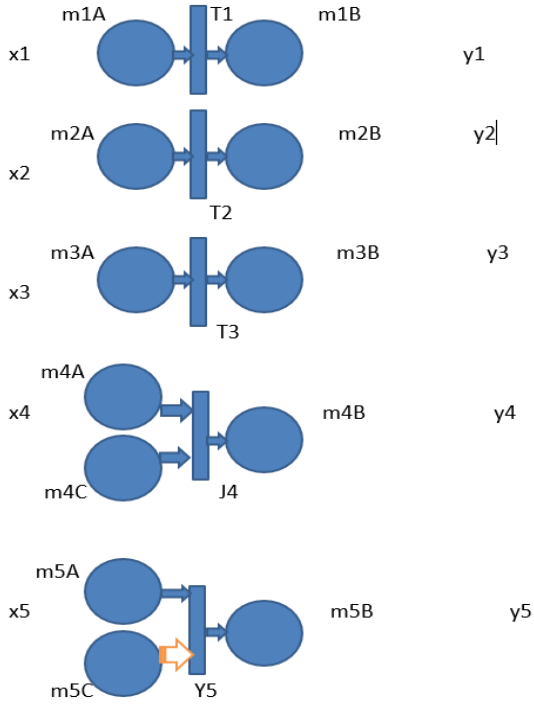


Fig. 7.17. Transitions case

representation to facilitate further conversion to VHDL code.

T1: m1A * x1 | - m1B
T2: m2A *! X2 | - m2B
T3: m3A * (x1 +! X2) | - m3B
J4: m4A * m4C | - m4B
Y5: m5A * (! X2 * x3) * m5C | - m5B

where x3 = x1 +! x2, and x5 =! x2 * x3

For transitions 3–5, the so-called predicates were used because the input signals are not single.

- Symbols – CONPAR syntax to describe transitions:
- T, J, Y (transitions) – transitions
- m (positions) – positions, places
- x (inputs) – input signals
- y (outputs) – output signals
- PreConditions | – PostConditions
- PreConditions – prerequisites that include input positions necessarily and optionally custom formulas/expressions or predicates and/or logical expressions
- PostConditions are postconditions that include initial positions that are activated only after the transitions are triggered
- * (logical AND) – logical AND
- + (logical OR) – logical OR
- ! (logical NOT) – logical NOT

CONPAR description:

```
.clock CLOCK
.input x1 x2 x3 x4 x5
.output y1 y2 y3 y4 y5
.part controller
.places m1 m2 m3 m4 m5
.transitions T1 T2 T3 J4 Y5
```

```
.net
T1: m1 * x1 | - y1
T2: m2 *! X2 | - y2
T3: m3 * (x1 +! X2) | - y3
J4: m4 * x4 | - y4
Y5: m5 * (! X2 * x3) | - y5
.ENetOutput
.marking
.e
```

VHDL code:

controller description

```
ENTITY controller IS
PORT (reset, x1, x2, x3, x4, x5, clock: IN BIT; y1, y2, y3, y4, y5: OUT BIT);

END controller;
```

description of architecture

```
ARCHITECTURE data IS
/ position signals, where Nm is the next position
SIGNAL m1, Nm1: BIT;
SIGNAL m2, Nm2: BIT;
SIGNAL m3, Nm3: BIT;
SIGNAL m4, Nm4: BIT;
SIGNAL m5, Nm5: BIT;
/ transition signals
SIGNAL T1: BIT;
SIGNAL T2: BIT;
SIGNAL T3: BIT;
SIGNAL J4: BIT;
SIGNAL Y5: BIT;
```

process description

```
BEGIN
PROCESS BEGIN
WAIT UNTIL clock'EVENT and clock = '1';
IF reset = '0' THEN
m1 <= Nm1;
m2 <= Nm2;
m3 <= Nm3;
m4 <= Nm4;
m5 <= Nm5;
ELSE
m1 <= '1';
```

```
    m2 <= ‘0’;
    m3 <= ‘0’;
    m4 <= ‘0’;
    m5 <= ‘0’;
    END IF;
    END PROCESS;

/ description of the data flow for transitions, output signals of exceptional
    moments
    T1

 <= m1 AND x1;
    T2 <= m2 AND NOT x2;
    T3 <= m3 AND x1 OR NOT x2;
    J4 <= m4 AND x4;
    Y5 <= m5 AND (NOT x2) AND x3;

y1 = T1;
    y2 = T2;
    y3 = T3;
    y4 = J4;
    y5 = Y5;

/ situation when transitions are not available
    ASSERT NOT (T1 = ‘0’ AND T2 = ’0’ AND T3 = ’0’ AND J4 = ’0’ AND Y5 = ’0’)

    REPORT “No Enabled Transitions”
    WARNING;
    END data;
```

Chapter 8:
Wireless Sensing and Actuation: Energy Related Aspects

Joan Peuteman
Katholieke Universiteit Leuven

8.1. Introduction

An increasing number of physical systems are able to communicate and interact with each other using the Internet of Things. The interplay between physical systems, mobile networks, wireless communication, computing resources, and artificial intelligence facilitates a large number of new applications. Intelligent transport systems, smart driving, Industry 4.0, smart cities[158], smart homes, smart living, and smart governance illustrate this large number of applications [164].

In order to realise all these smart applications, wireless sensor networks are needed, which connect a (large) number of sensor nodes. Figure 8.1 visualises a simplified block diagram of a typical sensor node [168]. Notice the distinction between digital sensors (e.g., air quality sensor, humidity sensor) and analogue sensors (e.g., temperature sensor, light sensor). The measurement data is sent to a microcontroller where the data can be stored and processed. Notice also the presence of an antenna, which allows the communication with other devices. All parts of the wireless sensor node consume energy; however, providing that energy is a technical challenge.

8.1.1. Application: a smart parking system

Although the range of applications of wireless sensor networks, based on an IoT technology, is very broad, we restrict ourselves to a single example: a smart parking system [162]. In many countries, the number of people living in the cities is increasing fast. The general growth of the population is one of the reasons for that increase, while a rural exodus is also an important factor. In multiple countries, young people are attracted by the modern way of life and the opportunities of employment available in the cities. This ever-increasing population density in the cities implies a number of problems and challenges.

In cities, the intensity of the traffic is increasing fast, which causes pollution by the large amount of exhaust gases released by vehicles, traffic jams on the very busy roads, and difficulties in finding a space to park the car. A smart parking system (although many different implementations exist), based on wireless sensor networks, helps to solve or reduce the difficulties in finding a space to park. It is indeed important to avoid unnecessary driving in the city centre to find a parking space. Unnecessary driving increases drivers' frustration, traffic congestion, and the number of car accidents.

First, it is important that all empty parking spaces are detected by sensors (e.g., using infrared sensors or ultrasound sensors). Each sensor module contains a microcontroller, which processes the data originating from the sensor (e.g., infrared) to determine whether the parking space is empty or not. The sensor modules send their information to a car park management centre. For instance, wireless Zigbee

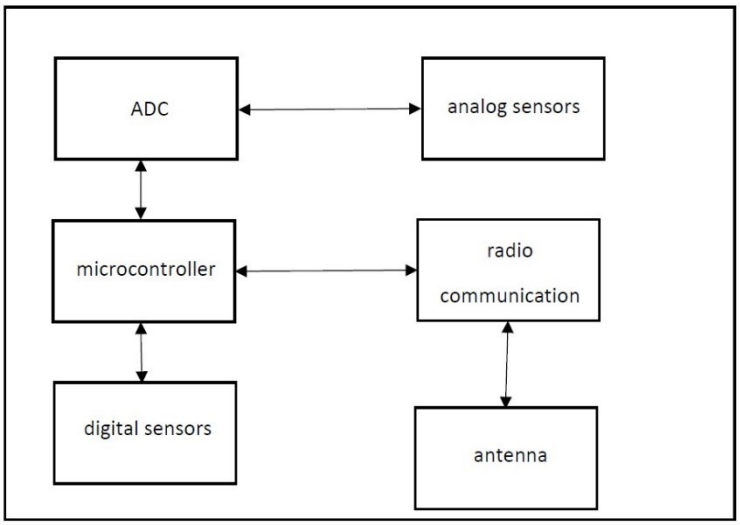


Fig. 8.1. Block diagram of a wireless sensor node

communication can be used, since Zigbee is a low-cost solution requiring only a limited amount of power to transmit the data. In comparison with other wireless communication protocols, the data transport capacity of Zigbee is limited, but in the present smart parking system, it is often sufficient. The car park management centre collects the data of all parking spaces. Based on that information and, other relevant factors such as internet connection, car drivers are able to view which park spaces are available and can reserve an empty parking space close to their actual location or close to their destination. The driver can use a laptop, a tablet or a smartphone to communicate and make the reservation.

Using the reservation system and an online payment system, the car driver reserves his/her parking space. Using Google Maps navigation, the driver is able to find his/her parking space as fast as possible generating a minimum of traffic load.

The smart parking system also needs a security module to avoid car parking in an unreserved parking place. For instance, a password is provided to drivers who have reserved a parking place. In case a driver parks his car in an unreserved parking place, this will be detected and an alert will be sent to the appropriate authorities, such as a car park attendant or a police officer.

8.1.2. Financial challenges

When realising smart applications based on wireless sensor nodes and wireless actuation, limiting the overall cost is very important. It is important to limit the initial cost in order to stimulate a start-up of the pilot phase of the project. After completing a pilot phase, it is also important to limit the costs needed to scale-up the system.

When considering the costs needed to realise IoT based projects, a distinction can be made between a number of components.

- The cost of the sensors/actuators, which are used to collect data, to process the data, and to implement the needed actions.
- The sensors need to communicate with each other implying the costs related to the network/communication infrastructure.
- Service infrastructure is needed to aggregate and render the information. It is important to minimise the costs related to this infrastructure. By using cloud services, for instance, the use of data storage or web servers simplifies the process, which reduces the cost.

It is also important to consider the salaries of the developers and the maintainers of the entire installation. Moreover, limiting the energy requirements is an important goal in order to limit the costs and to make the entire system more reliable.

8.2. Energy needs of wireless sensor nodes

As already mentioned, all components of a wireless sensor node, as visualised in Fig. 8.1, consume electrical power. The required power can be provided by the electrical grid, but quite often the sensor node is situated at a remote location or the sensor node is a mobile node. In these situations, providing power through the electrical grid is difficult, expensive, or even impossible. The sensor nodes can operate on primary batteries, but also the replacement of the batteries is often a difficult or even an impossible task. Reducing the power consumption in combination with a more compact battery energy storage extends the life expectancy of the battery but does not really solve the problem. By gathering energy from the environment (energy harvesting and energy scavenging) no replacement of batteries is needed anymore, i.e., the generated energy is used to feed the electronic equipment and to charge a secondary battery. The energy stored in this secondary battery allows the bridging of instants of time where the generated power is smaller than the consumed power [160].

8.2.1. Green Internet of Things

Based on the Internet of Things paradigm, an ever-increasing number of physical objects are able to communicate with each other. This allows sharing of information and making real-time decisions, with a minimum human input. The number of physical objects participating in the Internet of Things paradigm is even larger than

the Earth's population, implying that all together they consume huge amounts of energy. It is an ever-increasing challenge to reduce/limit the individual and total power consumption of these objects. However, it is important to reduce the power consumption of an individual object in order to increase its autonomy (e.g., in case they are fed by a battery) and to reduce/limit the total power consumption of all objects together in order to reduce/limit the environmental impact. By improving the energy efficiency, less energy is needed, implying less CO_2 emissions, i.e., a green or greener Internet of Things approach is obtained.

When reducing power consumption, software-based and hardware-based approaches are needed. In the present text, we restrict ourselves to a number of approaches, which facilitate the reduction of power consumption, but more information can be found in the literature. We make a distinction between

- radio optimisation techniques,
- sleep/wake up techniques,
- energy harvesting and wireless charging techniques,
- reduction of the data.

8.2.2. Radio optimisation techniques

When considering a wireless sensor node, as visualised in Fig. 8.1, and wireless sensor networks, the antenna-based radio communication is often the major energy consumption unit [157]. First, the electronic circuit must be powered but, second, the transmitted electromagnetic signal must be powered. In case of short communication distances, the power needed for the electronic circuit is larger than the power needed to transmit the signal. In case of a long-range communication, the power needed to transmit the signal using electromagnetic waves is larger than the power consumed by the electronic circuit.

Suppose the resulting communication is realised by combining a number of consecutive communications between adjacent nodes (daisy chain). By increasing the number of nodes, less electromagnetic power is needed to bridge the same total distance. Indeed, the distance between two adjacent nodes is inversely proportional to the number of nodes. The energy needed to communicate between two adjacent nodes is approximately proportional to the square of the distance between these two nodes. This implies that the energy needed to communicate between two adjacent nodes is approximately inversely proportional to the square of the number of nodes. Considering the number of consecutive communications and the energy needed to realise one single communication implies the total energy to communicate decreases as the number of nodes increases. Unfortunately, by increasing the number of nodes

the total delay also increases, because more hops will be required for data packet forwarding.

The number of nodes is not the only parameter which influences the energy consumption. Digital data can be sent using different modulation schemes. The power consumption also depends on the chosen modulation scheme.

8.2.3. Sleep/wake up strategy

By switching off non-active transceivers, an important amount of energy can be saved. Having switched off a transceiver, it operates in a so-called sleep mode. During the sleep mode, power consumption is considerably lower compared to the so-called active mode.

Suppose a duty cycle-based protocol has been used. Suppose during one cycle with length T, the transceiver is active during $\delta\ T$ (where the duty cycle $\delta < 1$) and the transceiver is in sleep mode during $(1 - \delta)\ T$. Suppose during the active mode, power P_{act} is consumed. Suppose during the sleep mode, power P_{sleep} is consumed and, of course, $P_{sleep} < P_{act}$. The average of the consumed power equals to

$$P_{avg} = \delta\ P_{act} + (1 - \delta)\ P_{sleep}. \tag{8.1}$$

It is useful to reduce P_{act} and P_{sleep}, but it is especially useful to reduce δ. Unfortunately, using a low duty cycle δ can lead to higher communication delays. Values of P_{act} and P_{sleep} vary in broad ranges. For instance, P_{sleep} can range from 1 μW to 0.1 mW, while P_{act} (transmitting electromagnetic waves) can range from 1 mW to 100 mW.

8.2.4. Energy harvesting and wireless charging techniques

Instead of using a primary battery (non-rechargeable), it is also possible to use a secondary battery (rechargeable) to supply the wireless sensor node with energy. Using energy harvesting or wireless charging techniques, it is possible to recharge the battery.

Energy harvesting relies on a broad range of energy resources, available around the wireless sensor node, and converts them into electrical energy. Small solar panels, small wind turbine generators, thermoelectric generators (e.g. based on the Seebeck effect), and electromagnetic generators provide the required electrical energy. Actually, small power renewable energy sources are used. More details about these approaches are provided below.

When wireless charging techniques are used, an electromagnetic field is used to transfer energy to the wireless sensor node. A transmitter device generates

a time-varying electromagnetic field, and a receiver captures energy from that electromagnetic field. By converting this captured energy into electrical energy, the electronic circuits can be fed without the use of wires or primary batteries.

8.2.5. Reduction of the data

As the amount of transmitted data decreases, the energy consumption also decreases. By reducing the sample rate of a sensor, the amount of measurement data will be reduced. A smaller amount of data needs to be processed, which reduces the energy needed to perform the computations by the processor. The amount of data which needs to be transmitted also decreases.

Also, by performing compression techniques, the amount of data to be transmitted decreases. Compressing the data at the transmitter side and decompressing the data at the receiver side also requires energy consumption, implying an optimum between computational energy and transmission energy is needed.

A combination of different approaches is needed to limit the total energy consumption [165]. Improved hardware implies that less energy is needed to transmit the same amount of raw data, in combination with decent data compression algorithms a further decrease in the required amount of energy is obtained.

8.3. Energy harvesting

Energy harvesting facilitates the conversion of ambient energy from the environment into electrical energy. The ambient energy can be very diverse:

- light energy (using solar panels);
- volume flow, i.e., a flow of liquids or gases (using a small wind turbine generator);
- kinetic energy, e.g., due to vibrations or movements of the energy harvesting device;
- thermal energy (using thermoelectric generators);
- electric or magnetic fields or radio waves from electrical equipment in the vicinity (using electromagnetic generators).

Quite often, the conversion efficiency is low and only low electrical powers are

generated. However, in combination with the already mentioned approaches used to reduce the energy consumption, these small amounts of energy can be sufficient to obtain a wireless sensor node which functions autonomously without the need to replace primary batteries.

In the literature, not only energy harvesting but also energy scavenging is considered. In general, the energy scavenging terminology is used when the ambient sources are unknown or very irregular. The energy harvesting terminology is used when the ambient sources are well characterised and are more regular.

In the present text, we restrict ourselves mainly to the use of light energy, vibrational energy, and thermal energy.

8.3.1. Solar and light energy

Solar cells can be used to capture power from the ambient light (in general, the power density is higher in an outdoor environment and smaller in an indoor environment). The generated energy can be stored in a secondary battery. The battery can possibly be combined with super-capacitors. Due to space limitations, and often also limited light intensities, the energy conversion efficiency of the solar cells is important. For instance, amorphous silicon solar cells have an efficiency of 11 %. Research is going on to obtain significantly higher efficiencies, for example, by using multiple band-gap solar cells (reaching efficiencies of, e.g., 35 %). A silicon solar cell typically has an open-circuit electrical potential of $0.7V$, which is low. A DC-DC converter can be used to boost the voltage level, but connecting a number of solar cells in series is also very common. Since a lot of wireless sensor nodes typically need $2V$, connecting three solar cells in series provides the needed voltage level.

8.3.2. Vibrational energy

When considering indoor applications, the light intensity is often too low to use solar cells in a satisfying way. Energy can also be harvested using the kinetic energy provided by vibrations. Different techniques facilitate the capturing of this vibrational energy. Piezoelectric materials can be used, i.e., some crystals have the ability to generate an electric potential in response to an applied mechanical stress [163]. The crystal contains dipoles and by compression or expansion, the electrical charges of the dipoles become aligned. This alignment leads to a net electric polarisation implying an electric potential across the crystal. In case of vibrations, the alternating mechanical compression and expansion leads to the generation of an AC voltage. An electronic circuit is needed to rectify and regulate the DC output voltage. This approach aims to provide a useful constant DC voltage.

An inductive system can also be used to harvest energy from available vibrations.

By moving a magnet in a wound coil, voltages are induced according to the law of Faraday. Capacitive systems also facilitate the harvesting of energy from vibrations. Vibrations are used to move one electrode of a capacitor, which facilitates the conversion of mechanical energy into electrical energy.

In the present text, a separate section will study a number of approaches to harvest energy using a capacitive system.

8.3.3. Thermal energy

Energy harvesting can also be based on thermal energy available[1] in the environment [161]. A thermoelectric power generator uses the temperature difference between two junctions to generate an electrical voltage. A thermoelectric generator based on the Seebeck effect is visualised in Fig. 8.2. A volt meter is used to measure the generated voltage.

The thermoelectric power generator in Fig 8.2 needs a heat source providing high temperature T_{hot} and a heat sink providing lower temperature T_{cold}. Two different conductor materials A and B are linked together to obtain junctions (a hot junction and a cold junction). At such a junction, the Seebeck effect provides an electrical voltage. By combining the voltages of both junctions, a total voltage

$$V = (\alpha_A - \alpha_B)(T_{hot} - T_{cold}) = \alpha_{AB}(T_{hot} - T_{cold}) \tag{8.2}$$

is obtained. Here, α_A and α_B are the Seebeck coefficients of conductors A and B (expressed in $V.K^{-1}$). By defining $\alpha_{AB} = \alpha_A - \alpha_B$, a voltage proportional with the temperature difference $T_{hot} - T_{cold}$ is obtained (when neglecting the impact of the temperature on α_A and α_B). The generated voltage increases either by increasing $T_{hot} - T_{cold}$ or by connecting several thermoelectric power generators in series.

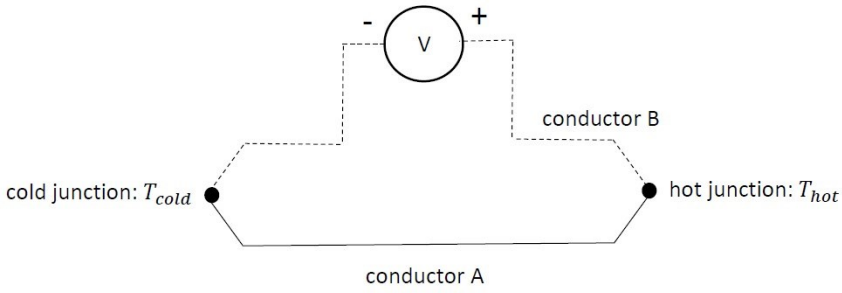


Fig. 8.2. Thermoelectric generator based on the Seebeck effect

[1] https://www.electrical4u.com/thermoelectric-power-generators-or-seebeck-power-generation/

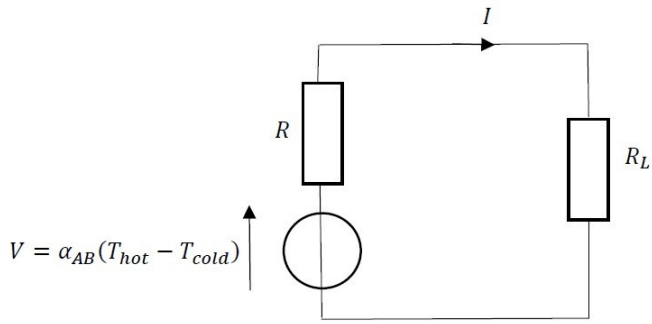


Fig. 8.3. Thermoelectric generator with electrical load

In case the volt meter is replaced by a load resistor R_L, a DC current I will flow and power will be sent to the resistor (or more general the electrical load). Notice, however, that the thermoelectric generator can be modelled as a non-ideal voltage source, as visualised in Fig. 8.3 (series connection of an ideal voltage source V and an internal resistor R).

Since current

$$I = \frac{V}{R+R_L} = \frac{\alpha_{AB}(T_{hot}-T_{cold})}{R+R_L}, \tag{8.3}$$

power P_L is sent to the load equals

$$P_L = R_L \frac{\alpha_{AB}^2 (T_{hot}-T_{cold})^2}{(R+R_L)^2}. \tag{8.4}$$

Power P_L is maximised when $R_L = R$ giving a maximum power

$$P_{L\,MAX} = \frac{\alpha_{AB}^2 (T_{hot}-T_{cold})^2}{4\,R}. \tag{8.5}$$

A larger power can be extracted when the internal resistor R is smaller. When decreasing the length of the conductors and the cross sections of the conductors in Fig. 8.2, a smaller internal resistor is obtained.

When connecting n thermoelectric generators in series, the generated voltage becomes $n\,\alpha_{AB}(T_{hot} - T_{cold})$ but also the internal resistor becomes nR. Maximum power will be extracted when $R_L = n\,R$.

8.4. Energy harvesting based on mechanical vibrations

As already mentioned, energy harvesting facilitates the conversion of ambient energy from the environment into electrical energy. The ambient energy source can be very diverse, i.e., light energy, energy available in a volume flow, thermal energy, energy available in electric or magnetic fields, energy available in radio waves, and

kinetic energy due to vibrations. In the present section, we focus on harvesting energy from vibrations [167].

8.4.1. Mechanical behaviour and the extracted power

Converting the energy originating from ambient vibrations usually occurs in two steps. The first step is a mechanical-to-mechanical conversion. The vibrations are converted into a relative motion between two elements based on a mass-spring-damper system. Secondly, a mechanical to electrical conversion is needed (using piezoelectric materials, using an inductive system, using a capacitive system).

8.4.1.1. Mass-spring-damper system

Since the ambient vibrations generally have a low amplitude, the use of a mass-spring-damper system is needed to amplify the relative movement of a mass compared to the original movement. Mechanical resonance is needed to have a sufficiently large relative movement which facilitates the realisation of the mechanical to electrical conversion in a decent way.[2]

Figure 8.4 visualises a seismic mass *m* connected with a spring having a spring constant *k* [170]. The movement of the mass is damped with a viscous damping constant *d*. The viscous damping models mechanical friction losses and the electrical energy extracted from the system. In the present calculations, we assume the mechanical friction losses can be neglected in comparison to the electrical energy extracted by the energy harvester. This assumption implies that in a steady state situation all power provided to the system in Fig. 8.4 will be converted into electrical energy.

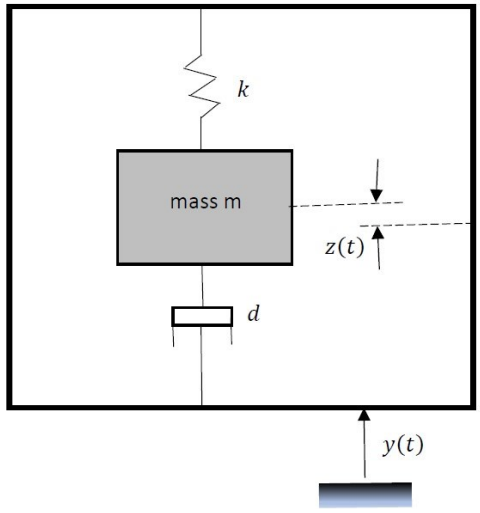


Fig. 8.4. Mechanical-to-mechanical conversion based on vibrations

[2] http://farside.ph.utexas.edu/teaching/315/Waves/node13.html

During the calculations, we also assume that the mass of the vibration source is very large in comparison to mass *m*. We also assume that the vibration source is an infinite source of power. These assumptions imply that the vibration noise $y(t)$ is not affected by movement $z(t)$ of mass *m*.

The behaviour of mass *m* can be modelled by the differential equation

$$m\,\ddot{z}(t) + d\,\dot{z}(t) + k\,z(t) = -m\,\ddot{y}(t)\,. \tag{8.6}$$

Here, *m* is the seismic mass, *k* is the spring constant, and *d* is the viscous damping constant. Notice that $z(t)$ is the movement of the vibration which provides all energy, i.e., the housing in Fig. 8.4 also vibrates with displacement $y(t)$. The relative motion of mass *m* with respect to the housing equals $z(t)$.

The force applied to the mass-spring-damper system equals $F(t) = -m\,\ddot{y}(t)$. The same force is also applied to mass *m*. Originating from the damper, force $d\,\dot{z}(t)$ is applied (viscous friction) and due to the spring force $k\,z(t)$ is applied.

In a real-life situation, displacement $y(t)$ can be irregular. But using Fourier theory, $y(t)$ can be considered to be a composition of a large (infinite) number of cosine functions having different pulsations ω. In the present analysis, we consider a sinusoidal excitational vibration $y(t) = Y\,cos(\omega t)$ which implies that $\ddot{y}(t) = -\omega^2\,Y\,cos(\omega t)$. The differential equation reduces to

$$m\,\ddot{z}(t) + d\,\dot{z}(t) + k\,z(t) = +m\,\omega^2\,Y\,cos(\omega t). \tag{8.7}$$

8.4.1.2. Natural pulsation

By defining the natural pulsation of the system as

$$\omega_n = \sqrt{\frac{k}{m}}\,, \tag{8.8}$$

the steady state solution of the differential equation equals

$$z(t) = Z\,cos(\omega t - \Phi) \tag{8.9}$$

with

$$Z = \frac{m\,\omega^2\,Y}{\sqrt{(k - m\omega^2)^2 + d^2\omega^2}} \tag{8.10}$$

and

$$\Phi = tan^{-1}\left(\frac{d\,\omega}{k - m\,\omega^2}\right)\,. \tag{8.11}$$

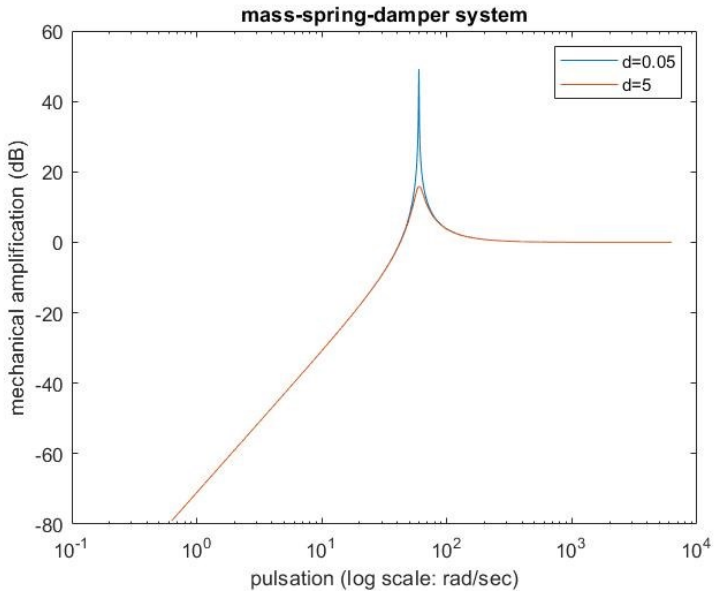


Fig. 8.5. Mechanical amplification of a mass-spring-damper system

Figure 8.5 visualises the evolution of ratio Z/Y as a function of ω in case m = 0.52 kg and k = 1847 N/m (although in small scale applications mass m can be considerably smaller, the parametres are inspired by the paper of A. V. Pedchenko et al. [166]). Two different values for the viscous damping constant are considered (d = 0.05 M/(m/s) and d = 5 N/(m/s).

Figure 8.5 expresses the mechanical amplification using a logarithmic scale $20\ log_{10}\left(\frac{Z}{Y}\right)$ expressed in dB. Notice also the logarithmic scale for pulsation ω. Close to the natural pulsation $\omega_n \cong 60$ rad/sec a large amplification is obtained. The lower the damping constant d, the larger the amplification. For pulsations lower than the natural pulsation, even an attenuation is obtained. For pulsations higher than the natural pulsation, the amplitudes of $y(t)$ and $z(t)$ are approximately the same.

In case the spectrum of excitational vibration $y(t)$ contains a large number of pulsations, only the pulsation(s) close to the natural pulsation will be amplified. Actually, a sufficiently large motion of mass m is expected and needed (with pulsation ω_n), which facilitates the mechanical to electrical energy conversion.

8.4.1.3. The average of the instantaneous power

As already mentioned, the force applied to the mass-spring-damper system equals $F(t) = -m\ \ddot{y}(t)$. The instantaneous power transferred to mass m is the product of the force and the velocity of the mass. The instantaneous power equals to

$$p(t) = -m\ \ddot{y}(t)\ (\dot{y}(t) + \dot{z}(t)). \tag{8.12}$$

With $y(t) = Y\ cos(\omega t)$, $\dot{y}(t) = -\omega\ Y\ sin(\omega t)$, $\ddot{y}(t) = -\omega^2\ Y\ cos(\omega t)$, $z(t) = Z\ cos(\omega t - \Phi)$ and $\dot{z}(t) = -\omega\ Z\ sin(\omega t - \Phi)$ the instantaneous power equals

$$p(t) = m\ \omega^2\ Y\ cos(\omega t)\ (-\omega\ Y\ sin(\omega t) - \omega\ Z\ sin(\omega t - \Phi)) \tag{8.13}$$

$$\boldsymbol{p(t) = m\ \omega^2\ Y\ cos(\omega t)\ (-\omega)\ ((Y + Z\ cos\Phi)\ sin(\omega t) - Z\ sin\Phi\ cos(\omega t))}. \tag{8.14}$$

By taking the average of the instantaneous power, P_{AVG} is obtained (taking into account the average of $cos(\omega t)\ sin(\omega t)$ equals zero and the average of $cos^2(\omega t)$ equals 0.5) with

$$P_{AVG} = m\ \omega^3\ Y\ sin\Phi\ \frac{Z}{2}\ . \tag{8.15}$$

Since $in\Phi = \frac{d\ \omega}{\sqrt{(k - m\omega^2)^2 + d^2\omega^2}}$, $Z = \frac{m\ \omega^2\ Y}{\sqrt{(k - m\omega^2)^2 + d^2\omega^2}}$, one obtains that

$$P_{AVG} = \frac{m\ \omega^4\ Y\ Z\ d}{2\sqrt{(k - m\ \omega^2)^2 + d^2\omega^2}}; \tag{8.16}$$

$$\boldsymbol{P_{AVG} = \frac{1}{2}\ \frac{m^2\ \omega^6\ Y^2\ d}{(k - m\ \omega^2)^2 + d^2\omega^2} = \frac{1}{2}\ \frac{(m^2\ \omega^6\ Y^2\ d)/k^2}{(1 - \frac{m}{k}\omega^2)^2 + \frac{d^2}{k^2}\omega^2}} \tag{8.17}$$

By defining the transducer damping factor

$$\zeta_t = \frac{d}{2\sqrt{k\ m}}, \tag{8.18}$$

one obtains that

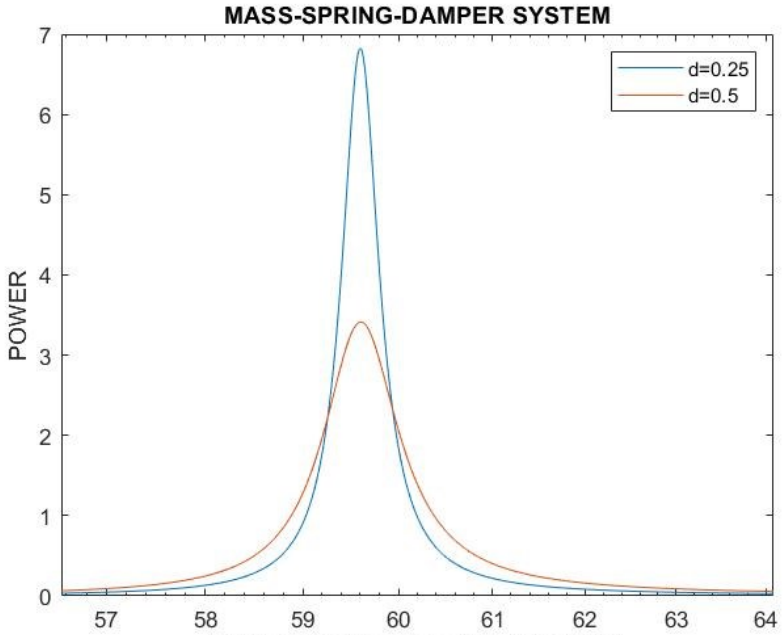


Fig. 8.6. Available power using a mass-spring-damper system

$$\frac{d^2}{k^2}\,\omega^2 = \frac{d^2}{k}\,\frac{\omega^2}{\omega_n^2}\,\frac{1}{m} = 4\,\zeta_t^2\,\frac{\omega^2}{\omega_n^2} = \left(2\,\zeta_t\,\frac{\omega}{\omega_n}\right)^2 \tag{8.19}$$

and that

$$\frac{m^2\,\omega^6\,Y^2\,d}{2\,k^2} = \frac{d}{2\sqrt{k\,m}}\,m\,\omega^3\,Y^2\,\left(\frac{\omega}{\omega_n}\right)^3 = m\,\zeta_t\,Y^2\,\left(\frac{\omega}{\omega_n}\right)^3\,\omega^3\,. \tag{8.20}$$

This implies that the average of the instantaneous power equals

$$P_{AVG} = \frac{m\,\zeta_t\,Y^2\,\left(\frac{\omega}{\omega_n}\right)^3\,\omega^3}{\left(1-\left(\frac{\omega}{\omega_n}\right)^2\right)^2+\left(2\,\zeta_t\,\frac{\omega}{\omega_n}\right)^2}\,. \tag{8.21}$$

Figure 8.6 visualises the evolution of the average power of the mass-spring-damper system as a function of ω in case $m = 0.52$ kg and k = 1847 N/m (with Y = 1 mm). Two different values for the viscous damping constant are considered (d = 0.25 N/(m/s) and (d = 0.5 N/(m/s)).

8.4.1.4. Generated electrical power

The damping factor determines the selectivity of the system. In case the frequencies of the vibration are well known and concentrated around one point, a low damping factor can be used providing a large power. Conversely, when the frequencies in $y(t)$ are varying with respect to time or when they are not well known, a higher damping factor is needed. By having a higher damping factor, the bandwidth of the system increases but the available power decreases.

In the calculations, we assumed that the mechanical friction losses can be neglected in comparison with the electrical energy extracted by the energy harvester. This assumption implies that, in a steady state situation, all power provided to the system in Fig. 8.4 will be converted into electrical energy.

When the energy harvester operates at its resonance pulsation ($\omega = \omega_n$), the power equals to

$$P_{AVG} = \frac{m\,Y^2\,\omega_n^2}{4\,\zeta_t} = \frac{m^2 Y^2 \omega_n^3}{2\,d}\,, \tag{8.22}$$

which implies that the power output is proportional to the cube of the pulsation. Notice also the impact of the transducer damping factor ζ_t. By reducing ζ_t, a larger power can be obtained. However, reducing the tranducer damping factor increases amplitude Z of displacement *z(t)* of mass m. Of course, the maximum distance the mass can move is limited by the size and geometry of the entire system.

8.4.1.5. Micro-electromechanical-systems (MEMS)

Figures 8.5 and 8.6 give a numerical example with large mass m. In order to have a higher natural pulsation ω_n, mass m needs to be smaller. Actually, MEMS (micro-electromechanical-systems) are obtained implying small sizes. A number of common

design rules can be taken into consideration:

- A sufficiently large mass m is needed within the available volume of the transducer (P_{AVG} is proportional with m^2 for a given ω_n and d).
- The allowed amplitude of the movement of the mass must be as large as possible.
- The spring should be designed to obtain a resonance pulsation which matches the pulsation (or a pulsation) of $y(t)$.
- The transducer damping factor ζ_t must be sufficiently small to allow mass m to move to the limit of its range.

8.4.2. Harvesting using a capacitive system

Different techniques facilitate the capturing of the vibrational energy: the use of piezoelectric materials, the use of inductive systems, the use of capacitive systems. In the present section, we restrict our focus to the use of capacitive systems.

The vibrational energy is harvested using a variable capacitor C. Consider a capacitor having two electrodes having area A. Between the electrodes, a dielectric having an absolute permittivity ε and thickness d implies a capacitor value

$$C = \frac{\varepsilon A}{d}. \tag{8.23}$$

By increasing d, the capacitor value C decreases. By decreasing d, the capacitor value C increases. When a DC voltage U is applied to the capacitor, charge $Q = CU$ is stored in the capacitor. The electric field in the dielectric stores energy and this energy equals to

$$E = \frac{C\,U^2}{2} = \frac{Q^2}{2\,C} = \frac{Q\,U}{2}. \tag{8.24}$$

Consider a capacitor where the position of a first electrode is fixed and the position of a second electrode can change in order to change d and C. Using the vibrational energy, the second electrode can move, which allows the harvesting of energy. A distinction can be made between two approaches [159]:

- a charge-constrained cycle,
- a voltage-constrained cycle.

8.4.2.1. Charge-constrained cycle

First, the working principle of a charge-constrained cycle will be considered.

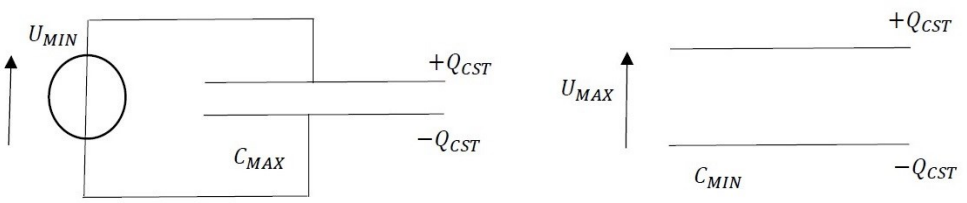


Fig. 8.7. Energy harvesting based on a charge-constrained cycle

Consider the situation visualised in Fig. 8.7. A capacitor has maximum value C_{MAX} (with a small gap between the electrodes) and using a DC voltage source U_{MIN}, the capacitor is charged with Q_{CST}. The energy stored in the capacitor equals to

$$E_2 = \frac{Q_{CST}\, U_{MIN}}{2} = \frac{Q_{CST}^2}{2\, C_{MAX}}. \tag{8.25}$$

When removing the voltage source, an open circuit is obtained implying that charge Q_{CST} remains constant. By moving one of the electrodes to obtain a larger gap d, the capacitor value decreases from C_{MAX} to C_{MIN}. With constant Q_{CST}, the voltage across the capacitor increases from U_{MIN} to U_{MAX} with

$$U_{MIN} = \frac{Q_{CST}}{C_{MAX}},\ U_{MAX} = \frac{Q_{CST}}{C_{MIN}}. \tag{8.26}$$

The energy stored in the capacitor increases from E_2 to

$$E_3 = \frac{Q_{CST}\, U_{MAX}}{2} = \frac{Q_{CST}^2}{2\, C_{MIN}}. \tag{8.27}$$

Once C_{MIN} and U_{MAX} have been reached, the electrical energy E_3 is extracted from the capacitor as visualised in Fig. 8.8. While discharging the capacitor, the capacitor provides a current (i.e., power and finally energy E_3) to the electrical load.

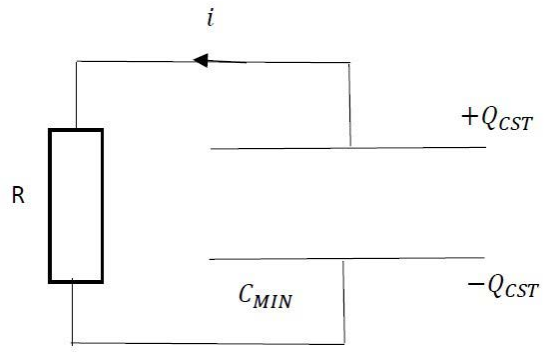


Fig. 8.8. Extracting electrical energy from the capacitor

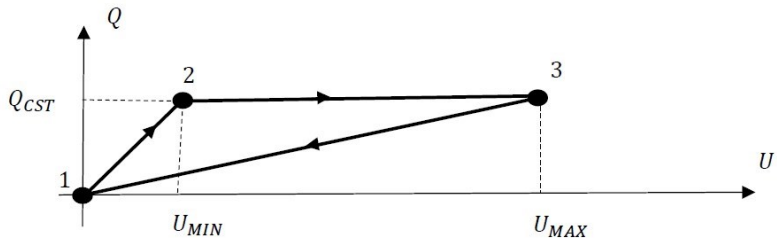


Fig. 8.9. Charge-constrained cycle

The entire charge-constrained cycle is visualised in Fig. 8.9. Notice first the uncharged capacitor in state 1. Notice the charging process with a transition from state 1 to state 2 requiring a voltage source to supply energy E_2. Notice the transition from state 2 to state 3, where mechanical energy is converted into electrostatic energy stored in the capacitor leading to a total stored energy E_3. When discharging the capacitor, as visualised in Fig. 8.8, E_3 is released. This implies that each cycle provides a net amount of electrical energy which equals

$$E_{Q=CTE} = E_3 - E_2 = \frac{1}{2}\, Q_{CST}^2 \left(\frac{1}{C_{MIN}} - \frac{1}{C_{MAX}}\right) = \frac{Q_{CST}}{2}\left(U_{MAX} - U_{MIN}\right). \tag{8.28}$$

Notice that the net amount of energy $E_{Q=CST}$ provided by the charge-constrained cycle corresponds to the area of the triangle in Fig. 8.9.

8.4.2.2. Voltage-constrained cycle

The working principle of a voltage-constrained cycle [169] corresponds to Figs. 8.10 and 8.11. Initially, a capacitor has maximum value C_{MAX} (with a small gap between the electrodes), and using a DC voltage source U_{CST}, the capacitor is charged with $Q_{MAX} = C_{MAX} U_{CST}$ (transition from state 1 to state 2 in Fig. 8.10). The energy stored in the capacitor equals

$$E_2 = \frac{Q_{MAX}\, U_{CST}}{2} = \frac{Q_{MAX}^2}{2\, C_{MAX}} = \frac{C_{MAX}\, U_{CST}^2}{2}. \tag{8.29}$$

The voltage source U_{CST} is kept constant. By moving one of the electrodes to obtain a larger gap d, the capacitor value decreases from C_{MAX} to C_{MIN} (transition from state 2 to state 3 in Fig. 8.10). With constant U_{CST}, the charge stored in the capacitor decreases from Q_{MAX} to Q_{MIN} with

$$Q_{MAX} = C_{MAX} U_{CST},\ Q_{MIN} = C_{MIN} U_{CST}. \tag{8.30}$$

The energy stored in the capacitor decreases from E_2 to

$$E_3 = \frac{Q_{MIN}\, U_{CST}}{2} = \frac{Q_{MIN}^2}{2\, C_{MIN}} = \frac{C_{Min}\, U_{CST}^2}{2}. \tag{8.31}$$

During the transition from state 2 at t_2 to state 3 at t_3, the capacitor provides

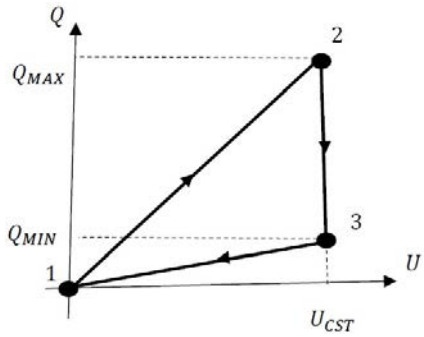


Fig. 8.10. Voltage-constrained cycle

current $i(t)$ to the voltage source. The energy supplied to the voltage source equals to

$$\int_{t_2}^{t_3} U_{CST}\, i(t)\, dt = U_{CST} \int_{t_2}^{t_3} i(t)\, dt = U_{CST}(Q_{MAX} - Q_{MIN}). \tag{8.32}$$

This implies that electrical energy $U_{CST}(Q_{MAX} - Q_{MIN})$ is supplied to the voltage source. Finally, also energy E_3 can be recovered, which is indicated by the transmission from state 3 to state 1 in Fig. 8.10.

Figure 8.11 implies that each cycle provides a net amount of electrical energy which equals to

$$E_{U=CTE} = U_{CST}(Q_{MAX} - Q_{MIN}) + E_3 - E_2\,, \tag{8.33}$$

which implies that

$$E_{U=CTE} = U_{CST}(Q_{MAX} - Q_{MIN}) + E_3 - E_2\,, \tag{8.34}$$

since $Q_{MAX} = C_{MAX}U_{CST}$ and $Q_{MIN} = C_{MIN}U_{CST}$, each cycle provides a net electrical energy which equals to

$$E_{U=CTE} = \frac{U_{CST}^2}{2}(C_{MAX} - C_{MIN})\,. \tag{8.35}$$

Notice that the net amount of electrical energy $E_{U=CST}$ provided by the charge-constrained cycle corresponds with the area of the triangle in Fig. 8.10.

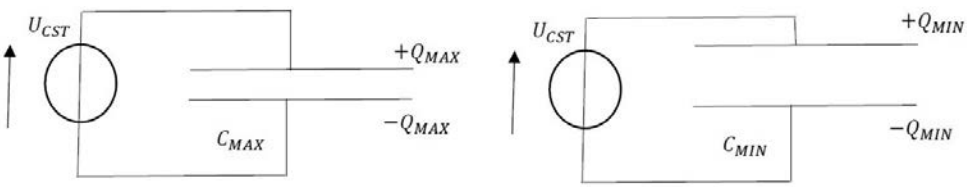


Fig. 8.11. Energy harvesting based on a voltage-constrained cycle

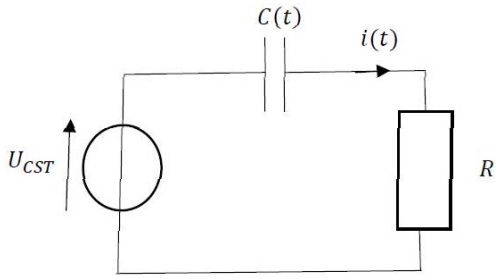


Fig. 8.12. Energy harvesting using a variable capacitor

8.4.2.3. Energy harvesting using a variable capacitor

Alternatively, energy can be harvested based on the circuit visualised in Fig. 8.12. The circuit in Fig. 8.12 contains constant voltage source U_{CST}, electrical load R and capacitor $C(t)$. In case the capacitor is constant, no current is flowing and no power is sent to load R. In case capacitor $C(t)$ is varying with respect to time, a current will flow and power will be sent to load R.

Suppose the area of the electrodes of the capacitor equals A and assume the dielectric has an absolute permittivity ε. One of the electrodes is moving, which implies that the distance between the electrodes is not constant. With displacement $z(t) = Z\ cos(\omega t)$, the distance between the electrodes equals $d(t) = d_0 + Z\ cos(\omega t)$ with $Z \ll d_0$. The capacitor value equals

$$C(t) = \frac{\varepsilon\ A}{d(t)} = \frac{\varepsilon\ A}{d_0 + Z\ cos(\omega t)} = C_0\ \frac{1}{(1+\frac{Z\ cos(\omega t)}{d_0})} \cong C_0\ (1 - \frac{Z\ cos(\omega t)}{d_0})\ . \tag{8.36}$$

With $C_0 = \frac{\varepsilon\ A}{d_0}$ and by defining

$$C_{MAX} = C_0\ (1 + \frac{Z}{d_0}), \quad C_{MIN} = C_0\ (1 - \frac{Z}{d_0}), \tag{8.37}$$

one obtains that

$$C(t) = (\frac{C_{MAX}+C_{MIN}}{2}) - (\frac{C_{MAX}-C_{MIN}}{2})\ cos(\omega t) = C_0 - (\frac{C_{MAX}-C_{MIN}}{2})\ cos(\omega t). \tag{8.38}$$

When considering capacitor $C(t)$ and voltage $u(t)$ across the capacitor, electrical charge $Q(t) = C(t)\ u(t)$ appears. By taking the derivative with respect to time, one obtains current

$$i(t) = \frac{d\ Q(t)}{d\ t} = C(t)\ \frac{d\ u(t)}{d\ t} + u(t)\ \frac{d\ C(t)}{d\ t}. \tag{8.39}$$

With $u(t)$ being the voltage across the capacitor in Fig. 8.12, the voltage law of Kirchhoff reveals that

$$u(t) = U_{CST} - R\ i(t)\ , \tag{8.40}$$

$$i(t) = -R\,C(t)\,\frac{d\,i(t)}{d\,t} + (U_{DC} - R\,i(t))\,\frac{d\,C(t)}{d\,t}. \tag{8.41}$$

or is equivalent to

$$(1 + R\,\frac{d\,C(t)}{d\,t})\,i(t) + R\,C(t)\,\frac{d\,i(t)}{d\,t} = U_{DC}\,\frac{d\,C(t)}{d\,t}. \tag{8.42}$$

By defining

$$\Delta C = \frac{C_{MAX} - C_{MIN}}{2}, \tag{8.43}$$

the time-dependent differential equation becomes

$$(1 + \omega R\,\Delta C\,sin(\omega t))i(t) + R(C_0 - \Delta C\,cos(\omega t))\frac{d\,i(t)}{d\,t} = U_{DC}\,\omega\,\Delta C\,sin(\omega t) \tag{8.44}$$

The time-dependent differential equation can be rewritten as

$$\frac{d\,i(t)}{d\,t} + \frac{(1+\omega R\,\Delta C\,sin(\omega t))}{R(C_0 - \Delta C\,cos(\omega t))}\,i(t) = \frac{U_{DC}\,\omega\,\Delta C\,sin(\omega t)}{R(C_0 - \Delta C\,cos(\omega t))}, \tag{8.45}$$

which is a differential equation of the form

$$\frac{d\,i(t)}{d\,t} + P(t)\,i(t) = Q(t) \tag{8.46}$$

with the obvious definitions for $P(t)$ and $Q(t)$. Calculations are omitted here, but by defining

$$v(t) = e^{\int P(t)\,dt}, \tag{8.47}$$

the differential equation can be solved by multiplying both sides of the equation with $v(t)$ giving

$$(\frac{d\,i(t)}{d\,t} + P(t)\,i(t))\,e^{\int P(t)\,dt} = \frac{d}{dt}(i(t)\,e^{\int P(t)\,dt}) = Q(t)\,e^{\int P(t)\,dt}. \tag{8.48}$$

This implies that the solution of the differential equation equals

$$i(t) = e^{-\int P(t)\,dt}\int Q(t)\,e^{\int P(t)\,dt}\,dt. \tag{8.49}$$

The electrical power sent to the resistor equals $R\,i^2(t)$.

8.4.2.4. Electret-based electrostatic harvester

When considering electrostatic converters, it is possible to avoid the use of the voltage sources in Figs. 8.7, 8.11 or 8.12 by using an electret-based converter [138]. Electrets are dielectric materials which are in a quasi-permanent electric polarisation state. Electrets are equivalent to permanent magnets (but electrostatic), i.e., the word "electret" comes from "electricity magnet".

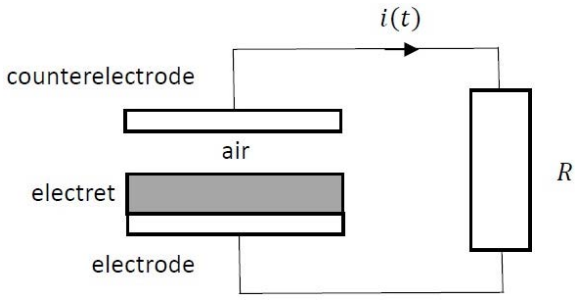


Fig. 8.13. Electret-based electrostatic harvester

Electret's polarisation is obtained by dipole orientation (instead of the orientation of magnetic domains in a permanent magnet) inside the electret. By combining an electret with two electrodes (named electrode and counterelectrode in Fig. 8.13), current $i(t)$ can be obtained if the counterelectrode is moving. This approach allows to convert mechanical energy into electrical energy without the need of initial electrical energy, as it is the case when no electret is used and a DC voltage source is needed. Several approaches exist, but the electret, electrode, and counterelectrode behave as a series connection of a constant voltage source and a variable capacitor (similar with the situation visualised in Fig. 8.12).

8.5. Wireless charging techniques

A transmitter device generates a time-varying electromagnetic field, and a receiver captures energy from that electromagnetic field. By converting this energy into electrical energy, the electronic circuits can be fed without the use of wires or primary batteries.

When considering an electromagnetic field, a distinction can be made between the near field and the far field.[3] When using the near field, power is only transferred over short distances. When considering the near field, a distinction can be made between inductive coupling (using a magnetic field) and capacitive coupling (using an electric field).

In the present text, the use of wireless powering will be restricted using the far field of an electromagnetic field in order to achieve a longer distance between the energy transmitter and the energy receiver of, e.g., a wireless sensor node. Although wireless powering is also possible using visible light from lasers, power transmission using radio waves can be made more directional by using electromagnetic waves having shorter wavelengths. The use of the microwave range is quite common, and the power receiving device (e.g., a wireless sensor node) contains a rectenna. The rectenna allows to convert the microwave energy into electrical energy (a DC voltage and a DC current are obtained).

[3]https://en.wikipedia.org/wiki/Wireless_power_transfer

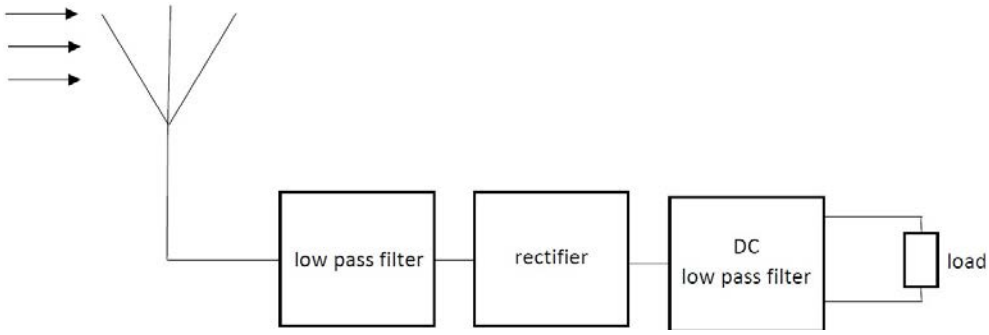


Fig. 8.14. Wireless charging using a rectenna

A rectenna is actually a rectifying antenna as visualised in Fig. 8.14. Notice first the receiving antenna converting electromagnetic waves into an electrical AC voltage. The AC voltage will be rectified, and, finally, a low pass filter is obtained to obtain a constant DC voltage implying a constant DC current in the electrical load.

8.6. Conclusions

The text underlined the practical use of wireless sensor nodes. Unfortunately, these wireless sensor nodes are often located at remote places, implying that it is a challenge to power them. Instead of using primary batteries (which need to be replaced from time to time), energy harvesting or wireless powering can be a decent solution. Since energy harvesting in general provides quite small powers and wireless powering also accounts for losses, it is important to limit the energy consumption of the wireless sensor node (or another electronic device) as much as possible.

Chapter 9:

Hardware and Software of Networks

Andrei Varuyeu
Gomel State University

9.1. The concept of a digital network

For the first time, computer networks appeared almost simultaneously with computers. This was due to the fact that the resource of computer time was extremely expensive and it was important to share its cost among several users. Users got the opportunity to prepare their data in parallel, which could then be processed either sequentially or in parallel (in the form of packets) by the blocks of the computing system.

This is how the principles of resource sharing and terminal systems emerged. These systems were widely used until the 80s of the 20th century, and some samples – almost until the beginning of the 21st century.

The rebirth of computer networks was caused by the practical need for data sharing by users of personal computing systems. From this point of view, the following definition is suitable for computer networks: a network is a system of independent computers connected to each other for the purpose of sharing data, peripherals, and other network resources.

Globally, the total number of digital network users will more likely grow from 4.5 billion in 2020 to 5.3 billion by 2023, at CAGR a projected growth rate of 6 percent. In relative terms, this represents 57 percent of the global population in 2020 and 66 percent of the global population by 2023 (Fig. 9.1).

Initially, all networks could be divided into two classes:

- data exchange networks or information networks;
- data processing networks or computer networks.

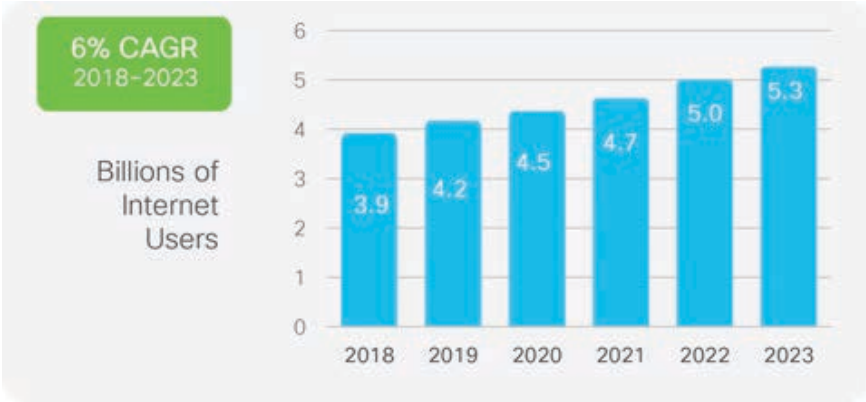


Fig. 9.1. Global Internet user growth [171]

Information networks included systems for transmitting signals, messages, data ,and other types of information. Distributed and unified computing systems belonged to the data processing networks. But, since distributed processing requires the use of information exchange mechanisms, this line has gradually been erased and at the moment all computer networks are both information and computational. Therefore, the more general term – digital networks is often used.

Digital networks can be viewed from different perspectives:

- for a running program, a network is a complex system of routes for transmitting data and resources for processing them;
- for the user, a computer network is a tool for accessing network resources;
- for the manager, the network is a means of managing production processes;
- for a network designer, it is a set of standards and requirements that must be observed during project implementation.

A modern digital network has the following properties:

- an excellent combination of "performance-usability-cost" of computing resources;
- sharing of data and devices;
- online access to extensive corporate information;
- use of external data;
- integration of information systems.

However, there are also problems associated with the implementation of networks, for example, complex programming for distributed systems, ensuring software compatibility, ensuring the reliability of information transfer, ensuring security. The area of using computer networks today is constantly expanding, it includes science, education, business, entertainment.

The performance of a network is often measured by the rate of communication that can be realized in its environment. This approach is based on the fact that different types of network services have different requirements for network bandwidth (Fig. 9.2). The combination of several types of services within a single network structure imposes additional requirements on the equipment used and the supporting software systems. Thus, modern networks are complex hardware and software systems.

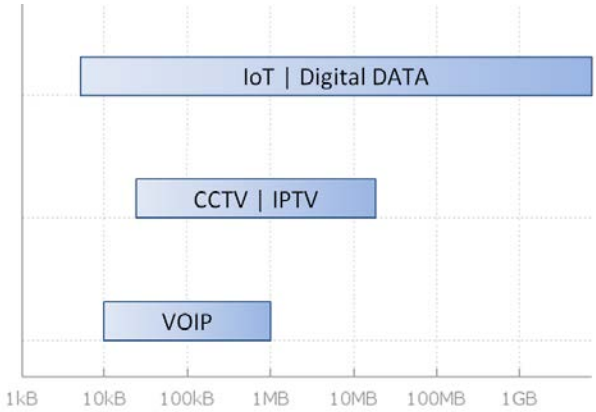


Fig. 9.2. The level of network services requirements for network bandwidth

Forward-looking projections, the number of devices and connections is growing faster (10 % CAGR) than the population (CAGR 1.0 %). The growth in the number of Internet users is accelerating (CAGR 6 %). This trend is accelerating the growth of the average number of devices and connections per person. The growing number of D2D applications such as video surveillance, smart meters and sensors, transportation management, health monitoring, and package or asset tracking are making an important contribution to the growth of devices and connections. By 2023, D2D connections will account for half or more than half of all devices and connections.

D2D connections are becoming the fastest growing category of devices and connections, growing nearly 2.4-fold (19 percent CAGR) to 14.7 billion connections by 2023. Smartphones will be in the second place in terms of growth with a 7 percent CAGR (growth rate 1.4). Connected TVs are set to be in second place in terms of growth (CAGR around 6 percent) to 3.2 billion by 2023. At the same time, the number

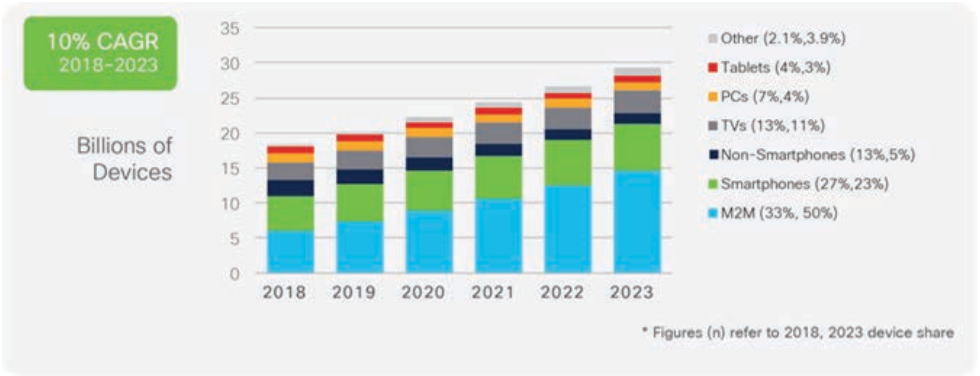


Fig. 9.3. Global device and connection growth [171]

of PCs will decrease (decrease by 2.3 percent). However, in the forecast period and by the end of 2023, there will be more PCs than tablets (1.2 billion PCs versus 840 million tablets).

By 2023, the share of individual users in the total number of devices, including both stationary systems and mobile, will be 74 percent, and the share of business will be 26 percent. The share of individual users will grow at a slightly slower pace, at 9.1 % CAGR, compared to the business segment, which will grow at 12.0 % CAGR (Fig. 9.3).

9.2. Virtualization of network nodes and network segments

Traditionally, corporate servers are built as a server operating system (OS) installed on dedicated hardware. All of the server's RAM, processing power, and hard disk space are used to implement the services (for example, for the Internet, postal services, etc.)

The main issue with this configuration is reliability. If any component fails, the service provided by this server becomes unavailable. This is called a single point of failure. The same problem applies to hosts that relay traffic. The second problem is that dedicated server resources are often underutilized. Dedicated servers sit idle for a long time, waiting for the need to provide the specific service they provide. These servers waste energy and take up more space than their service format allows. This situation is called server sprawl.

Network node virtualization reclaims unused resources and consolidates the number of required servers and network devices. It also allows multiple operating systems to be used on a single hardware platform. Additionally, it becomes possible

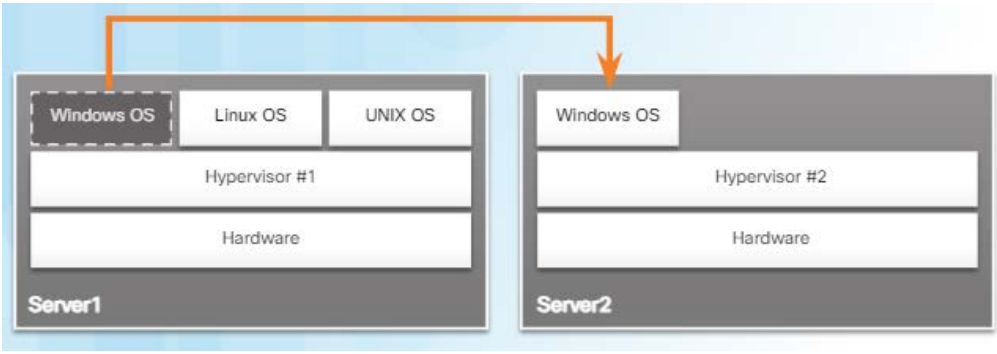


Fig. 9.4. Migration of the Windows virtual server

to balance the load. For example, consider the situation in Fig. 9.4. Let us assume that Server1 is running low on resources. To make more resources available, the

management console moves the Windows instance to the hypervisor on Server2.

Software defined networking (SDN) is a network architecture that has been developed to virtualize the network. SDN moves the control plane from each network device to a central network intelligence and policy-making entity called the SDN controller.

There is no conceptual difference between running a network node and running a virtual machine. Network connections can be virtualized at the level of software links between operating environments at the hypervisor level.

9.3. Organization of the computing process in a network structure

Computing processes in digital networks can be divided into three classes: centralized, decentralized and distributed.

The main difference between the centralized and the decentralized model is the distribution of functions between the parties. Figure 9.5 shows some possible options for architectures of software systems with "thin clients" of different levels of "thickness".

Web layered architecture, thin client, and ultrathin client are examples of a centralized service model.

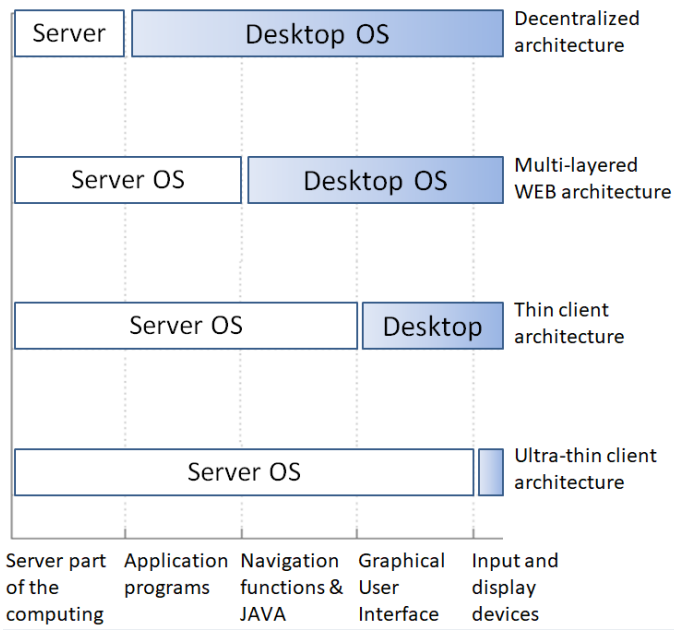


Fig. 9.5. Four levels of thickness of the "thin client"

An ultra-thin client (terminal) can only display a raster image and transmit information from input devices to the server. The Window system is implemented using the resources of the server to which the terminals are connected. This architecture places a heavy load on the server processor, which limits the number of concurrent clients. To reduce the load on the server, requests from terminals are grouped and processed in parallel.

Thin clients are devices capable of supporting the Windowing system (for example, X-Window). In this case, the amount of information transmitted to the client is significantly reduced in comparison with the previous architecture.

The next level houses Java Stations, which combine a web browser interface with the ability to download and run Java applets and standalone Java applications. This adds the ability to download programs over the network and execute them locally (on the client) to the I/O functions.

Redistribution of client functions in any of the considered architectures increases network traffic and load on server resources. The solution to this problem was found in the use of specialized devices for typical network services and the approach of these devices to clients.

Modern network equipment is designed to perform the functions of processing software systems (Fig. 9.6).

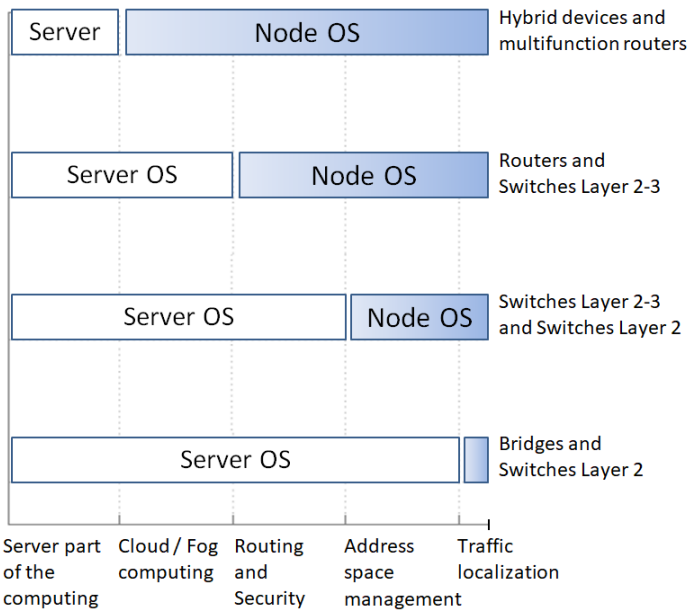


Fig. 9.6. Four levels of redistribution of server functions

The number of protocol and application functions that can be ported to network devices depends to a large extent on the support by the embedded operating systems of the selected devices for the required protocols and network services, as well as the amount of free resources (CPU, RAM, etc.). This model of resource reallocation belongs to the technology of Fog Computing.

Fog computing is a distributed computing infrastructure in which data, computation, applications and storage are hosted on a network between the data source and the cloud. Fog computing brings the features and capabilities of the cloud closer to the point where the data appears, thus enabling data localization.

9.4. The structure of the network environment of fog computing

Fog computing is part of the distributed computing infrastructure for the IoT network model, which defines an isolated segment that is closer to the network perimeter relative to the IoT device. It allows devices to access data, run the necessary code to process that data, and, in special cases, make immediate decisions. Data does not need to be sent over network connections online. The possibility of their intermediate accumulation and primary processing is provided. IoT devices are capable of working when network connections are lost, which increases fault tolerance. Sensitive data are stored within the boundaries where they are needed, which increases the level of security (Fig. 9.7).

To ensure the health of the system, it is necessary to use the components of the application support platform, which implement the infrastructure for hosting applications, primary data banks, and ensuring application mobility between cloud and fog computing environments. Devices that provide both cable and wireless connection to the network environment of IoT devices are natural intermediaries.

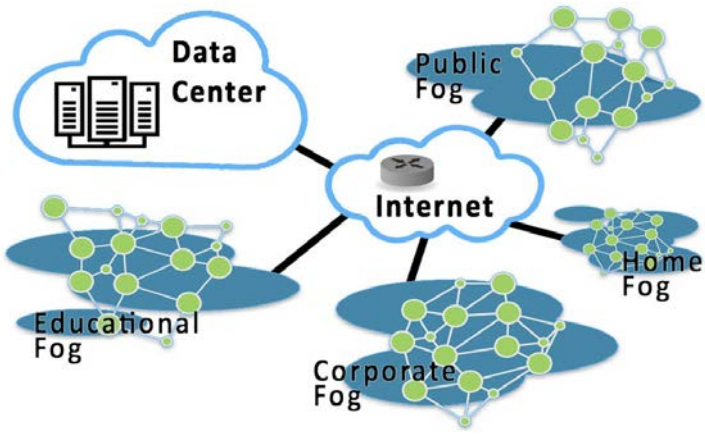


Fig. 9.7. The network model of fog computing

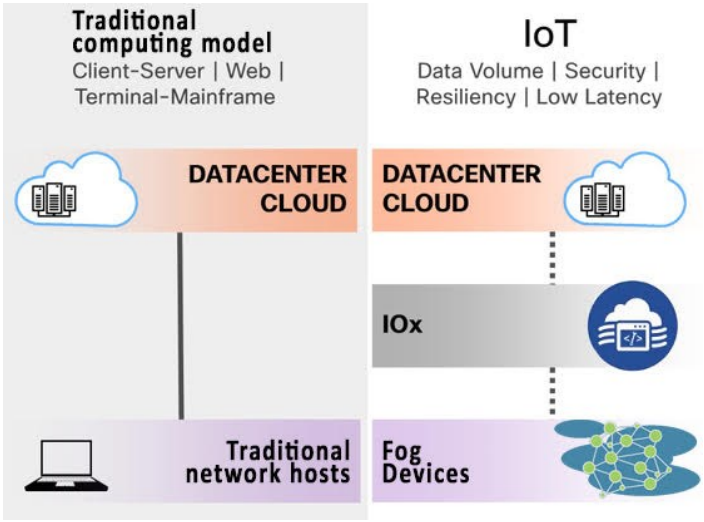


Fig. 9.8. Changing the boundaries of actions of requests in fog computing IoT

The operating system of such a network device should have a high level of versatility, since its computational power ensures the replacement of the functionality moved beyond the boundaries of IoT devices. For example, Cisco IOx is a Cisco software solution for its devices that combines Cisco IOS and Linux functionality, which allows routers to place applications near objects that these applications manage and need to get monitored, analysed and optimized (Fig. 9.8).

The number of protocol and application functions that can be ported to network devices depends to a large extent on the support by the embedded operating systems of the selected devices for the required protocols and network services as well as the amount of free resources (CPU, RAM, etc.).

9.5. Software-defined networking architecture

It is assumed that in solving practical problems, the number of IoT devices and the volume of their traffic exceed the resource capacity of communication channels on the way: network perimeter – cloud server. In this case, the number of these devices changes dynamically, both upwards and downwards, depending on the needs of the problem being solved. Consequently, settings of network connection devices should be dynamically changed, which allows combining subsets of IoT devices into a single network segment or removing them from this segment. In order for the network to have these properties, a software-defined networking architecture is used.

Software-defined networking is a data network where the network management level is separated from data transmission devices and is implemented by software. Formally, this is one of the ways to virtualize computing resources, allowing more flexibility to solve the issue of limiting access to the physical environment of data transfer.

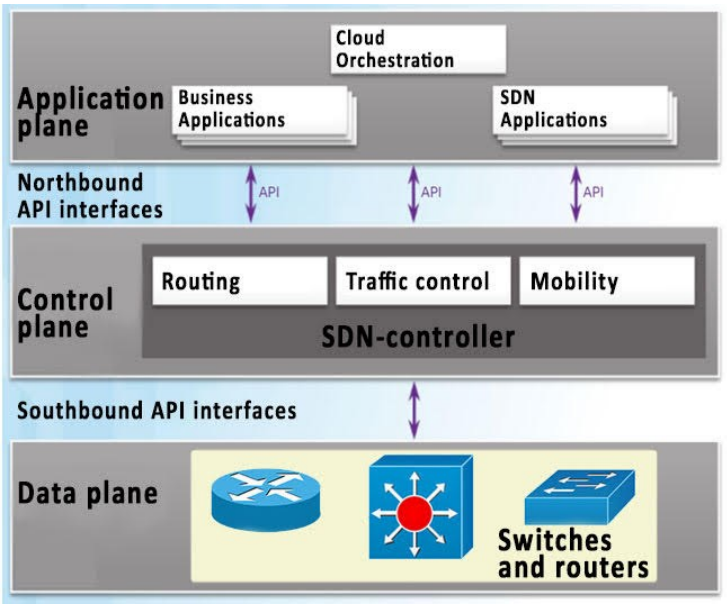


Fig. 9.9. Three-tier model of the SDN architecture

Centralized management of multiple network devices reduces the probability of an error in assigning access and reduces network maintenance time in the case of changes in security policies or communication protocols. The SDN architecture demarcates the target and control data flows that arise at the data transfer level, control level, and application level (Fig. 9.9).

Traffic localization within the data plane allows to free up the resource capacity of the CPU network device for organizing centralized control, coordinating the modes of collaboration or solving related tasks, which is the purpose of the control plane.

The decision point about authorizing access of the terminal equipment to the physical interface and assigning the characteristics of the communication channel being organized is actually transferred to the server or the nearest software controller side. For example, Cloud-Fog Middleware can become such a controller.

Cloud-Fog Middleware is deployed in the cloud and acts as middleware between the cloud and the nebula, i.e., the set of fog control nodes. It processes resource allocation tasks, query placement tasks, and transfers tasks into a specific nebula for execution. Requests with tasks processed in the cloud are delay-independent tasks requested from neighbouring nebulae. These tasks can be resource-intensive (for example, big data analysis or image processing) and, therefore, should be performed in the cloud with almost infinite resources.

9.6. Specifying the boundaries of the cloud environment

The hierarchical application service architecture allows you to combine hardware solutions, customer service software interfaces (applications), and software-implemented (virtualized) network services to improve the efficiency of service to the final equipment.

The architecture consisting of a nebula controller (Control Plane, Level 1) and multi-level fog nodes (Control Plane, Levels 2, 3, and 4) is shown in Fig. 9.10. The approach provides hierarchical optimization of network traffic when migrating services. In the figure, we demonstrate a fully connected and completely fog process, where fog nodes are integrated into the structure to implement end-user services.

Widespread local fog nodes are used in the example. These are mobile devices (Control Plane, Layer 4), where such a fog node transmits video content wirelessly from device to device (D2D Interface) for mobile devices with high and similar traffic requirements with each other for creation an D2D network.

A neighbouring fog node, for example, a base station or an access point (Control Plane, Layer 3), supports from several dozen to several hundred local fog nodes. Above them, there should be a regional fog node, for example, a base band unit or Internet Service Provider (Control Plane, Level 2), which controls coordination throughout the city. The cloud or connection to it (Control Plane, level 1) is the top of such a multi-level architecture.

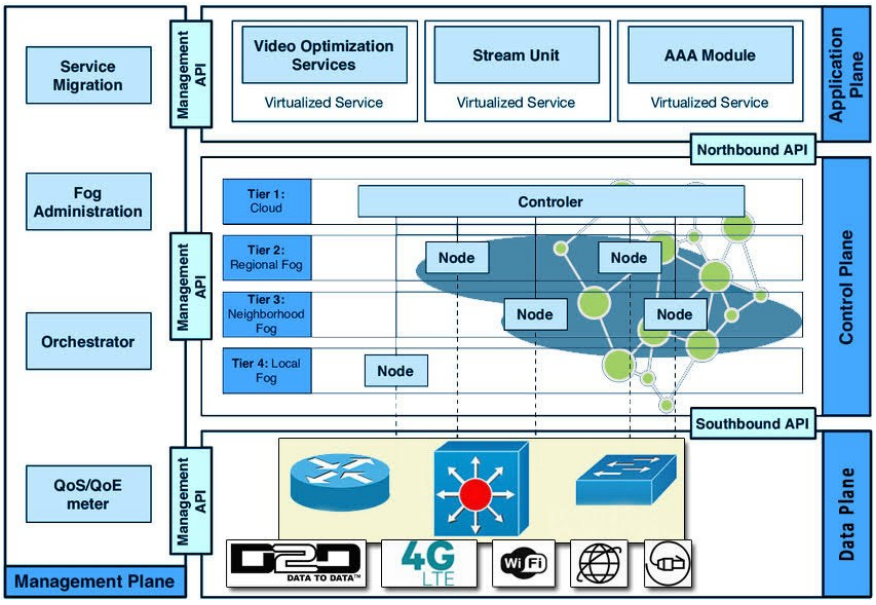


Fig. 9.10. Fog computing layered architecture

9.7. Building solutions for tunnelling IPv6 traffic in an IPv4 environment

To successfully implement the migration of virtual hosts, a stable network infrastructure is required. In the context of the procedures for transitioning the network layer addressing system to IPv6, a solution should be considered.

If the results of the Dual Stack integration experiment fail and/or the provider's traffic completely transitions to IPv6 data transfer mode (which will be a likely practice for newly created network connections), the network system administrator will be able to use tunnelling solutions to connect IPv6 network environments over the IPv4 network. An example of such transient protocol support zones is shown in Table 9.1. In the example used, a break in the continuous IPv6 addressing field (IP break) is used.

Table 9.1. Example of a Network Map with Different Levels of IPv4/IPv6 Support

Version IP	Segment 1	Segment 2	Segment 3	Segment 4			Segment 5	Segment 6
IPv4	+	+	+	+	+	+	–	–
IPv6	+	+	+	+	–	+	+	+

To solve the problems of hybrid tunnelling, it is necessary to have an effective connection between network segments using the means of the second IP-protocol. In Segment 3, it is enough to declare an IPv6 over IPv4 tunnel. The IPv4 channel is used to deliver the IPv6 traffic. According to the terms of the example, after passing the tunnel, all IPv6 clients will be able to implement full two-way communications.

It should be emphasized that when encapsulating the tunnelled traffic, the IPv4 bearer packet adds its address data on top of the transmitted IPv6 packet. Since the size of the resulting network packet (after encapsulation) should not exceed the 1500 byte limit, to do this, it is necessary on the sender's side to reduce the field of the data block by the size of the embedded service data, which in the illustrated example is 24 bytes.

Table 9.2 shows the sections available for OSPF versions to run before tunnelling is configured.

Table 9.2. Discontinuities in OSPF Protocol Versions

	Segment 1	Segment 2	Segment 3	Segment 4			Segment 5	Segment 6
OSPFv2	+	+	+	+	+	+	–	–
OSPFv3	+	+	+	+	–	+	+	+

As you can see from the figure, OSPFv2 works only in the IPv4 address zone, and OSPFv3 in the IPv6 address zone. After building the tunnels, the picture changes slightly (Table 9.3).

Table 9.3. Extension of the OSPFv3 Protocol Coverage Area

	Segment 1	Segment 2	Segment 3	Segment 4			Segment 5	Segment 6
OSPFv2	+	+	+	+	+	+	–	–
OSPFv3	+	+	+	+			+	+

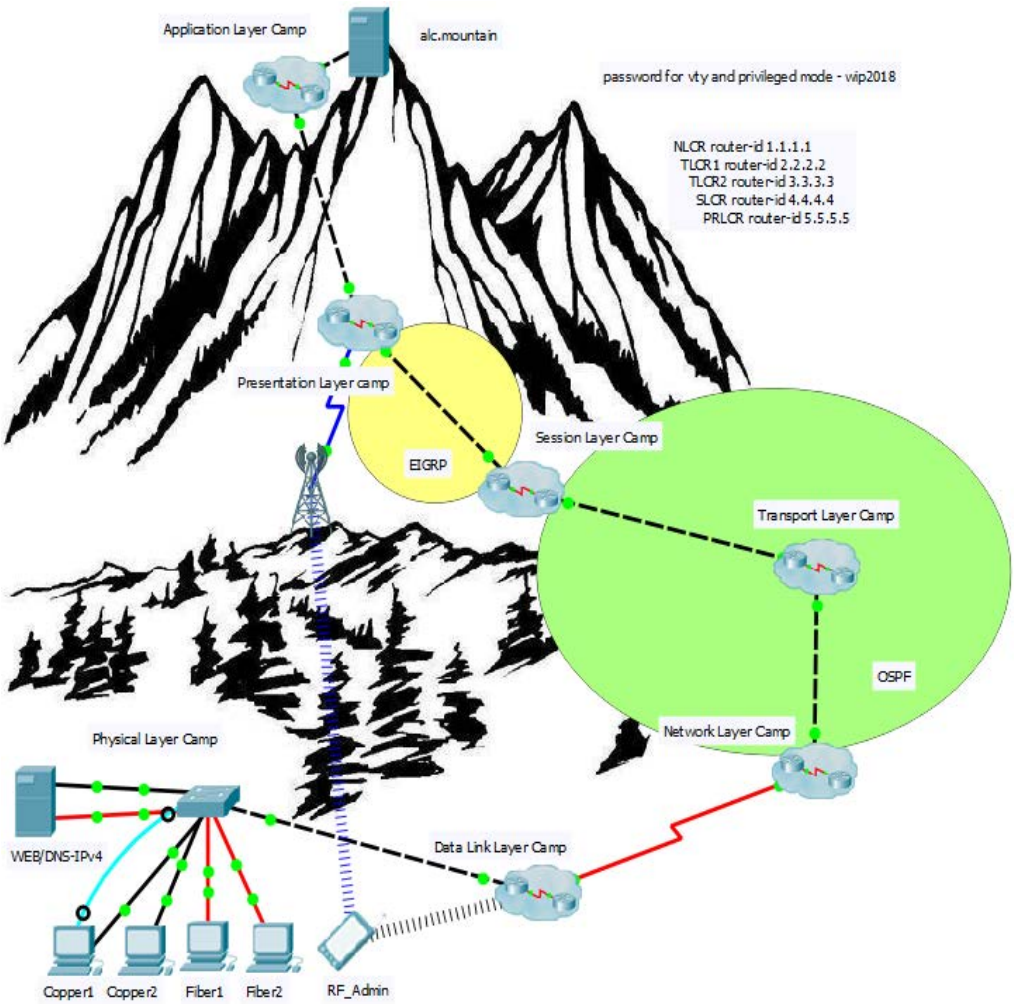


Fig. 9.11. Working window for solving the problem in the Cisco Packet Tracer environment

Since the dynamic routing protocols require the collection of overhead data about the network topology in order to find the optimal route for forwarding IP packets, the parallel operation of two versions of OSPF can increase the load on communication channels. Each OSPF protocol version sends out link-layer messages describing routers and networks, which together form a link-state database (LSDB) on each router.

It should be noted that the data for building LSDB is collected at the link level, and the main work of the protocol versions is at the network level. Therefore, it is possible to optimize service traffic.

The example considered was used as a qualifying problem for the second qualifying round of the competition "Technologies of data transmission in local and global networks" of the XI International Olympiad "IT-Planet" on March 17, 2018 (Fig. 9.11).

In this implementation mode, DualStack does not work throughout the network, but only on the side of traffic generation and maintenance.

References

[1] R. Poovendran, K. Sampigethaya, S. K. S. Gupta, I. Lee, K. V. Prasad, D. Corman, J. L. Paunicka, "Special Issue on Cyber–Physical Systems" in *Proceedings of the IEEE*, vol. 100, no. 1, pp. 6–12, 2012.

[2] https://www.nsf.gov/

[3] E. Kyriakides, M. Polycarpou (Eds.), *Intelligent Monitoring, Control, and Security of Critical Infrastructure Systems Studies in Computational Intelligence*, Springer, Berlin Heidelberg, 2015, pp. 201–316, doi:10. 1007/978- 3- 662- 44160- 2_11

[4] National Cyber Security Strategy 2016–2021, HM Government, UK.

[5] J. Brandenburger, V. Colla, S. Cateni, A. Vignali, F. Ferro, and J. Melcher (2018). *Applying Big Data Concepts to Improve Flat Steel Production Processes*. 10.1007/978-981-10-8476-8_1.

[6] V. S. Morkun, V. V. Tron, S. A. Goncharov, N. S. Podgorodetskiy, *Energy-efficient automated control of the ore concentration process with recognition of its technological varieties* - Kryvyi Rih: FOP Burovaya O. A., 2014, 326 p.

[7] N. Morkun, V. S. Morkun, V. V. Tron, "Optimal control of iron ore concentration" V – LAP LAMBERT Academic Publishing, Saarbrucken, Deutschland, 2015. 310 p.

[8] V. S. Protsuto, "Automated control systems for technological processes of concentrating factories": Nedra, 1987, 253 p.

[9] S. A. Ayvazyan, V. M. Bukhstaber, I. S. Enyukov, L. D. Meshalkin, Applied Statistics: *Classification and Dimension Reduction*: Ref. ed. / / ed. S.A. Ayvazyan. – M.: Finance and Statistics, 1989. 607 p.

[10] L. van der Maaten, E. O. Postma, H. J. van den Herik, "Dimensionality reduction: A comparative review," Technical Report TiCC TR 2009-005, [available online:] http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.125.6716&rep=rep1&type=pdf.

[11] T. Cox and M. Cox (Eds.), *Multidimensional scaling*, Chapman & Hall, London, UK, 1994. 10 p.

[12] J. B. Kruskal. *Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis*. Psychometrika. 1964. 29. pp. 1–27.

[13] Dijkstra E. W. A note on two problems in connection with graphs / E. W. Dijkstra // Numerische Mathematik. 1959. 1. pp. 269–271.

[14] S. Lafon, A. B. Lee, "Diffusion maps and coarse-graining: A unified framework for dimensionality reduction, graph partitioning, and data set parameterization," IEEE Transactions on Pattern Analysis and Machine Intelligence, 28(9), 2006. pp. 1393–1403.

[15] H. Moodi, D. Bustan, "On Identification of Nonlinear Systems Using Volterra Kernels Expansion on Laguerre and Wavelet Function," *Chinese Control and Decision Conference,* Xuzhou, China, 26–28 May 2010. pp. 1141–1145.

[16] L. Aronovich, K. Mushkin, O. Sonin, (2014). Read-ahead processing in networked client-server architecture.

[17] V. Golik, V. Komashchenko, V. Morkun (2015) *Feasibility of using the mill tailings for preparation of self-hardening mixtures*, Metallurgical and Mining Industry, No 3, pp. 38–41.

[18] A. Kupin, I. Kuznetsov (2016), *Informatsiyna tehnologiya dlya grupovoyi diagnostiki asinhronnih elektrodviguniv na osnovi spektralnih harakteristik ta intelektualnoyi klasifikatsiyi*: monografiya, FOP Chernyavskiy D. O., 200 p.

[19] M. RuEmann, M. Lorenz, P. Gerbert, M. Waldner, J. Justus, P. Engel and M. Harnisch, (2015), *Industry 4.0: The Future of Productivity and Growth in Manufacturing Industries*, The Boston Consulting Group, pp. 4–10, [available online:] https://image-src.bcg.com/Images/Industry_40_Future_of_Productivity_April_2015_tcm9-61694.pdf.

[20] Mohammed, Z., Ahmed, E. (2017), Internet of Things Applications, Challenges and Related Future Technologies, World Scientific News, №67(2), pp. 126–148.

[21] Chamberlin, B. (2016), Healthcare Internet of Things: 18 trends to watch in 2016, IBM Center for Applied Insights, Avaible at: https://ibmcai.com/2016/03/01/healthcare-internet-of-things-18-trends-to-watch-in-2016.

[22] Lutsenko, I., Fomovskaya, E., (2015) Synthesis of cybernetic structure of optimal spooler, Metallurgical and Mining Industry, No. 9, pp. 297–301.

[23] Chamberlin, B. (2016), Healthcare Internet of Things: 18 trends to watch in 2016, IBM Center for Applied Insights, Avaible at: https://ibmcai.com/2016/03/01/healthcare-internet-of-things-18-trends-to-watch-in-2016/.

[24] Gubbia, J., Buyyab, R., Marusic, S., Palaniswami, M. (2013), Internet of Things (IoT): A vision, architectural elements, and future directions, Future Generation Computer Systems, No. 29, 1645–1660.

[25] IoT connections outlook (2017), Ericsson mobility report. November 2017, 5–20, [available online:] https://www.ericsson.com/assets/local/mobility-report/documents/2017/ericsson-mobility-report-november-2017-central-and-eastern-europe.pdf.

[26] Chong, J. (2017) LG's new smart devices strengthen its IoT game, Digital News Asia, Available at: http://www.lg.com/uk/support/solutions/washingmachines / smart-thinq.

[27] Kulagin, M., Volkov, I. (2016), Promyishlennyiy internet na praktike: udalennaya diagnostika stankov s ChPU s pomoschyu tehnologii Winnum, CAD/cam/cae Observer, Issue 6 (106), 20–25.

[28] Zolfaghari, S., Noor, S., Mehrjou, M., Marhaban, M., Mariun, N. (2018), Broken Rotor Bar Fault Detection and Classification Using Wavelet Packet Signature Analysis Based on Fourier Transform and Multi-Layer Perceptron Neural Network, Applied sciences, Vo. 8, Issue 1(25), 1–21. doi: 10.3390/app8010025.

[29] Holler, J., Tsiatsis, V., Mulligan, C., Karnouskos, S., Avesand, S., Boyle, S. (2014), From Machine-to-Machine to the Internet of Things. Introduction to a New Age of Intelligence, Elsevier, 100–125. doi: https://doi.org/10.1016/B978-0-12-407684-6.00017-6.

[30] Kupin, A., Kuznetsov, D., Muzyka, I., Paraniuk, D., Serdiuk, O., Suvorov, O., Dvornikov, V. (2018). The concept of a modular cyberphysical system for the early diagnosis of energy equipment. Eastern-European Journal of Enterprise Technologies. 4. 71–79. 10.15587/1729-4061.2018.139644.

[31] Sposib diagnostuvannya elektrodviguna: pat. 81128U Ukraine. Byul. №12/25.06.2013.

[32] Kupin, A., Vdovichenko, I., Muzyka, I., Kuznetsov, D. (2017), Development of an intelligent system for the prognostication of energy produced by photovoltaic cells in smart GRID systems, Eastern-European Journal of Enterprise Technologies, No. 5/8 (89), 4–9. doi:10.15587/1729-4061.2017.112278.

[33] A. Ross, *The Industries of the future*, Simon & Schuster Publications ISBN-10: 1476753660 (Ukrainian version of Росс Алек. Індустрії майбутнього / пер. з англ. Н. Кошманенко. – К. : Наш формат), 2017.

[34] «The United States accused Russia of attacking the NotPetya virus and promised consequences,» (in Ukrainian only, as electronic news article: США звинуватили Росію в атаці вірусу NotPetya і пообіцяли наслідки), [available online:] https://www.unian.ua/politics/10009301-ssha-zvinuvatili-rosiyu-v-ataci-virusu-notpetya-i-poobicyali-naslidki.html.

[35] O. K. Yudin, S. S., Buchyk, "Analysis of threats to state information resources," (in Ukranian only as: Юдін, С.С. Бучик Аналіз загроз державним інформаційним ресурсам) in *Problems of informatization and management* (Проблеми інформатизації та управління.), 2013, No. 4 (44) pp. 93–99. [available online:] http://jrnl.nau.edu.ua/index.php/PIU/article/view/6404.

[36] X.800 : Security architecture for Open Systems Interconnection for CCITT applications**,** [available online:] https://www.itu.int/rec/T-REC-X.800-199103-I/en.

[37] *Open Systems Interconnection Basic Reference Model* OSI // [available online:] (in Ukrainian) https://uk.wikipedia.org/wiki/Мережева_модель_OSI.

[38] "The Convention of the European Council on Cybercrime of 23.11.2001 number 994, ratified by Ukraine from 07.09.2005 number 2824-IV // [electronic resource] (in Ukrainian only as: "Конвенція Ради Європи про кіберзлочинність від 23.11.2001 № 994: ратифіковано Законом України від 07.09.2005 № 2824-IV), [available online:] https://zakon.rada.gov.ua/laws/show/994_575.

[39] L. V. Palaeva, A. M. Khafizov, A. M. Gilyazetdinova, A. R. Vakhitova, K. N. Davydova., E. R. Sirotina, "The main types of cyber-attacks on automated process control systems and means of protection against them," in *Fundamental research.* No. 10–3, 2017, pp. 507–511 (in Russian) [available online:] http://www.fundamental-research.ru/ru/article/view?id=41866.

[40] «Трояни» та сніфери: якими бувають віруси та як від них позбутися? // [Електронний ресурс]. Режим доступу: https://cybercalm.org/novyny/troyany-ta-snifery-yakymy-buvayut-virusy-ta-yak-vid-nyh-pozbutysya.

[41] Розвинена стала загроза // [Електронний ресурс]. Режим доступу: https://uk.wikipedia.org/wiki/Розвинена_стала_загроза#Відомі_приклади.

[42] У Windows 10 з'явиться своя «пісочниця» // [Електронний ресурс]. Режим доступу: https://news.finance.ua/ua/news/-/440947/u-windows-10-zyavytsya-svoya-pisochnytsya.

[43] Что такое Cyber-Kill Chain и почему ее надо учитывать в стратегии защиты // [Електронний ресурс]. Режим доступу: https://habr.com/ru/company/panda/blog/327488.

[44] Експерт: кіберзахист – це не параноя // [Електронний ресурс]. Режим доступу: https://www.bbc.com/ukrainian/features-39364360.

[45] Reviews for Endpoint Detection and Response Solutions // [Електронний ресурс]. Режим доступу: https://www.gartner.com/reviews/market/endpoint-detection-and-response-solutions.

[46] ESET LiveGrid // [Електронний ресурс]. Режим доступу: https://help.eset.com/ess/10/uk-UA/idh_config_charon.html.

[47] Arbor DDoS Protection // [Електронний ресурс]. Режим доступу: https://www.netscout.com/ddos-protection.

[48] Cisco Advanced Malware Protection Solution Overview // [Електронний ресурс]. Режим доступу: https://www.cisco.com/c/en/us/solutions/collateral/enterprise-networks/advanced-malware-protection/solution-overview-c22-734228.html?dtid=osscdc000283.

[49] Полювання на кіберзагрози // [Електронний ресурс]. Режим доступу: https://uk.wikipedia.org/wiki/Полювання_на_кіберзагрози.

[50] Сетевая телеметрия Cisco против киберугроз // [Електронний ресурс]. Режим доступу: https://habr.com/ru/company/cisco/blog/229073.

[51] S. M. Rinaldi, J. P. Peerenboom, T. K. Kelly, "Identifying, understanding, and analyzing critical infrastructure interdependencies" in *IEEE Control Systems Magazine*, Vol. 21, No. 6, pp. 11–25, 2001.

[52] J. Clarke, "Managing risks to energy sector critical infrastructure" in P.E. Cornel (Ed.): *Energy Security and Security Policy: NATO and the Role of International Security Actors in Achieving Energy Security*, Series: NATO School Research, pp. 59–64, Nov. 2007.

[53] R. Setola, V. Rosato, E. Kyriakides, E. Rome (Eds.): *Managing the Complexity of Critical Infrastructures A Modelling and Simulation Approach*, Series: Studies in Systems, Decision and Control, Vol. 90, Springer, ISBN 978-3-319-51042-2, 2017.

[54] J. E. Contreras-Ocaña, Y. Chen, U. Siddiqi, B. Zhang, "Non-Wire Alternatives: An Additional Value Stream for Distributed Energy Resources," in *IEEE Transactions on Sustainable Energy*, Vol. 11, No. 3, pp. 1287–1299, July 2020.

[55] A. Kotsonias, M. Asprou, L. Hadjidemetriou, E. Kyriakides, "State estimation for distribution grids with a single point grounded neutral conductor" in *IEEE Trans. Instrumentation and Measurement*, Vol. 69, No. 10, pp. 8167–8177, Oct. 2020.

[56] A. Kotsonias, L. Hadjidemetriou, E. Kyriakides, "Power flow for a four-wire radial low voltage distribution grid with a single point grounded neutral" in *Proc. IEEE ISGT*, Bucharest, Romania, 2019, pp. 1–5.

[57] A. Charalambous, L. Hadjidemetriou, L. Zacharia, A. Bintoudi, A. Tsolakis, D. Tzovaras, E. Kyriakides, "Phase balancing and reactive power support services for microgrids" in *Applied Sciences*, 9(23), 5067, pp. 1–17, 2019.

[58] L. Hadjidemetriou, A. Charalambous, E. Kyriakides, "Control scheme for phase balancing low-voltage distribution grids," in *Proc. IEEE SEST*, Porto, Portugal, 2019, pp. 1–6.

[59] A. Safayet, P. Fajri, I. Husain, "Reactive power management for overvoltage prevention at high PV penetration in low voltage distribution system" in *IEEE Trans. Ind. Appl*., Vol. 53, No. 6, pp. 5786–5794, 2017.

[60] E. Demirok, P. C. González, K. H. B. Frederiksen, D. Sera, P. Rodriguez, R. Teodorescu, "Local Reactive Power Control Methods for Overvoltage Prevention of Distributed Solar Inverters in Low-Voltage Grids," in *IEEE Journal of Photovoltaics*, Vol. 1, No. 2, pp. 174–182, Oct. 2011.

[61] L. Hadjidemetriou, E. Kyriakides, and F. Blaabjerg, "A robust synchronization to enhance the power quality of renewable energy systems," in *IEEE Trans. Industrial Electronics*, vol. 62, no. 8, pp. 4858-4868, Aug. 2015.

[62] A. Charalambous, L. Hadjidemetriou, E. Kyriakides, M. Polycarpou, "Voltage and frequency support scheme for storage systems in distribution grids" in *IEEE Trans. Power Electronics*, pp. 1–8, Dec. 2020.

[63] L. Tziovani, L. Hadjidemetriou, C. Charalampous, S. Timotheou, E. Kyriakides, "Modelling and energy management of a flywheel storage system for peak shaving applications" in *Proc. IEEE ISGT*, Hague, 2020, pp. 1–5.

[64] A. Sangwongwanich et al., "Reliability Assessment of PV Inverters with Battery Systems Considering PV Self-Consumption and Battery Sizing" in *2018 IEEE ECCE*, Portland, OR, 2018, pp. 7284–7291.

[65] L. Tziovani, P. Kolios, L. Hadjidemetriou, E. Kyriakides, "Grid friendly operation with profit and reliability maximization of a hybrid photovoltaic-storage system" in *Proc. IEEE SEST*, Porto, Portugal, 2019, pp. 1–6.

[66] B. Sun, Z. Huang, X. Tan, D. H. K. Tsang, "Optimal Scheduling for Electric Vehicle Charging with Discrete Charging Levels in Distribution Grid" in *IEEE Transactions on Smart Grid*, Vol. 9, No. 2, pp. 624–634, March 2018.

[67] A. G. Phadke, T. S. Thorp, *Synchronized Phasor Measurements and Their Applications*. R2: Analysis of data quality issues in WAMC systems. New York: Springer, 2008.

[68] A. J. Roscoe, B. Dickerson, K. E. Martin, "The amended standard C37.118.1a and its implications for frequency-tracking m-class Phasor Measurement Units (PMUs)" in *IEEE AMPS 2014*, Aachen, pp. 1–6, Sept. 2014.

[69] S. Kirti, Z. Wang, A. Scaglione, R. Thomas, "On the Communication Architecture for Wide-Area Real-Time Monitoring in Power Networks" in the *40th Annual Hawaii International Conference on System Sciences* (HICSS'07), Waikoloa, HI, pp. 119–129, 2007.

[70] B. Naduvathuparambil, M. C. Valenti, A. Feliachi, "Communication delays in wide area measurement systems" in the *Thirty-Fourth Southeastern Symposium on System Theory*, Huntsville, AL, USA, pp. 118–122, 2002.

[71] D. Dotta, A. S. e Silva, I. C. Decker, "Wide-area measurements-based two-level control design considering signal transmission delay," in *IEEE Trans. Power Systems*, Vol. 24, No. 1, pp. 208–216, 2009.

[72] B. Chaudhuri, B. C. Pal, "Robust damping of multiple swing modes employing global stabilizing signals with a TCSC," in *IEEE Trans. Power Systems*, Vol. 19, No. 1, pp. 499–506, 2004.

[73] H. Wu, K. S. Tsakalis, G. T. Heydt, "Evaluation of time delay effects to wide-area power system stabilizer design," in *IEEE Trans. Power Systems*, Vol. 19, No. 4, pp. 1935–1941, 2004.

[74] I. Kamwa, R. Grondin, Y. Hebert, "Wide-area measurement based stabilizing control of large power systems-a decentralized/hierarchical approach," in *IEEE Trans. Power Systems*, Vol. 16, No. 1, pp. 136–153, 2001.

[75] T. Surinkaew, I. Ngamroo, "Hierarchical co-ordinated wide area and local controls of DFIG wind turbine and PSS for robust power oscillation damping," in *IEEE Trans. Sustainable Energy*, Vol. 7, No. 3, pp. 943–955, 2016.

[76] M. Beiraghi, A. M. Ranjbar, "Additive model decision tree-based adaptive wide-area damping controller design," in *IEEE Systems Journal*, Vol. 12, No. 1, pp. 328–339, 2018.

[77] F. Okou, L.-A. Dessaint, O. Akhrif, "Power system stability enhancement using a wide area signals based hierarchical controller," in *IEEE Trans. Power Systems*, Vol. 20, No. 3, pp. 1465–1477, 2005.

[78] M. E. Raoufat, K. Tomsovic, S. M. Djouadi, "Virtual actuators for wide-area damping control of power systems," in *IEEE Trans. Power Systems*, Vol. 31, No. 6, pp. 4703–4711, 2016.

[79] D. Ke, C. Y. Chung, "Design of probabilistically-robust wide-area power system stabilizers to suppress inter-area oscillations of wind integrated power systems," in *IEEE Trans. Power Systems*, Vol. 31, No. 6, pp. 4297–4309, 2016.

[80] J. Lian, R. Huang, S. Wang, R. Fan, M. A. Elzondo, H. Kirkham, J. Hansen, L.D. Marinovici, D. Schoenwald, F. Wilches-Bernal, "Universal Wide-area Damping Control for Mitigating Inter-area Oscillations in Power Systems," in *Pacific Northwest National Laboratory*, Washington, 2017.

[81] D. Gautam, V. Vittal, T. Harbour, "Impact of increased penetration of DFIG-based wind turbine generators on transient and small signal stability of power systems," in *IEEE Trans. Power Systems*, Vol. 24, No. 3, pp. 1426–1434, 2009.

[82] D. A. Halamay, T. K. A. Brekken, A. Simmons, S. McArthur, "Reserve requirement impacts of large-scale integration of wind, solar, and ocean wave power generation," in *IEEE Trans. Sustainable Energy*, Vol. 2, No. 3, pp. 321–328, 2011.

[83] C. Liu, G. Cai, W. Ge, D. Yang, C. Liu, Z. Sun, "Oscillation analysis and wide-area damping control of DFIGs for renewable energy power systems using line modal potential energy," in *IEEE Trans. Power Systems*, Vol. 33, No. 3, pp. 3460–3471, 2018 .

[84] Y. Wang, C. Lim, P. Shi, "Adaptively adjusted event-triggering mechanism on fault detection for networked control systems," in *IEEE Trans. Cybernetics*, Vol. 47, No. 8, pp. 2299–2311, 2017.

[85] B. J. Pierre et al., "Design of the Pacific DC Intertie wide area damping controller," in *IEEE Trans. Power Systems*, Vol. 34, No. 5, pp. 3594–3604, 2019.

[86] F. Wilches-Bernal et al., "Time delay definitions and characterization in the pacific DC intertie wide area damping controller," in *Proc. IEEE PES General Meeting*, Chicago, 2017.

[87] M. Long, C. Hwa and J.Y. Hung, "Denial of service attacks on network-based control systems: impact and mitigation," in *IEEE Trans. Industrial Informatics*, Vol. 1, No. 2, pp. 85–96, 2005.

[88] P. Demetriou, M. Asprou, J. Quiros-Tortos, E. Kyriakides, "Dynamic IEEE test systems for transient analysis," in *IEEE Systems Journal*, Vol. 11, No. 4, pp. 2108–2117, 2017.

[89] P. M. Anderson, A. A. Fouad, *Power System Control and Stability*, 2nd Edition, Wiley-IEEE Press, 2002.

[90] C. Dufour, J. Bélanger, "On the use of real-time simulation technology in smart grid research and development," in *IEEE Trans. Industry Applications*, Vol. 50, No. 6, pp. 3963–3970, 2014.

[91] M. S. Almas, L. Vanfretti, "RT-HIL implementation of the hybrid synchrophasor and GOOSE-based passive islanding schemes," in *IEEE Trans. Power Delivery*, Vol. 31, No. 3, pp. 1299–1309, 2016.

[92] L. Bottaccioli et al., "A flexible distributed infrastructure for real-time cosimulations in smart grids," in *IEEE Trans. Industrial Informatics*, Vol. 13, No. 6, pp. 3265–3274, 2017.

[93] O.-R. Technologies, "Hardware-In-the-Loop," OPAL-RT Technologies, 2019. [Online]. Available: https://www.opal-rt.com/hardware-in-the-loop/. [Accessed 23 10 23].

[94] O.-R. Technologies, "Power Hardware-In-the-Loop," OPAL-RT Technologies, 2019. [Online]. Available: https://www.opal-rt.com/power-hardware-in-the-loop/. [Accessed 23 10 2019].

[95] L. Zacharia, *Development of Robust and Effective Coordination Methodologies for Power Systems*, Nicosia: University of Cyprus, 2020.

[96] L. Deka, M. Chowdhury "Transportation cyber-physical systems," *Elsevier Science,* pp. 2–3, 2018.

[97] Shi Jianjuna, Wu Xub, G. Jizhenc, Ch. Yangzhoua, "The analysis of traffic control cyber-physical systems," in 2013 13th COTA International Conference of Transportation Professionals (CICTP 2013), Shenzhen, China, 2013.

[98] "FLOW solution overview," DataFromSky, pp. 4–11, 2020.

[99] www.datafromsky.com

[100] DataFromSky Viewer, user guide, pp. 4–15.

[101] www.goodvisionlive.com

[102] "Vyznachennia intensyvnosti rukhu ta skladu transportnoho potoku," *in Ukranian*, "Determination of traffic frequency and traffic stream mix," DSTU 8824:2019, pp. 17–23, 2019.

[103] NIST. SP. 1500-201, Framework for Cyber-Physical Systems, Vol. 1 (2017), p. 79. DOI: https://doi.org/10.6028/NIST.SP.1500-201.

[104] Ahmadzai Ahmadi, Chantal Cherifi, Vincent Cheutet, Yacine Ouzrout. A Review of CPS 5 Components Architecture for Manufacturing Based on Standards. 11th IEEE International Conference on Software, Knowledge, Information Management and Applications (SKIMA 2017), Dec 2017, Colombo, Sri Lanka. 6 p.

[105] Madhan, E. S., Uttam Ghosh, Deepak K. Tosh, K. Mandal, E. Murali, and Soumalya Ghosh. "An improved communications in cyber physical system architecture, protocols and applications." In *2019 16th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*, pp. 1–6. IEEE, 2019.

[106] Jawhar, Imad, Jameela Al-Jaroodi, Hassan Noura, and Nader Mohamed. "Networking and Communication in Cyber Physical Systems." In *2017 IEEE 37th International Conference on Distributed Computing Systems Workshops (ICDCSW)*, pp. 75–82. IEEE, 2017.

[107] Dey, Nilanjan, Amira S. Ashour, Fuqian Shi, Simon James Fong, and João Manuel RS Tavares. "Medical cyber-physical systems: A survey." *Journal of medical systems* 42, No. 4 (2018): 1–13.

[108] Li, Tao, Feng Tan, Qixin Wang, Lei Bu, Jian-Nong Cao, and Xue Liu. "From offline toward real-time: A hybrid systems model checking and CPS co-design approach for medical device plug-and-play (MDPnP)." In *2012 IEEE/ACM Third International Conference on Cyber-Physical Systems*, pp. 13–22. IEEE, 2012.

[109] Mohanty, Saraju P. "Advances in Transportation Cyber-Physical System (T-CPS)." *IEEE Consumer Electronics Magazine* 9, No. 4 (2020): 4–6.

[110] Vierhauser, Michael, Jane Cleland-Huang, Sean Bayley, Thomas Krismayer, Rick Rabiser, and Pau Grünbacher. "Monitoring CPS at runtime-A case study in the UAV domain." In *2018 44th Euromicro Conference on Software Engineering and Advanced Applications (SEAA)*, pp. 73–80. IEEE, 2018.

[111] Zheng, Yu, Guanxue Wang, Zhongxiang Chen, Yan Liu, and Xiong Shen. "A finite state machine based diagnostic expert system of large-scale autonomous unmanned submarine." In *2018 IEEE 8th International Conference on Underwater System Technology: Theory and Applications (USYS)*, pp. 1–6. IEEE, 2018.

[112] Hamdaoui, Youssef and Abdelilah Maach. "A cyber-physical power distribution management system for smart buildings." In *Proceedings of the Mediterranean Symposium on Smart City Applications*, pp. 538–550. Springer, Cham, 2017.

[113] Guo, Ping, Puwadol Oak Dusadeerungsikul, and Shimon Y. Nof. "Agricultural cyber physical system collaboration for greenhouse stress management." *Computers and electronics in agriculture* 150 (2018): 439–454.

[114] de Carvalho, Rafael Viana, Cláudio de Oliveira e Silva, Arlindo Rodrigues Galvão Filho, Filipe de Souza Lima Ribeiro, Leonardo Brodbeck Chaves, and Clarimar José Coelho. "Cyber-phiscial Systems with Petri Nets to Model Hydropower Control." In *2020 International Conference on Electrical, Communication, and Computer Engineering (ICECCE)*, pp. 1–6. IEEE, 2020.

[115] Rodrigues, Joel JPC, and Amjad Gawanmeh, eds. *Cyber-Physical Systems for Next-Generation Networks*. IGI Global, 2018.

[116] Capatina, D., I. Stoian, T. Sanislav, O. Ghiran, E. Stancel, and I. Filip. "Integration techniques of the embedded distributed systems using programming environments and industrial standard communication protocols." In *2006 IEEE International Conference on Automation, Quality and Testing, Robotics*, Vol. 1, pp. 430–435. IEEE, 2006.

[117] Moiş, George, Silviu Folea, Teodora Sanislav, and Liviu Miclea. "Communication in cyber-physical systems." In *2015 19th International Conference on System Theory, Control and Computing (ICSTCC)*, pp. 303–307. IEEE, 2015.

[118] Huang, Pei, Li Xiao, Soroor Soltani, Matt W. Mutka, and Ning Xi. "The evolution of MAC protocols in wireless sensor networks: A survey." *IEEE communications surveys & tutorials* 15, No. 1 (2012): 101–120.

[119] Woolley, Martin. "Bluetooth Core Specification Version 5.2 Feature Overview." *Bluetooth SIG: Kirkland, WA, USA* (2020).

[120] WiFi Alliance, Generational Wi-Fi® User Guide. October 2018.

[121] Oughton, Edward J., William Lehr, Konstantinos Katsaros, Ioannis Selinis, Dean Bubley, and Julius Kusuma. "Revisiting wireless internet connectivity: 5G vs Wi-Fi 6." *Telecommunications Policy* 45, No. 5 (2021): 102127.

[122] Qiao, Lei, Zhe Zheng, Wenpeng Cui, and Liang Wang. "A survey on Wi-Fi HaLow technology for Internet of Things." In *2018 2nd IEEE Conference on Energy Internet and Energy System Integration (EI2)*, pp. 1–5. IEEE, 2018.

[123] Wang, Tao, Hong Xiao, and Lianglun Cheng. "Sensing as a Service-A Service-Oriented Collaborative Sensing Framework for Detecting Composite Events in Industrial Cyber-Physical System.", In Proceedings of the International Conference on Advances in Computer Technology, Information Science and Communications (CTISC 2019), pp. 264–271, SCITEPRESS, 2019.

[124] Eugen Pop and Daniela Gîfu. "A Cyber-Physical Systems Oriented Platform Using Web Services." In *2019 22nd International Conference on Control Systems and Computer Science (CSCS)*, pp. 602–609. IEEE, 2019.

[125] Pedapudi, Anupam Vamsi, Sai Teja Kurapati, Gayathri Narayanan, and Dhanesh G. Kurup. "Performance Analysis of Sensor Communications for Agriculture Systems using SDR Platform." In *2020 5th IEEE International Conference on Recent Advances and Innovations in Engineering (ICRAIE)*, pp. 1–6. IEEE, 2020.

[126] Rawat, Danda B. and Chandra Bajracharya. *Vehicular cyber physical systems*. Cham: Springer, 2017.

[127] Lee, Jay, Moslem Azamfar, Jaskaran Singh, and Shahin Siahpour. "Integration of digital twin and deep learning in cyber-physical systems: towards smart manufacturing." *IET Collaborative Intelligent Manufacturing* 2, No. 1, 2020, pp. 34–36.

[128] Edward A. Lee, Sanjit A. Seshia. Introduction to Embedded Systems, A Cyber-Physical Systems Approach. 2011. http://LeeSeshia.org.

[129] Patricia Derler, Edward A. Lee, and Alberto Sangiovanni Vincentelli "Modeling Cyber-Physical Systems".

[130] Strassburger S., "On the Role of Simulation and Simulation Standards in Industry 4.0," Simulation Innovation Workshop (SIW), February 11–15, 2019, Florida Hotel & Conference Center at the Florida Mall, Orlando, FL. – Orlando, Fla.: SISO, Simulation Interoperability Standards Organization, p. 12.

[131] https://www.tutorialspoint.com/embedded_systems/es_overview.htm

[132] https://intuit.ru/studies/courses/10620/1104/lecture/24050

[133] https://www.researchgate.net/publication/331248990_On_the_Role_of_Simulation_and_Simulation_Standards_in_Industry_40

[134] https://www.avsystem.com/blog/what-is-iot-architecture/

[135] High Level Architecture (HLA), Release 3.0, AIOTI WG03 – IoT Standardisation. European Communities, 2017, p. 41.

[136] Nutt G.J., "Evaluation Nets for Computer Systems Performance Analysis." FJCC, AFIPS PRESS, Vol. 41, 1972, pp. 279–286.

[137] Kazymyr V., V. Prila O., Usik A., Sysa D., "New paradigm of model-oriented control in IoT," Information and Software Technologies. In: 25th International Conference, ICIST 2019, Vilnius, Lithuania, October 10–12, 2019, Proceedings. pp. 605–614.

[138] https://coderlessons.com/tutorials/kompiuternoe-programmirovanie/vstraevaemye-sistemy/vstraevaemye-sistemy#5

[139] Kazimir V. V. The Embedded model system EMS / V. V. Kazimir, G. A. Sira, I. I. Musketeer // Bulletin of the Chernigiv State Technological University, Chernigiv, 2011. No. 3 (51). pp. 144–153.

[140] Kazimir V. V. Modeling of synthetic environment for system security assessment / V. V. Kazimir, G. A. Sira // Sixth scientific-practical conference with international participation "Mathematical and simulation modeling of MODS'2011 systems" Abstracts. Chernihiv - 2011. pp. 415–419.

[141] Sira G. A. Dynamic verification of information systems based on simulation models of synthetic environment / G. A. Sira // International Scientific Conference "Intelligent Decision Making Systems and Problems of Computational Intelligence ISDMCI'2011". Abstracts. Evpatoria, 2011. Vol. 1, pp. 113–117.

[142] Feller V. Introduction to simulation and the language of SLAM-II [Text] / V. Feller // M .: Mir, 1987, 738 p.

[143] Description of the markup language Extensible Markup Language (XML) [Electronic resource] // Access mode: http://www.w3.org/XML/

[144] The site dedicated to the JSON format [Electronic resource] // Access mode: http://www.json.org/

[145] SQLite library website [Electronic resource] // Access mode: http://www.sqlite.org/

[146] Zamyatina Ye.B. Modern theories of simulation. Special course / E. B. Zamyatin // Perm State University. Textbook, 2007, p. 119.

[147] http://www.pnml.org/

[148] https://www.arm.com/why-arm

[149] https://www.edgefx.in/arm-microcontroller-architecture-and-its-programming/

[150] https://zooterra-msk.ru/en/cattle/processory-arm-osobennosti-arhitektury-otlichiya-i-perspektivy-vse/

[151] Zlatanov N., "ARM Architecture and RISC Applications," 2016.

[152] Dinesh M., Saravanan P., "FPGA based real time monitoring system for agricultural field," International Journal of Electronics and Computer Science Engineering, Vol. 1, 2006, pp. 1514–1519.

[153] Palagin A., Yakovlev Y., "Design Features of Computer Systems on an FPGA Crystal," Mathematical Machines and Systems, No. 2, 2017, pp. 3–14.

[154] Kalachev A., "Multi-core configurable computing platform Zynq-7000," Modern electronics. No. 1, 2013, pp. 22–31.

[155] http://www.uco.es/~ff1mumuj/h_intro.htm

[156] R. T. Aljardiri, L. Y. Taha, P. Ivey, Electrostatic Energy Harvesting Systems: A Better Understanding of Their Sustainability, Journal of Clean Energy Technologies, Vol. 5, No. 5, September 2017.

[157] M. H. Alsharif, S. Kim, N. Kuruoglu, Energy Harvesting Techniques for Wireless Sensor Networks/Radio-Frequency Identification: A Review, Symmetry, MDPI, Vol. 11, Issue 7, 865, doi:10.3390/sym11070865, www.mdpi.com/journal/symmetry , July 2019.

[158] P. Ballon, Smart Cities: hoe technologie onze steden leefbaar houdt en slimmer maakt, Lannoo Campus, Tielt, Belgium, Nov. 2016, ISBN 978-94-014-2938-2 (in Dutch).

[159] S. Boisseau, G. Despesse, B. Ahmed Seddik, Electrostatic Conversion for Vibration Energy Harvesting, Small-Scale Energy Harvesting, Intech, 2012.

[160] A. Ciuffoletti, Low-Cost IoT: A Holistic Approach (pp. 6–24): in: B. Kantarci, S. Oktug (eds.), Wireless Sensor and Actuator Networks for Smart Cities, printed edition of the Special Issue Published in Journal of Sensor and Actuator Networks, MDPI, Basel, Switzerland, 2018, ISBN 978-3-03897-424-6, www.mdpi.com/journal/jsan.

[161] D. Enescu, Thermoelectric Energy Harvesting: Basic Principles and Applications, open access peer reviewed chapter, 2019, DOI: 10.5772/intechopen.83495 (downloaded from: https://www.intechopen.com/books/green-energy-advances/thermoelectric-energy-harvesting-basic-principles-and-applications).

[162] A. Hilmani, A. Maizate, L. Hassouni, Designing and Managing a Smart Parking System Using Wireless Sensor Networks (pp. 25 - 44): in: B. Kantarci, S. Oktug (eds.), Wireless Sensor and Actuator Networks for Smart Cities, printed edition of the Special Issue Published in Journal of Sensor and Actuator Networks, MDPI, Basel, Switzerland, 2018, ISBN 978-3-03897-424-6, www.mdpi.com/journal/jsan.

[163] C.A. Howells, Piezoelectric energy harvesting, Energy Conversion and Management, Elsevier, Vol. 50, 2009, pp. 1847–1850.

[164] B. Kantarci, S. Oktug, Special Issue: Wireless Sensor and Actuator Networks for Smart Cities (pp. 1 – 5): in: B. Kantarci, S. Oktug (eds.), Wireless Sensor and Actuator Networks for Smart Cities, printed edition of the Special Issue Published in Journal of Sensor and Actuator Networks, MDPI, Basel, Switzerland, 2018, ISBN 978-3-03897-424-6, www.mdpi.com/journal/jsan.

[165] J. Morley, K. Widdicks, M. Hazas, Digitalisation and data demand: The impact of Internet traffic on overall and peak electricity consumption, Energy Research & Social Science, Elsevier, Vol. 38, pp. 128–137, 2018.

[166] A. V. Pedchenko, E. J. Barth, Broad frequency vibration energy harvesting control approach based on the maximum power transfer theorem, ASME 2013 Dynamic Systems and Control Conference, Palo Alto, California, USA, October 21–23, 2013, paper DSCC 2013-3981.

[167] S. S. Rao, Vibration of continuous systems, John Wiley & sons, Hoboken, New Jersey, 2007, ISBN 978-0-471-77171-5.

[168] D. Steingart, Power Sources for Wireless Sensor Networks (pp. 267–286): in: S. Priya, D.J. Inman (eds.), Energy Harvesting Technologies, Springer Science + Business Media, New York, 2009, doi 10.1007/978-0-387-76464-1_9, ISBN 978-0-387-76463-4.

[169] E. O. Torres, G. A. Rincon-Mora, Energy budget and high gain strategies for voltage constrained electrostatic harvesters, IEEE International Symposium on Circuits and Systems (ISCAS), May 24–27, 2009, Taipei, Taiwan.

[170] C. B. Williams, R. B. Yates, Analysis of a micro-electric generator for microsystems, Sensors and Actuators, Vol. 52, 1996, pp. 8–11.

[171] https://www.cisco.com/c/en/us/index.html.