5

Communication and Data Management in Industry 4.0

Maria del Carmen Lucas-Estañ[1], Theofanis P. Raptis[2], Miguel Sepulcre[1], Andrea Passarella[2], Javier Gozalvez[1] and Marco Conti[2]

[1]UWICORE Laboratory, Universidad Miguel Hernández de Elche (UMH), Elche, Spain
[2]Institute of Informatics and Telematics, National Research Council (CNR), Pisa, Italy
E-mail: m.lucas@umh.es; theofanis.raptis@iit.cnr.it; msepulcre@umh.es; andrea.passarella@iit.cnr.it; j.gozalvez@umh.es; marco.conti@iit.cnr.it

The Industry 4.0 paradigm alludes to a new industrial revolution where factories evolve towards digitalized and networked structures where intelligence is spread among the different elements of the production systems. Two key technological enablers to achieve the flexibility and efficiency sought for factories of the future are the communication networks and the data management schemes that will support connectivity and data distribution in Cyber-Physical Production Systems. Communications and data management must be built upon a flexible and reliable architecture to be able to efficiently meet the stringent and varying requirements in terms of latency, reliability and data rates demanded by industrial applications. To this aim, this chapter presents a hierarchical communications and data management architecture, where decentralized and local management decisions are coordinated by a central orchestrator that ensures the efficient global operation of the system. The defined architecture considers a multi-tier organization, where different management strategies can be applied to satisfy the different requirements in terms of latency and reliability of different industrial applications. The use of virtualization and softwarization technologies as RAN Slicing and

Cloud RAN will allow to achieve the flexibility, scalability and adaptation capabilities required to support the high-demanding and diverse industrial environment.

5.1 Introduction

In future industrial applications, the Internet of Things (IoT) with its communications and data management functions will help shape the operational efficiency and safety of industrial processes through integrating sensors, data management, advanced analytics, and automation into a mega-unit [1]. The future and significant participation of intelligent robots will enable effective and cost-efficient production, achieving sustainable revenue growth. Industrial automation systems, emerging from the Industry 4.0 paradigm, count on sensors' information and the analysis of such information [2]. As such, connectivity is a crucial factor for the success of industrial Cyber-Physical-Systems (CPS), where machines and components can talk to one another. Moreover, in the context of Industry 4.0 and to match the increased market demand for highly customized products, traditional pilot lines designed for mass production are now evolving towards more flexible "plug & produce" modular manufacturing strategies based on autonomous assembly stations [3], which will make increased use of massive volumes of Big Data streams to support self-learning capabilities and will demand real-time reactions of increasingly connected mobile and autonomous robots and vehicles. While conventional cloud solutions will be definitely part of the picture, they will not be enough. The concept of centrally organized enterprises at which large amounts of data are sent to a remote data center do not deliver the expected performance for Industry 4.0 scenarios and applications. Recently, moving service supply from the cloud to the edge has enabled the possibility of meeting application delay requirements, improves scalability and energy efficiency, and mitigates the network traffic burden. With these advantages, decentralized industrial operations can become a promising solution and can provide more scalable services for delay-tolerant applications [4].

Two technological enablers of Industry 4.0 are: (i) the communication infrastructure that will support the ubiquitous connectivity of Cyber-Physical Production Systems (CPPS) and (ii) the data management schemes built upon the communication infrastructure that will enable efficient data distribution within the Factories of the Future [5]. In the industrial environment, a wide set of applications and services with very different communication requirements will coexist, being one of the most demanding verticals with respect to the number of connected nodes, ultra-low latencies, ultra-high reliability, energy efficiency, and ultra-low communication costs [6]. The varying and stringent communication and data availability requirements of the industrial applications pose an important challenge for the design of the communication network and of the data management systems. The communication network and the data management strategy must be built upon a flexible architecture capable of meeting the communication requirements of the industrial applications, with particular attention on time-critical automation.

The architecture reviewed in this chapter is the reference communications and data management architecture of the H2020 AUTOWARE project [7]. The main objective of AUTOWARE is to build an open consolidated ecosystem that lowers the barriers of small, medium- & micro-sized enterprises (SMMEs) for cognitive automation application development and application of autonomous manufacturing processes. Communications and data management are two technological enablers within the AUTOWARE Framework (Figure 5.1 and presented in detail in Chapter 2). Within the AUTOWARE framework, the AUTOWARE Reference Architecture establishes four layers: Enterprise, Factory, Workcell/Production Line, and Field Devices. In addition, the AUTOWARE Reference Architecture also includes two transversal layers: (i) the Fog/Cloud layer, since applications or services

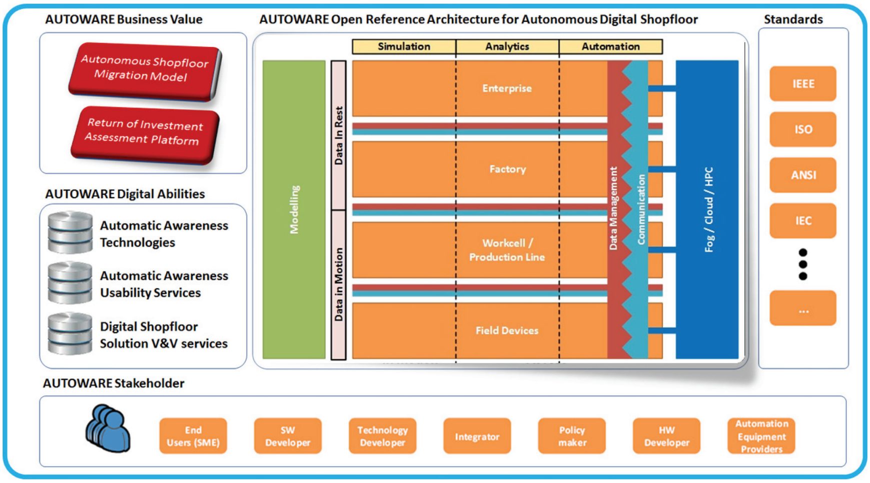


Figure 5.1 The AUTOWARE framework.

in all the layers can be included or implemented in the Fog/Cloud, and (ii) the Modelling layer, since different technical components inside the different layers can be modelled, and it could be possible to have modeling approaches that take the different layers into account. The communications and data management architecture proposed in AUTOWARE supports the communication network and the data management system and enables the data exchange between the different AUTOWARE components, exploiting the Fog and/or Cloud concepts. It provides communication links between devices, entities, and applications implemented in different layers, and also within the same layer. Within the AUTOWARE Reference Architecture (defined in the H2020 AUTOWARE Project), the communication network and data management system can be represented as a transversal layer that interconnects all the functional layers of the AUTOWARE Reference Architecture (see Figure 5.2). The communications and data management architecture presented in this chapter provides the communication and data distribution capabilities required by the different systems or platforms developed within the AUTOWARE framework.

AUTOWARE proposes the use of a heterogeneous network that integrates different communication technologies covering the industrial environment. The objective is to exploit the abilities of different wired and wireless communication technologies to meet the broad range of communication requirements posed by Industry 4.0 in an efficient and reliable way. To this aim, inter-system interferences between different wireless technologies operating in the same unlicensed frequency band need to be monitored and controlled, as well as inter-cell interferences for wireless technologies using the licensed

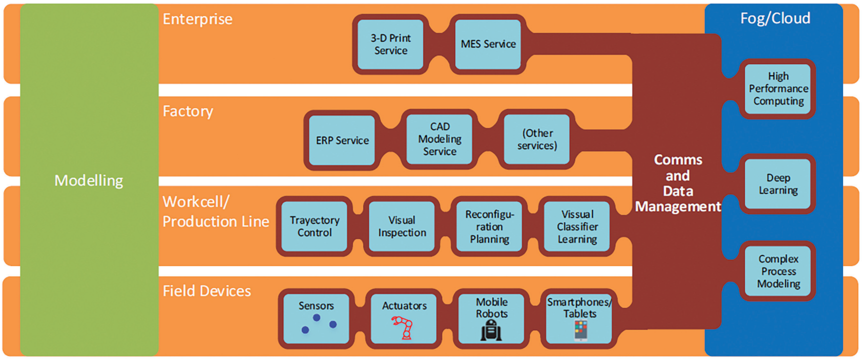


Figure 5.2 Communication network and data management system into the AUTOWARE Reference Architecture.

spectrum. From a data management standpoint, real-time data availability requirements, optimized utilization of IT resources (particularly for SMMEs), and data ownership constraints call for distributed data management schemes, whereby data are stored, replicated, and accessed from multiple locations in the network, depending on data generation and data access patterns, as well as the status of physical resources at the individual nodes.

To efficiently integrate the different communication technologies in a unique network and handle the data management process, we adopt a software-defined hierarchical approach where a central entity guarantees the coordination of local and distributed managers resulting in a mix of centralized management (orchestration) and decentralized operation of the communication and data management functions. Communication links are organized in different virtual tiers based on the performance requirements of the application they support. Different communications and data management strategies can then be applied at each tier to meet the specific communication and data availability requirements of each application. To implement the proposed hierarchical and multi-tier management architecture, we consider the use of RAN (Radio Access Network) Slicing and Cloud RAN as technological enablers to achieve the flexibility, scalability, and adaptation architectural capabilities needed to guarantee the stringent and varying communication and data distribution requirements of industrial applications.

This chapter is organized as follows. Section 5.2 presents the requirements imposed by Industry 4.0 to the communications and data management system. Section 5.3 reviews communication architectures proposed for Industrial Wireless Networks, and Section 5.4 presents traditional and current trends on the design of data management strategies in industrial environments. Section 5.5 presents the proposed communications and data management architecture, and the technological enablers considered to build up the architecture, RAN Slicing and Cloud RAN. Section 5.6 describes the possibilities offered by the proposed hierarchical architecture to implement hybrid management schemes to introduce flexibility in the management of wireless connections while maintaining a close coordination with a central network manager. Section 5.7 presents examples of early adoption of communication and data management concepts supported by the suggested architecture. How the reference communications and data management architecture fits into the overall AUTOWARE framework is presented in Section 5.8. Section 5.9 summarizes and concludes the chapter.

5.2 Industry 4.0 Communication and Data Requirements

Industry 4.0 poses a complex communication environment because of the wide set of different industrial applications and services that will coexist, all of them demanding very different and stringent communication requirements. The 5GPPP classifies industrial use cases in five families, each of them representing a different subset of communication requirements in terms of latency, reliability, availability, throughput, etc. [6]. Instant process optimization based on real-time monitoring of the manufacturing performance and the quality of produced goods is one of the most demanding use case families in terms of latency and reliability. Some of the sensors may communicate at low bitrates but with ultra-low latency and ultra-high reliability, whereas vision-controlled robot arms or mobile robots may require reliable high-bandwidth communication. Inside the factory, there are also applications or services without time-critical requirements, such as the localization of assets and goods and logistic processes, non-time critical quality control, or data capturing for later usage in virtual design contexts. The challenge in this second use case family is to ensure high availability of the wireless networks, given the harsh industrial environment. Remotely controlling digital factories requires end-to-end communications between remote workers and the factory. This use case family could simply involve the use of tablets or smartphones, or more complex scenarios with augmented reality devices that facilitate the creation of virtual back office teams that exploit the collected data for preventives analytics. In this use case family, there is a less stringent need for low latency, but high availability is key to ensure that emergency maintenance actions can take place immediately. The fourth use case family identified involves the connectivity between different production sites as well as with further actors in the value chain (e.g. suppliers, logistics) seamlessly. A high level of network and service availability and reliability including wireless link is one of the key requirements. The last use case family identified by the 5G-PPP considers that factories will play an important role in the provisioning of the connected goods that are produced, for which autonomy is a key requirement. Table 5.1 summarizes the communication requirements for each of the five use case families identified by the 5G-PPP.

The International Society of Automation (ISA) and ETSI also highlight the diverse communication requirements of industrial applications. For example, ISA establishes safety, control, and monitoring applications in six different classes based on the importance of message timeliness [9]. ETSI has also investigated the communication requirements of industrial automation

Table 5.1 5G-PPP use case families for manufacturing [6]

Use Case Family	Representative Scenarios	Latency	Reliability	Bandwidth
1. Time-critical process optimization inside factory	• Real-time closed-loop communication between machines to increase efficiency and flexibility • 3D augmented reality applications for training and maintenance • 3D video-driven interaction between collaborative robots and humans	Ultra-low	Ultra-high	Low to high
2. Non-time-critical in-factory communication	• Identification/tracing of objects/goods inside the factory • Non-real-time sensor data capturing for process optimization • Data capturing for design, simulation, and forecasting of new products and production processes	Less critical	High	Low to high
3. Remote control	• Remote quality inspection/diagnostics • Remote virtual back office	Less critical	High	Low to high
4. Intra-/Inter-enterprise communication	• Identification/tracking of goods in the end-to-end value chain • Reliable and secure interconnection of premises (intra-/inter-enterprise) • Exchanging data for simulation/design purposes	Ultra-low to less critical	High	Low to high
5. Connected goods	• Connecting goods during product lifetime to monitor product characteristics, sensing its surrounding context and offering new data-driven services	Less critical	Low	Low

in [10] and differentiated two types of applications. The first type involves the use of sensors and actuators in industrial automation and its main requirement is the real-time behavior or determinism. The second type of applications involves the communication at higher levels of the automation hierarchy, e.g. at the control or enterprise level, where throughput, security, and reliability become more important. Automation systems are subdivided into three main classes (manufacturing cell, factory hall, and plant level) with different needs in terms of latency (from 5 to 20 ms). Their requirements in terms of latency, update time, and number of devices can notably differ between them (see Table 5.2). However, all three classes require a 10^{-9} packet loss rate and a 99.999% application availability.

The timing requirements depend on different factors. As presented by the 5GPPP in [6], process automation industries (such as oil and gas, chemicals, food and beverage, etc.) typically require cycle times of about 100 ms. In factory automation (e.g. automotive production, industrial machinery, and consumer products), typical cycle times are 10 ms. The highest demands

Table 5.2 Performance requirements for three classes of communication in industry established by ETSI [10]

	Manufacturing Cell	Factory Hall	Plant Level
Indoor/outdoor application	Indoor	Mostly indoor	Mostly outdoor
Spatial dimension L×W×H (m^3)	10×10×3	100×100×10	1000×1000×50
Number of devices (typically)	30	100	1000
Number of parallel networks (clusters)	10	5	5
Number of such clusters per plant	50	10	1
Min. number of locally parallel devices	300	500	250
Network type	Star	Star/Mesh	Mesh
Packet size (on air, byte)	16	200	105
Max. allowable latency (end-to-end) incl. jitter/retransmits (ms)	5 ± 10%	20 ± 10%	20 ± 10%
Max. on-air duty cycle related to media utilization	20%	20%	20%
Update time (ms)	50 ± 10%	200 ± 10%	500 ± 10%
Packet loss rate (outside latency)	10^{-9}	10^{-9}	10^{-9}
Spectral efficiency (typically) (bis/s/Hz)	1	1.18	0.13
Bandwidth requirements (MHZ)	8	34	34
Application availability	Exceeds 99.999%		

Table 5.3 Timing requirements for motion control systems [6]

Requirement	Value
Cycle time	1 ms (250 μs . . . 31.25 μs)
Response time/update time	. . . 100 μs
Jitter	<1 μs . . . 30 ns
Switch latency time	. . . 40 ns
Redundancy switchover time	<15 μs
Time synchronization accuracy	. . . 100 ns

Table 5.4 Communication requirements for some industrial applications [5]

	Motion Control	Condition Monitoring	Augmented Reality
Latency/cycle time	250 μs–1 ms	100 ms	10 ms
Reliability (PER)	1e-8	1e-5	1e-5
Data rate	kbit/s–Mbit/s	kbit/s	Mbit/s–Gbit/s

are set by motion control applications (printing machines, textiles, paper mills, etc.) requiring cycle times of less than 1 ms with a jitter of less than 1 μs. For motion control, current requirements are shown in Table 5.3. Table 5.4 also shows the communication requirements of three relevant application examples (extracted from [5]) that illustrate the range of diverging and stringent communications requirements imposed by Industry 4.0.

These requirements have been confirmed within AUTOWARE. The communication requirements of several industrial use cases that are being developed within AUTOWARE have been analyzed. For example, in the PWR Pack AUTOWARE use case presented in [11], a stringent latency bound of 1 ms with a data rate lower than 100 kb/s is imposed to transmit commands from a Programmable Logic Controller (PLC) to a robot to control the servomotors and the movement of the robot, while 1–100 Mb have to be transmitted per image from a camera to a 3D visualization system tolerating a maximum 5 ms latency. On the other hand, the communication between a fixed robot and a component supplier mobile robotic platform within the neutral experimentation facility for collaborative robotics that is being developed by IK4-Tekniker [12] requires robust, flexible, and highly reliable wireless communication with latency bounded to some hundreds of milliseconds to guarantee the coordination and interoperation of both robots.

Due to the fact that the application functions should be applicable to different types of network nodes, they cannot rely only on specific communication functions, but include additional functions like smart data distribution and management. It should be worth noting that the ultimate

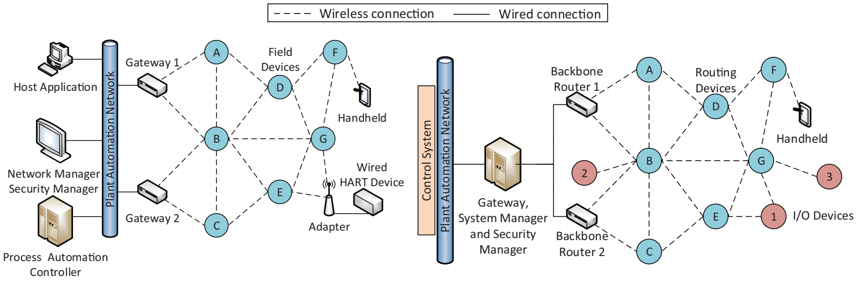
Table 5.5 Additional requirements for different application scenarios [13, 14]

	Desired Value	Application Scenario
Connectivity	300.000 devices per AP	Massive M2M connectivity
Battery life	>10 years	Hard-to-reach deployments
Reliability	99.999%	Protection and control
Seamless and quick connectivity	–	Mobile devices

Industry 4.0 application performance is the result of the concurrent operation and synergies across communication architectures and data distribution strategies. Table 5.5 shows some additional requirements for different application scenarios that impose additional constraints to manage the communications network and impose specific constraints to data management schemes [13, 14]. A massive M2M (machine to machine) connectivity will require an Access Point (AP) to support hundreds of thousands of field devices, with obvious limitations on the data rates each can support, and thus on rates at which they are enquired for (new) data. Maintenance for such large connectivity should be very low; thus, a very long battery period for such devices will be a necessity. A battery life for wireless devices greater than 10 years will mean that many hard-to-reach sensors and actuators could only sustain very low data rates. Reliability will play a critical role in industrial requirements with safety protection and control applications, calling for resilient data management schemes. In addition to all these requirements, a network should also be able to provide pervasive connectivity experience for the devices that may transition from outdoors to indoors location in a mobile scenario. Finally, data availability issues impose other specific requirements. For example, depending on applications, data might not be replicated outside of a set of devices or a geographical area for ownership reasons. Data might have to be replicated, instead, on other groups of nodes for data availability. Conversions across data formats might be needed, to guarantee interoperability across different factory or enterprise systems. All these issues belong to the broader concept of data sovereignty that is the main focus of the Industrial Data Space (IDS) initiative [15].

5.3 Industrial Wireless Network Architectures

Traditionally, communication networks in industrial systems have been based on wired fieldbuses and Ethernet-based technologies, and often on proprietary standards such as HART, PROFIBUS, Foundation Fieldbus H1, etc. While

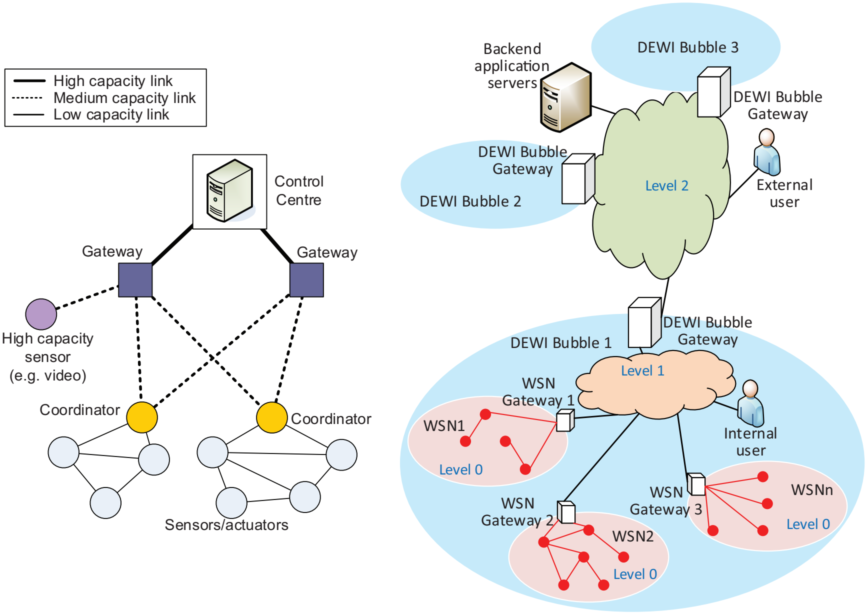


a) WirelessHart network architecture b) ISA100.11a network architecture

Figure 5.3 Examples of centralized management architectures.

wired technologies can provide high communications reliability, they are not able to fully meet the required flexibility and adaptation of future manufacturing processes for Industry 4.0. Wireless communication technologies present key advantages for industrial monitoring and control systems. They can provide connectivity to moving parts or mobile objects (robots, machinery, or workers) and offer the desired deployment flexibility by minimizing and significantly simplifying the need of cable installation. Operating in unlicensed frequency bands, WirelessHART, ISA100.11a, and IEEE 802.15.4e, are some of the wireless technologies developed to support industrial automation and control applications. These technologies are based on the IEEE 802.15.4 physical and MAC (Medium Access Control) layers, and share some fundamental technologies and mechanisms, e.g., a centralized network management and Time Division Multiple Access (TDMA) combined with Frequency Hopping (FH). Figure 5.3 shows the network architecture for WirelessHart and ISA100.11a. In both examples, there is a central network management entity referred to as Network Manager in a WirelessHart network and System Manager in the ISA100.11a network that is in charge of the configuration and management at the data link and network levels of the communications between the different devices (gateways, routers, and end devices).

The main objective of having a centralized network management is to achieve high communications reliability levels. However, the excessive overhead and reconfiguration time that results from collecting state information by the central manager (e.g. the Network Manager in a WirelessHart network or the System Manager in a ISA100.11a network) and distributing management decisions to end devices limits the reconfiguration and scalability capabilities of networks with centralized management, as highlighted in [16] and [17]. To overcome this drawback, the authors of [17–21] proposed to divide a large network into multiple subnetworks and considered a hierarchical management architecture. In this context, each subnetwork has its own manager that deals with the wireless dynamics within its subnetwork. A global entity is in charge of the management and coordination of the entire network with the subnetwork managers. Proposals in [19–21] rely on hierarchical architectures and also propose the integration of heterogeneous technologies to efficiently guarantee the wide range of different communication requirements of industrial applications; the need of using heterogeneous technologies in manufacturing processes was already highlighted by ETSI in [10]. For example, the approach proposed in [19], and shown in Figure 5.4(a), considers the deployment of several subnetworks in the lowest level of the industrial network architecture connecting sensors and actuators. The deployed devices collect data and send it to a central control and management system, which is located at the highest level of the network architecture. This IWN integrates and exploits various wireless technologies with different communication capacities at different levels of the architecture. Coordinators at each subnetwork act as sink nodes and collect data from different low-bandwidth sensors and transmit it to gateway nodes using higher-bandwidth wireless technologies. The gateway nodes are usually deployed so that they can collect and transmit data from various sink nodes to the central control and management system through high-bandwidth technologies. Another example is the network architecture proposed in the framework of the DEWI (Dependable Embedded Wireless Infrastructure) project [22]. The DEWI hierarchical architecture [20] is depicted in Figure 5.4(b). This architecture is based on the concept of DEWI Bubbles. A DEWI Bubble is defined as a high-level abstraction of a set of industrial wireless sensor networks (WSN) located in proximity with enhanced inter-operability, technology reusability, and cross-domain development. In ref. [20], standard interfaces are defined to allow WSNs that can implement different communication technologies to exchange information among them. Each WSN has its own Gateway that is in charge of the WSN management and protocol translation. The use of resources at different WSNs inside a Bubble is coordinated by a higher-level gateway that also provides protocol translation functionalities for the WSN under its support. Communication between different Bubbles is possible through their corresponding Bubble Gateways. Interfaces, services, and interoperability features of the different nodes and gateways are described in [20]. Ref. [20] is focused on IoT systems and provides connectivity to



a) IWN architecture adapted from [19]   b) IWN architecture adapted from [20]

Figure 5.4 Examples of hierarchical IWN architectures.

a large number of communication devices. However, it does not particularly consider applications with very stringent latency and reliability requirements.

Another interesting hierarchical management architecture that considers the use of heterogeneous wireless technologies is presented in [21], and has been developed in the framework of the KoI project [23]. The architecture presented in [21] proposes a two-tier management approach for radio resource coordination to support mission-critical wireless communications. To guarantee the capacity and scalability requirements of the industrial environment, ref. [21] considers the deployment of multiple small cells. Each of these small cells can implement a different wireless technology, and has a Local Radio Coordinator (LRC) that is in charge of the fine-grained management of radio resources for devices in its cell. On a higher level, there is a single Global Radio Coordinator (GRC) that carries out the radio resource management on a broader operational area and coordinates the use of radio resources by the different cells to avoid inter-system (for wireless technologies using unlicensed bands) and inter-cell (for those working on licensed bands) interference among them. In ref. [21], the control plane and the data plane are split following the Software-Defined Networking (SDN) principle. Control management is carried out in a centralized mode at LRCs and the GRC. For the data plane, centralized and assisted Device-to-Device (D2D) modes are considered within each cell.

5G networks are also being designed to support, among other verticals, Industrial IoT systems [24]. To this end, the use of Private 5G networks is proposed [25]. Private 5G networks will allow the implementation of local networks with dedicated radio equipment (independent of traffic fluctuation in the wide-area macro network) using shared and unlicensed spectrum, as well as locally dedicated licensed spectrum. The design of these Private 5G networks to support industrial wireless applications considers the implementation of several small cells to cover the whole industrial environment integrated in the network architecture as shown in Figure 5.5. Private 5G networks will have to support Ultra Reliable Low Latency Communications (URLLC) for time-critical applications, and Enhanced Mobile Broadband services for augmented/virtual reality services. In addition, the integration of 5G networks with Time Sensitive Networks (TSN)[1] is considered to guarantee deterministic end-to-end industrial communications, as presented in [24]. Figure 5.6 summarizes these key capabilities of Private 5G networks for Industrial IoT systems.

The reference communication and data management architecture designed in AUTOWARE is very aligned with the concepts that are being studied for Industrial 5G networks. The support of very different communication requirements demanded for a wide set of industrial applications (from time-critical applications to ultra-high demanding throughput applications) and the integration of different communication technologies (wired and wireless) are key objectives of the designed AUTOWARE communication and data management architecture to meet the requirements of Industry 4.0. In fact, AUTOWARE focuses on the design of a communication architecture that is able to efficiently meet the varying and stringent communication

---

[1] TSN is a set of IEEE 802 Ethernet sub-standards that aim to achieve deterministic communication over Ethernet by using time synchronization and a schedule that is shared between all the components (i.e. end systems and switches) within the network. By defining various queues based on time, TSN ensures a bounded maximum latency for scheduled traffic through switched networks, thereby guaranteeing the latency of critical scheduled communication. Additionally, TSN supports the convergence of having critical and non-critical communication sharing the same network, without interfering with each other, resulting in a reduction of costs (reduction of required cabling).

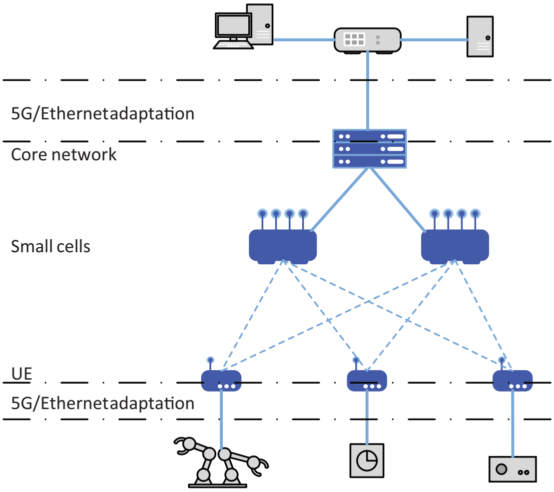


Figure 5.5 Private 5G Networks architecture for Industrial IoT systems [24].

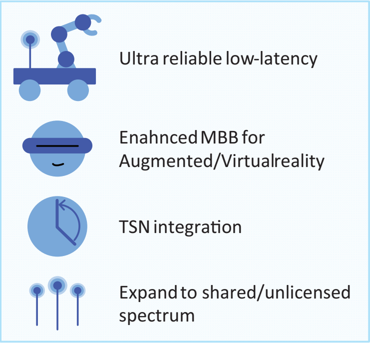


Figure 5.6 Key capabilities of Private 5G Networks for Industrial IoT systems [24].

requirements of the wide set of applications and services that will coexist within the factories of the future; in contrast to the architectures proposed in [20] and [21], which are mainly designed to guarantee communication requirements of a given type of service (to provide connectivity to a large number of communication devices in [20], and mission-critical wireless communications in [21]). In addition, this work goes a step further and analyzes the requirements of the communication architecture from the point of view of the data management and distribution.

5.4 Data Management in Industrial Environments

Traditionally, industrial application systems tend to be entirely centralized. For this reason, distributed data management has not been studied extensively in the past, and the emphasis has been put on the efficient wireless and wired communication within the industrial environment. The reader can find state-of-the-art approaches on relevant typical networks in [19, 26–28].

However, there have been some interesting works on various aspects of the data management process, e.g., end-to-end latency provisioning. In [29], the authors present a centralized routing method, and, consequently, they do not use proxies, data handling special nodes, or hierarchical data management. In [30], the authors address different optimization objectives, focusing on minimizing the maximum hop distance, rather than guaranteeing it as a hard constraint. Also, they assume a bounded number of proxies and they examine only on the worst-case number of hops. In [31], the authors present a cross-layer approach, which combines MAC-layer and cache management techniques for adaptive cache invalidation, cache replacement, and cache prefetching. In [32], the authors consider a different data management objective: replacement of locally cached data items with new ones. As the authors claim, the significance of this functionality stems from the fact that data queried in real applications is not random but instead exhibits locality characteristics. Therefore, the design of efficient replacement policies, given an underlying caching mechanism, is addressed. In [33], although the authors consider delay aspects and a realistic industrial IoT model (based on WirelessHART), their main objective is to bound the worst-case delay in the network. Also, they do not exploit the potential presence of proxy nodes, and consequently, they stick to the traditional, centralized industrial IoT setting. In [34], the authors consider a multi-hop network organized in clusters and provide a routing algorithm and cluster partitioning. Our distributed data management concepts and algorithms can work on top of this approach (and of any clustering approach), for example, by allocating the role of proxies to cluster-heads. In fact, clustering and our solutions address two different problems.

5.5 Hierarchical Communication and Data Management Architecture for Industry 4.0

The network architecture presented in this chapter is designed to provide flexible and efficient connectivity and data management in Industry 4.0.

AUTOWARE proposes a hierarchical management architecture that supports the use of heterogeneous communication technologies. The proposed architecture also establishes multiple tiers where communication cells are functionally classified; different tiers establish different requirements in terms of reliability, latency, and data rates and impose different constraints on the management algorithms and the flexibility to implement them.

5.5.1 Heterogeneous Industrial Wireless Network

As presented in Section 5.2, industrial applications demand a wide range of different communication requirements that are difficult to be efficiently satisfied with a single communication technology. In this context, the proposed architecture exploits the different capabilities of the available communication technologies (wired and wireless) to meet the wide range of requirements of industrial applications. For example, unlicensed wireless technologies such as WirelessHART, ISA100.11a, or IEEE 802.15.4e must implement mechanisms to minimize the interference generated to other potential devices sharing the same band, as for example, listen-before-talk-based channel access schemes. Although these wireless technologies are suitable to efficiently meet the requirements of non-time-critical monitoring or production applications, they usually fail to meet the stringent latency and reliability requirements of time-critical automation and control applications. In addition, these technologies were designed for static and low-bandwidth deployments, and the digitalization of industries requires significantly higher bandwidth provisioning and the capacity to integrate moving robots and objects in the factory. On the other hand, cellular standards operating on licensed frequency bands introduced in Release 14 [35] mechanisms for latency reduction in order to support certain delay critical applications. Moreover, Factories of the Future represent one of the key verticals for 5G-PPP, and 5G technologies are being developed to support a large variety of applications scenarios, targeting URLLC with a latency of about 1 ms and reliability of $1–10^{-9}$ [36]. Also, Private LTE and Private 5G networks will be relevant technologies to be used in industrial environments [25]. As a complement of wireless technologies, the use of wired communication technologies, as for example TSN, can also be considered for communication links between static devices.

In this context, we propose that several subnetworks or cells (we will use the term cell throughout the rest of the document) implementing heterogeneous technologies cover the whole industrial plant (or several plants). We adopt and use the concept of cell to manage the communications and data management resources and improve the network scalability. Different cells can use different communication technologies. Cells using different communication technologies could overlap in space. Also, cells using the same technology but in a different channel could cover the same area (or partially). Each network node is connected to the cell that is able to most efficiently satisfy its communication needs. For example, WirelessHART can be used to monitor a liquid level and control a valve, while 5G communications can be employed for time-critical communications between a sensor and an actuator. TSN could be a good candidate to implement long-distance backhaul links between static devices. Figure 5.7 illustrates the concept of cells in the proposed heterogeneous architecture with five cells implementing two different technologies. Technology 1 and Technology 2 could represent WirelessHART and 5G technologies. Technology 3 is used to connect each cell through a local management entity, referred to as Local Manager (LM), to a central management entity represented as Orchestrator in Figure 5.7 (roles of LMs and the Orchestrator in the proposed reference communication and data management architecture are presented in the next section), and it could be implemented with TSN (the communication link between LMs and the Orchestrator could also be implemented by a multi-hop link using also heterogeneous technologies for improved flexibility and scalability (e.g., IEEE 802.11 and TSN)).

Cells implementing wireless communication technologies that operate in unlicensed spectrum bands can suffer from inter-system and intra-system interferences. Mechanisms to detect external interferences are needed, and cells need to be coordinated to guarantee interworking and coexistence between concurrently operating technologies. Cells implementing a communication technology using licensed spectrum, as for example, LTE or 5G networks, are also possible. Although the use of licensed spectrum bands guarantees communications free of external interference, planning and coordination among multiple cells is still needed to control inter-cell interference. Considering the highly dynamic and changing nature of industrial environments, coordination among cells need to be carried out dynamically in order to guarantee the stringent communication requirements of industrial automation processes.

5.5.2 Hierarchical Management

The proposed reference communication and data management architecture considers a hierarchical structure that combines local and decentralized

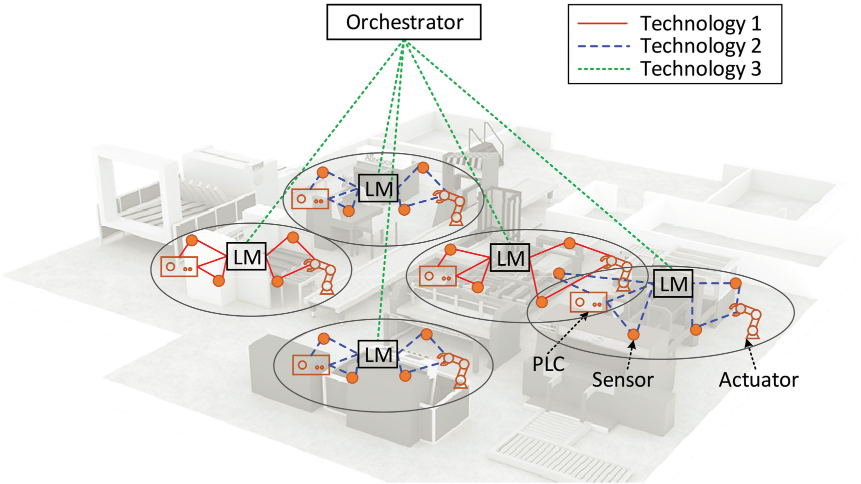


Figure 5.7 Hierarchical and heterogeneous reference architecture to support CPPS connectivity and data management.

management with centralized decisions to efficiently use the available communication resources and carry out the data management in the system. The management structure is depicted in Figure 5.7, and the functions of the two key components, the Orchestrator and the LMs, are next described.

5.5.2.1 Hierarchical communications

The Orchestrator is in charge of the global coordination of the radio resources assigned to the different cells. It establishes constraints to the radio resource utilization that each cell has to comply with in order to guarantee coordination and interworking of different cells, and finally guarantee the requirements of the industrial applications developed in the whole plant. For example, the Orchestrator must avoid inter-cell interferences between cells implementing the same licensed technology. It must also guarantee interworking among cells implementing wireless technologies using unlicensed spectrum bands in order to avoid inter-system interferences, as for example, dynamically allocating non-interfering channels to different cells based on the current demand. LMs are implemented at each cell. An LM is in charge of the local management of the radio resources within its cell and makes local decisions to ensure that communication requirements of nodes in its cell are satisfied.

As shown in Figure 5.8, LMs are in charge of management functions such as Radio Resource Allocation, Power Control, or Scheduling. These functions

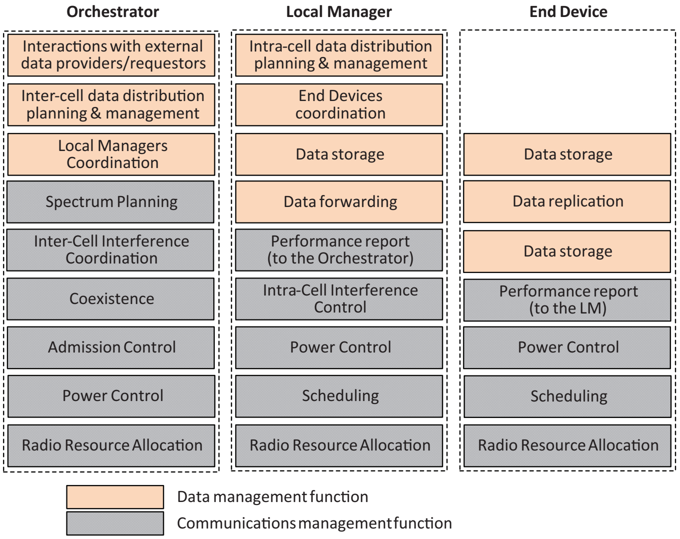


Figure 5.8 Communication and data management functions in different entities of the hierarchical architecture.

locally coordinate the use of radio resources among the devices attached to the same cell and require very short response times. Intra-Cell Interference Control needs to be carried out also by the LM if several transmissions are allowed to share radio resources within the same cell. LMs also report the performance levels experienced within its cell to the Orchestrator. Thanks to its global vision, the Orchestrator has the information required and the ability to adapt and (re-)configure the whole network. For example, under changes in the configuration of the industrial plant or in the production system, the Orchestrator can reallocate frequency bands to cells implementing licensed technologies based on the new load conditions or the new communication requirements. It could also establish new interworking policies to control interferences between different cells working in the unlicensed spectrum. The Orchestrator can also establish constraints about the maximum transmission power or the radio resources to allocate to some transmissions to guarantee the coordination between different cells. It is also in charge of the Admission Control. In this context, the Orchestrator also decides to which cell a new device is attached to consider the communication capabilities of the device, the communication requirements of the application, and the current operating conditions of each cell.

The described hierarchical communication and data management architecture corresponds to the control plane. We consider that control plane and user plane[2] are separated. Therefore, although a centralized management is adopted within a cell, nodes in proximity might communicate directly using D2D communications. In some cells, end-devices might also participate in management functions, for example, if distributed radio resource allocation algorithms are considered for D2D communications in 5G cells. End devices can also participate in other management functions such as Power Control or Scheduling (see Figure 5.8).

5.5.2.2 Data management

The Orchestrator plays an important role in facilitating the development of novel smart data distribution solutions that cooperate with cloud-based service provisioning and communication technologies. Smart proactive data storage/replication techniques can be designed, ensuring that data is located where it can be accessed by appropriate decision makers in a timely manner based on the performance of the underlying communication infrastructure. Consequently, the Orchestrator serves as a great opportunity to implement different types of data-oriented automation functions at reduced costs, like interactions with external data providers or requestors, inter-cell data distribution planning, and management and coordination of the LMs.

On the other hand, it is widely recognized that entirely centralized solutions to collect and manage data in industrial environments are not always suitable [38, 39] This is due to the fact that in order to assure quick reaction, process monitoring and automation control may span among multiple physical locations. Additionally, the adoption of IoT technologies with the associated massive amounts of generated data makes decentralized data management inevitable. A significant challenge is that, when data are managed across multiple physical locations, data distribution needs to be carefully designed, so as to ensure that industrial process control is not affected by the well-known issues related to communication delays and jitters [26, 40].

For data management, allocation of roles on the Orchestrator, LMs, and individual devices is less precisely defined in general, and can vary significantly on a per-application and per-scenario basis. In general, we expect

---

[2] The User Plane carries the network user traffic, i.e., the data that is generated and consumed by the AUTOWARE applications and services. The Control Plane carries signaling traffic, and is critical for the correct operation of the network. For example, signaling messages would be needed to properly configure a wired/wireless link to achieve the necessary latency and reliability levels to support an application. They would also be needed to intelligently control the data management process. The Control Plane therefore is needed to enable the user data exchange between the different AUTOWARE components.

that the Orchestrator would decide on which cells (controlled by one LM each) data need to be available and thus replicated. Also, it would decide out of which cells they must not be replicated due to ownership reasons. It would implement, in collaboration with cloud platforms, authentication of users across cells and, when needed, data transcoding functions. Thus, we expect the Orchestrator to be responsible for managing the heterogeneity issues related to managing data across a number of different cells, possibly owned and operated by different entities. LMs would manage individual cells. They would typically decide where, inside the cell, data need to be replicated, stored, and moved dynamically, based on the requirements of the specific applications, and the resources available at the individual nodes. Note that data will in general be replicated across the individual nodes, and not exclusively at the LMs, to guarantee low delays and jitters, which might be excessive if the LMs operate as unique centralized data managers. In some cases, end-devices can also participate in management functions, for example, by exploiting D2D communications to directly exchange data between them, implementing localized data replication or storage policies. In those cases, the data routing is not necessarily regulated centrally, but can be efficiently distributed, using appropriate cooperation schemes. In the architecture, therefore, the control of data management schemes can be performed centrally at the Orchestrator, locally at the LMs, or even at individual devices, as appropriate. Data management operations become distributed, and they exploit devices that lie between source and destination devices, like the use of proxies for data storage and access.

5.5.3 Multi-tier Organization

In the proposed reference communication and data management architecture, cells are organized in different tiers depending on the communication requirements of the industrial application they support. LMs of cells in different tiers consider the use of different management algorithms to efficiently meet the stringent requirements of the different industrial applications they support. For example, regarding scheduling, a semi-persistent scheduling algorithm could be applied in LTE cells to guarantee ultra-low latency communications; semi-persistent scheduling algorithms avoid delays associated to the exchange of signaling messages to request (from the device to the base station or eNB) and grant (from the base station or eNB to the device) access to the radio resources. However, semi-persistent scheduling algorithms might not be adequate for less demanding latency requirements due to the potential

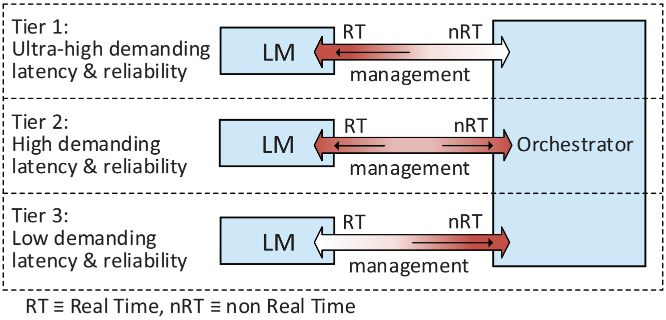


Figure 5.9 LM–Orchestrator interaction at different tiers of the management architecture.

underutilization of radio resources. The different requirements in terms of latency and reliability of the application supported by a cell also affect the exact locations where data should be stored and replicated. For example, in time-critical applications, the lower the data access latency bound is, the closer to the destination the data should be replicated.

The requirements of the nodes connected to a cell also influence the type of interactions between the LM of the cell and the Orchestrator. LMs of cells that support communication links with loose latency requirements can delegate some of their management functions to the Orchestrator. For these cells, a closer coordination between different cells could be achieved. Management decisions performed by LMs based on local information are preferred for applications with ultra-high demanding latency requirements (see Figure 5.9).

5.5.4 Architectural Enablers: Virtualization and Softwarization

Efficiency, agility, and speed are fundamental characteristics that future communication and networking architectures must accomplish to support the high diverging and stringent performance requirements of future communication systems (including but not limited to the industrial ones) [41]. In this context, the communication and data management architecture proposed within this chapter considers the use of RAN Slicing and Cloud RAN as enabling technologies to achieve the sought flexibility and efficiency.

5.5.4.1 RAN slicing

The proposed architecture considers the use of heterogeneous communication technologies. The assignment of communication technologies to industrial applications does not need to necessarily be a one-to-one matching. There is a clear trend nowadays in designing wireless technologies such that they can support more than one type of application even belonging to different "verticals", each of them with possibly radically different communication requirements. For example, LTE or 5G networks can be used to satisfy the ultra low-latency and high-reliability communications of a time-critical automation process. In addition, the same networks could also support applications that require high-throughput levels, such as virtual reality or 4K/8K ultra-high-definition video. This is typically achieved through network virtualization and slicing, to guarantee isolation of (virtual) resources and independence across verticals, or across applications in the same vertical.

In the proposed architecture, each cell can support several industrial applications with different communication requirements. The industrial applications supported by the same cell might require different management functions or techniques to satisfy their different requirements in terms of transmission rates, delay, or reliability. Moreover, it is important to ensure that the application-specific requirements are satisfied independently of the congestion and performance experienced by the other application supported by the same cell, i.e., performance isolation needs to be guaranteed between different applications. For example, the amount of traffic generated by a given application should not negatively influence the performance of the other application. In this context, we propose the use of RAN Slicing to solve the above-mentioned issues. RAN Slicing is based on SDN (Software-Defined Networking) and NFV (Network Function Virtualization) technologies, and it proposes to split the resources and management functions of an RAN into different slices to create multiple logical (virtual) networks on top of a common network [42]. Each of these slices, in this case, virtual RANs, must contain the required resources needed to meet the communication requirements of the application or service that such slice supports. As presented in [42], one of the main objectives of RAN Slicing is to assure isolation in terms of performance. In addition, isolation in terms of management must also be ensured, allowing the independent management of each slice as a separated network. As a result, RAN Slicing becomes a key technology to deploy a flexible communication and networking architecture capable of meeting the stringent and diverging communication requirements of industrial applications, and in particular, those of URLLC.

In the proposed architecture, each slice of a physical cell is referred to as virtual cell, as shown in Figure 5.10. Virtual cells resulting from the split of the same physical cell can be located at different levels of the

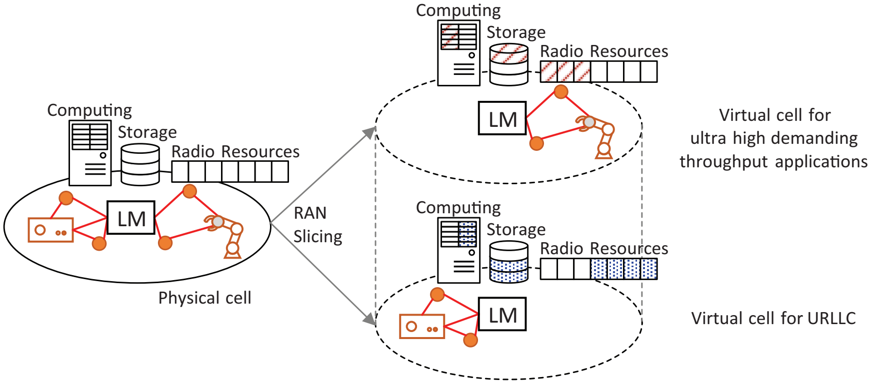


Figure 5.10 Virtual cells based on RAN Slicing.

multi-tier architecture depending on the communication requirements of the applications. Each virtual cell implements the appropriate functions based on the requirements of the application supported and must be assigned the RAN resources required to satisfy the requirements of the communication links it supports.

RAN resources (e.g., data storage, computing, radio resources, etc.) must be allocated to each virtual cell considering the operating conditions, such as the amount of traffic, the link quality, etc. The amount of RAN resources allocated to each virtual cell must be therefore dynamically adapted based on the operating conditions. Within the proposed reference architecture, the Orchestrator is the management entity in charge of creating and managing RAN slices or virtual cells. Thanks to the reports received from the LMs, the Orchestrator has a global view of the performance experienced at the different (virtual) cells. As a result, it is able to decide the amount of RAN resources that must be assigned to each virtual cell to guarantee the communication requirements of the applications.

With respect to data management functions, they will operate on top of the virtual networks generated by RAN Slicing. However, note that the requirements posed by data management will determine part of the network traffic patterns. Therefore, RAN Slicing defined by the Orchestrator might consider the traffic patterns resulting from data management operations, in order to optimize slicing itself.

5.5.4.2 Cloudification of the RAN

Cloud-based RAN (or simply Cloud RAN) is a novel paradigm for RAN architectures that applies NFV and cloud technologies for deploying

RAN functions [43]. Cloud RAN splits the base station into a radio unit, known as Radio Remote Head (RRH), and a signal-processing unit referred to as Base Band Unit (BBU) [44]. The key concept of Cloud RAN is that the signal processing units, i.e., the BBUs, can be moved to the cloud. Cloud RAN shifts from the traditional distributed architecture to a centralized one, where some or all of the base station processing and management functions are placed in a central virtualized BBU pool (a virtualized cluster which can consist of general purpose processors to perform baseband processing and that is shared by all cells) [43]. Virtual BBUs and RRHs are connected by a fronthaul network. Centralizing processing and management functions in the same location improves interworking and coordination among cells; virtual BBUs are located in the same place, and exchange of data among them can be carried out easier and with shorter delay.

We foresee Cloud RAN as the baseline technology for the proposed architecture, to implement hierarchical and multi-tier communication management. Cloud RAN will be a key technology to achieve a tight coordination between cells in the proposed architecture and to control inter-cell and inter-system interferences. As presented in [45] and [46], Cloud RAN can support different functional splits that are perfectly aligned with the foreseen needs of industrial applications; some processing functions can be executed remotely while functions with strong real-time requirements can remain at the cell site. In the proposed communication and data management architecture, the decision about how to perform this functional split must be made by the Orchestrator considering the particular communication requirements of the industrial applications supported by each cell (see Figure 5.11).

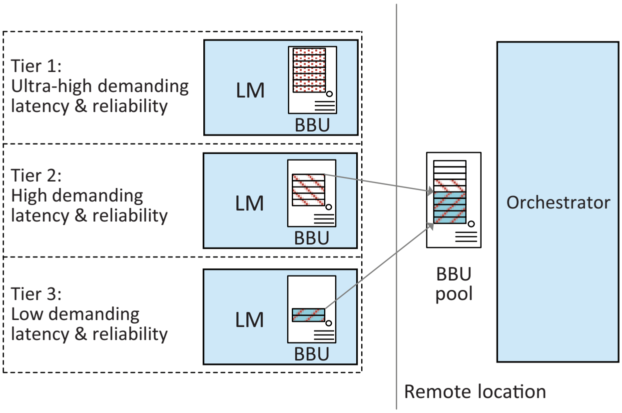


Figure 5.11 Cloudification of the RAN.

The Cloud RAN architectural paradigm allows for hardware resource pooling, which also reduces operational cost, by reducing power and energy consumption compared to traditional architectures [43], which results in an attractive incentive for industrial deployment. The cloudification of the RAN will also leverage RAN Slicing on a single network infrastructure and will increase flexibility for the construction of on-demand slices to support individual service types or application within a cell.

5.6 Hybrid Communication Management

Communication systems must be able to support the high dynamism of industrial environment, which will result from the coexistence of different industrial applications, different types of sensors, the mobility of nodes (robots, machinery, vehicles, and workers), and changes in the production demands. Industry 4.0 then demands flexible and dynamic communication networks able to adapt their configuration to changes in the environment to seamlessly ensure the communication requirements of industrial applications. To this end, communication management decisions must be based on current operating conditions and the continuous monitoring of experienced performance. The proposed hierarchical communication and data management architecture allows the implementation of hybrid communication management schemes that integrate local and decentralized management decisions while maintaining a close coordination through a central management entity (the Orchestrator in the reference AUTOWARE architecture) with global knowledge of the performance experienced in the whole industrial communication network. The hybrid communication management introduces flexibility in the management of wireless connections and increases the capability of the network to detect and react to local changes in the industrial environment while efficiently guaranteeing the communication requirements of industrial applications and services supported by the whole network.

In hybrid management schemes, management entities must interact to coordinate their decisions and ensure the correct operation of the whole network. Figure 5.12 represents the interactions between the management entities of the hierarchical architecture: the Orchestrator, LMs, and end-devices (as presented in Section 5.2, end-devices might also participate in the communication management). Boxes within each management entity represent different functions executed at each entity:

- Local measurements: This function measures physical parameters on the communication link, as for example, received signal level (received signal strength indication or RSSI), signal-to-noise ratio (SNR), etc. In addition, this function also measures and evaluates the performance experienced in the communication, as for example, throughput, delay, packet error ratio (PER), etc. This function is performed by each entity on its communication links.

- Performance gathering: This function collects information about the performance experienced at the different cells. This function is performed at the LMs, which collect performance information gathered by end-devices within its cell, and also at the Orchestrator, which receives performance information gathered by the LMs.
- Reasoning: The reasoning function processes the data obtained by the local measurements and the performance gathering functions to synthesize higher-level performance information. The reasoning performed at each entity will depend on the particular application supported (and the communication requirements of the application) and also on the particular management algorithm implemented. For example, if a cell supports time-critical control applications, the maximum value of latency experienced by the 99 percentile of packets transmitted might be of interest, while the average throughput achieved in the communication could be required to analyze the performance of a 3D visualization application.
- Reporting: This function sends periodic performance reports to the management entity in the higher hierarchical level. Particularly, end-devices send periodic reports to the LMs, which in turn report performance information to the Orchestrator.
- Global/local/communication management decision: This function executes the decision rule or decision policy. This function can be whatever of the communication management functions shown in Figure 5.8: for example, Admission Control or Inter-Cell Interference Coordination algorithms can be executed as the Global management decision function in the Orchestrator, Power Control or Radio Resource Allocation within a cell can be executed as the Local management decision function in the LMs, and Scheduling or Power Control can be executed as the Communication management decision function at the end-devices.

As shown in Figure 5.12, an end-device performs local measurements of the quality and performance experienced in its communication links. This local data (1) is processed by the reasoning function that provides high-level performance information (2a) that is reported to the LM in its cell (3). This high-level performance information can also be used by the end-device (2b) to get a management decision (4) and configure its communication

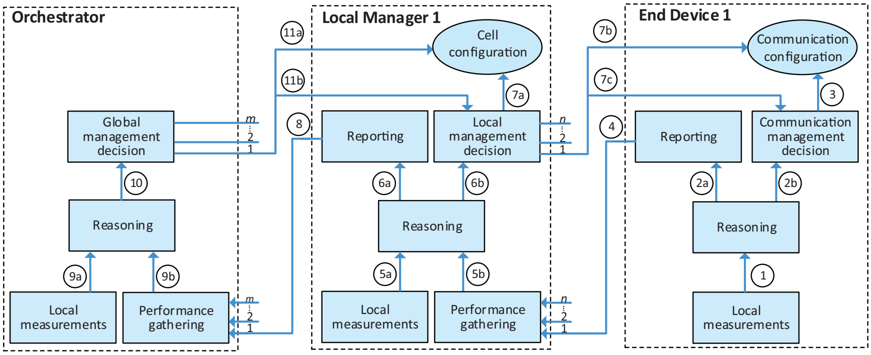


Figure 5.12 Hybrid communication management: interaction between management entities.

parameters in the case that the end-device has management capabilities. In this case, the management decisions taken by different end-devices in the same cell are coordinated by the LM in the cell, which can also configure some communication parameters of the end-devices (7b). Decisions taken by end-devices are constrained by the decisions taken by the LM (7c). If end-devices do not have management capabilities, the communication parameters for the end-devices are directly configured by the LM (8b). The Local management decisions taken by each LM are based on the performance information gathered by all end-devices in its cell (from 1 to *n* devices in the figure), and also on local measurements performed by the own LM. This data (5a and 5b) is processed by the reasoning function in the LM, and the resulting high-level performance information (6b) is used to take a local management decision and configure the communication parameters of the end-devices in its cell (7a, 7b, and 7c). Each LM also reports to the Orchestrator the processed information about the performance experienced in its cell (8). The Orchestrator receives performance information from all the LMs (from 1 to *m* LMs in the figure). The performance information gathered by the LMs (9b), together with local measurements performed by the Orchestrator in its communication links with the LMs (9a), is processed by the reasoning function in the Orchestrator. The high-level performance information (10) is used by the Orchestrator to achieve a global management decision and configure radio resources to use at each cell (11a). The global management decisions made by the Orchestrator constrain the local management decisions made by the LMs (11b) to guarantee the coordination among the different LMs in the network, and finally ensure the communication requirements of the industrial applications and services supported by the network.

5.7 Decentralized Data Distribution

The smart data management process provided by the architecture interacts with the underlying networking protocols. In order to provide both efficient data access and end-to-end delay guarantees, one of the technical components of the architecture is a dedicated decentralized data distribution. The main idea behind the decentralized data distribution is decoupling the Network plane from the Data plane. The data-enabled architecture functions selectively move data to different network areas and devise methods on how the data requests should be served, given a known underlying routing protocol. More specifically, the role of the decentralized data distribution component is three-fold:

1. It investigates where and when the data should be moved, and to which network areas.
2. It decides which network nodes can serve as special nodes and assume more responsibilities with respect to data management.
3. It indicates how the available data will be distributed and delivered to the individual network devices requesting it.

Note that the architecture enables the storing and replication of data between (i) (potentially mobile) nodes in the factory environment (e.g., the mobile nodes of the factory operators, nodes installed in work cells, nodes attached to mobile robots, etc.); (ii) edge nodes providing storage services for the specific (areas of the) factory; and (iii) remote cloud storage services. All the three layers can be used in a synergic way, based on the properties of the data and the requirements of the users requesting it. Depending on these properties, data processing may need highly variable computational resources. Advanced scheduling and resource management strategies lie at the core of the distributed infrastructure resources usage. However, such strategies must be tailored to the particular algorithm/data combination to be managed. Differently from the past, the scheduling process, instead of looking for smart ways to adapt the application to the execution environment, now aims at selecting and managing the computational resources available on the distributed infrastructure to fulfill some performance indicators.

The suggested architecture can be used in order to efficiently deploy the data management functions over typical industrial IoT networks. Initial results show that the decentralized data management scheme of the proposed architecture can indeed enhance various target metrics when applied to various industrial IoT networking settings. In the following subsections, we briefly review some recent examples, where the decentralized data distribution concepts resulted in an enhanced network performance.

5.7.1 Average Data Access Latency Guarantees

Assuming that applications in industrial IoT networks require that there is (i) a set of producers generating data (e.g., IoT sensors), (ii) a set of consumers requiring those data in order to implement the application logic (e.g., IoT actuators), and (iii) a maximum latency $\boldsymbol{L}_{\max}$ that consumers can tolerate in receiving data after they have requested them; the decentralized data management module (DML) offers an efficient method for regulating the data distribution among producers and consumers. The DML selectively assigns a special role to some of the network nodes, that of the proxy. Each node that can become a proxy potentially serves as an intermediary between producers and consumers, even though the node might be neither a producer nor a consumer. If properly selected, proxy nodes can significantly reduce the average data access latency; however, when a node is selected as a proxy, it has to increase its storing, computational, and communication activities. Thus, the DML minimizes the number of proxies, to reduce as much as possible the overall system resource consumption. In [47], we have provided an extensive experimental evaluation, both in a testbed and through simulations, and we demonstrated that the proposed decentralized data management (i) guarantees that the access latency stays below the given threshold and (ii) significantly outperforms traditional centralized and even distributed approaches, in terms of average data access latency guarantees.

5.7.2 Maximum Data Access Latency Guarantees

Another representative example of decentralized data management is the exploitation of the presence of a limited set of pre-installed proxy nodes, which are more capable than resource-limited IoT devices in the resource-constrained network (e.g., fog nodes). Different to the previous example, here we focused on network lifetime and on maximum (instead of average) data access latencies. The problem we addressed in [48] is the maximization of the network lifetime, given the proxy locations in the network, the initial limited energy supplies of the nodes, the data request patterns (and their corresponding parameters), and the maximum latency that consumer nodes can tolerate since the time they request data. We proved that the problem is computationally hard and we designed an offline centralized heuristic algorithm for identifying which paths in the network the data should follow and on which proxies they should be cached, in order to meet the latency constraint and to efficiently prolong the network lifetime. We implemented the method and evaluated its performance in a testbed, composed of IEEE 802.15.4-enabled network nodes. We demonstrated that the proposed heuristic (i) guarantees data access latency below the given threshold and (ii) performs well in terms of network lifetime with respect to a theoretically optimal solution.

5.7.3 Dynamic Path Reconfigurations

As in the previous examples, we assume that applications require a certain upper bound on the end-to-end data delivery latency from proxies to consumers and that at some point in time, a central controller computes an optimal set of multi-hop paths from producers to proxies and from proxies to consumers, which guarantee a maximum delivery delay, while maximizing the energy lifetime of the network (i.e., the time until the first node in the network exhaust energy resources). In this example, we focus on maintaining the network configuration in such a way that application requirements are met after important network operational parameters change due to some unplanned events (e.g., heavy interference, excessive energy consumption), while guaranteeing an appropriate utilization of energy resources. In [49], we provided several efficient algorithmic functions that locally reconfigure the paths of the data distribution process, when a communication link or a network node fails. The functions regulate how the local path reconfiguration should be implemented and how a node can join a new path or modify an already existing path, ensuring that there will be no loops. The proposed method can be implemented on top of existing data forwarding schemes designed for industrial IoT networks. We demonstrated through simulations the performance gains of our method in terms of energy consumption and data delivery success rate.

5.8 Communications and Data Management within the AUTOWARE Framework

The reference communication and data management architecture of AUTOWARE supports the control plane of the communication network and the data management system. As shown in Figure 5.13, end (or field)-devices such as sensors, actuators, mobile robots, etc., are distributed throughout the factory plant participating in different industrial processes or tasks.

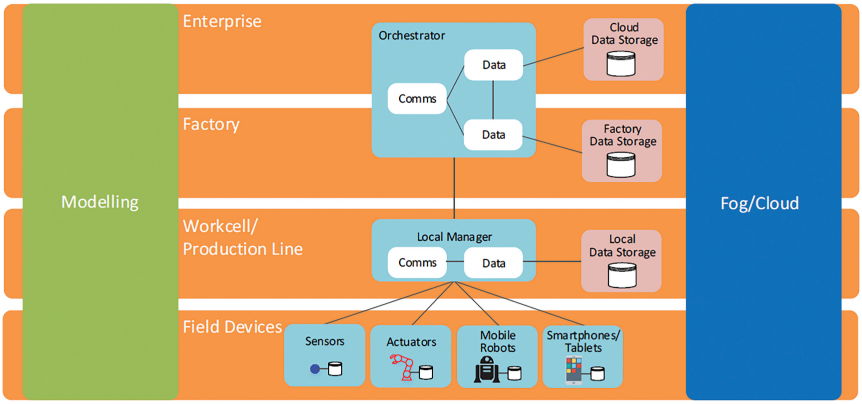


Figure 5.13 Integration of the hierarchical and multi-tier heterogeneous communication and data management architecture into the AUTOWARE Reference Architecture.

These field devices are then included within the Field Devices Layer of the AUTOWARE Reference Architecture defined in Chapter 10. Various LMs can be implemented at different workcells or production lines to locally manage the communication resources and data in the different communication cells deployed in the industrial plant. These management nodes are included in the Workcell/Production Line Layer, and they form a distributed management infrastructure that operates close to the field devices. As previously presented, both the Orchestrator and the LMs have communication and data management functionalities.

From the point of view of communications, the Orchestrator is in charge of the global management of the communication resources used by the different cells deployed within a factory plant. When there is only one industrial plant or when there are multiple but independent plants (from the communications perspective), the main communication functions of the Orchestrator are in the Factory Layer. However, if different industrial plants are deployed and they are close enough so that the operation of a cell implemented in a plant can affect the operation of a different cell in the other plant, then the Orchestrator should be able to manage the communication resources of the different plants. In this case, some of its communication functions should be part of the Enterprise Layer. Based on the previous reasoning, the Orchestrator and, in particular, the communication management function within the Orchestrator should be flexible and be able to be implemented in the Factory and the Enterprise Layer.

From the point of view of data storage, management, and distribution, the data can be circulated and processed at different levels of the architecture, depending on the targeted use case and the requirements that the industrial operator is imposing on the application. For example, if the requirements necessitate critical and short access latency applications (e.g., Table 5.5), such as condition monitoring, then imposing data transfers back and forth between the Field Layer, the Workcell/Production Line Layer, and the Factory Layer may lead to severe sub-optimal paths, which in turn negatively affect the overall network latency. At the same time, those transfer patterns will lead to poor network performance, as field devices often have to tolerate longer response times than necessary. In this case, the data can be stored and managed at the lower layers of the architecture, with the LMs in the role of the data coordinator. Another example is when the requirements necessitate the employment of computationally more sophisticated methods on larger volumes of data that can only be performed by stronger devices than those at the Field Layer, such as 3D object recognition or video tracking, which come with vast amounts of data. In this case, the data can be forwarded, stored, and processed in the higher levels of the architecture, the Factory Layer, or the Enterprise Layer, with the Orchestrator in the role of the data coordinator.

5.9 Conclusions

A software-defined heterogeneous, hierarchical, and multi-tier communication management architecture with edge-powered smart data distribution strategies has been presented in this chapter to support ubiquitous, flexible, and reliable connectivity and efficient data management in highly dynamic Industry 4.0 scenarios where multiple digital services and applications are bound to coexist. The proposed architecture exploits the different abilities of heterogeneous communication technologies to meet the broad range of communication requirements demanded by Industry 4.0 applications. Integration of the different technologies in an efficient and reliable network is achieved by means of a hybrid management strategy consisting of decentralized management decisions coordinated by a central orchestrator. Local management entities organized in different virtual tiers of the architecture can implement different management functions based on the requirements of the application they support. The hierarchical and multi-tier communication management architecture enables the implementation of cooperating, but distinct management functions to maximize flexibility and efficiency to meet the stringent and varying requirements of industrial applications. The proposed architecture considers the use of RAN Slicing and Cloud RAN as enabling technologies to meet reliably and effectively future Industry 4.0 autonomous assembly scenarios and modular plug & play manufacturing systems. The technological enablers of the communications and data management architecture were identified as part of the AUTOWARE framework, both in the user plane and in the control plane of the AUTOWARE reference architecture.

Acknowledgments

This work was funded by the European Commission through the FoF-RIA Project *AUTOWARE: Wireless Autonomous, Reliable and Resilient Production Operation Architecture for Cognitive Manufacturing* (No. 723909).

References

[1] V. K. L. Huang, Z. Pang, C. J. A. Chen and K. F. Tsang, "New Trends in the Practical Deployment of Industrial Wireless: From Noncritical to Critical Use Cases", in IEEE Industrial Electronics Magazine, vol. 12, no. 2, pp. 50–58, June 2018.

[2] T. Sauter, S. Soucek, W. Kastner and D. Dietrich, "The Evolution of Factory and Building Automation", in IEEE Industrial Electronics Magazine, vol. 5, no. 3, pp. 35–48, September 2011.

[3] *How Audi is changing the future of automotive manufacturing*, Feb. 2017. Available at https://www.drivingline.com/. Last access on 2017/12/01.

[4] C. H. Chen, M. Y. Lin and C. C. Liu, "Edge Computing Gateway of the Industrial Internet of Things Using Multiple Collaborative Microcontrollers", in IEEE Network, vol. 32, no. 1, pp. 24–32, January–February 2018.

[5] Plattform Industrie 4.0, "Network-based communication for Industrie 4.0", *Publications of Plattform Industrie 4.0*, April 2016. Available at http://www.plattform-i40.de. Last access on 2017/10/20.

[6] 5GPPP, *5G and the Factories of the Future*, October 2015.

[7] H2020 AUTOWARE project website: http://www.autoware-eu.org/.

[8] M. C. Lucas-Estañ, T. P. Raptis, M. Sepulcre, A. Passarella, C. Regueiro and O. Lazaro, "A software defined hierarchical communication and data management architecture for industry 4.0", *14th Annual Conference on Wireless On-demand Network Systems and Services* (WONS), Isola 2000, pp. 37–44, 2018.

[9] P. Zand, et al., “Wireless industrial monitoring and control networks: The journey so far and the road ahead”, *Journal of Sensor and Actuator Networks*, vol. 1, no. 2, pp. 123–152, 2012.

[10] ETSI, “Technical Report; Electromagnetic compatibility and Radio spectrum Matters (ERM); System Reference Document; Short Range Devices (SRD); Part 2: Technical characteristics for SRD equipment for wireless industrial applications using technologies different from Ultra-Wide Band (UWB)”, ETSI TR 102 889-2 V1.1.1, August 2011.

[11] E. Molina, et al., “The AUTOWARE Framework and Requirements for the Cognitive Digital Automation”, in *Proc. of the 18th IFIP Working Conference on Virtual Enterprises (PRO-VE)*, Vicenza, Italy, September 2017.

[12] M.C. Lucas-Estañ, J.L. Maestre, B. Coll-Perales, J. Gozalvez, “An Experimental Evaluation of Redundancy in Industrial Wireless Communications”, in *Proc. 2018 IEEE 23rd International Conference on Emerging Technologies and Factory Automation* (ETFA 2018), Torino, Italy, 4–7 September 2018.

[13] A. Varghese, D. Tandur, “Wireless requirements and challenges in Industry 4.0”, in *Proc. 2014 International Conference on Contemporary Computing and Informatics (IC3I)*, pp. 634–638, 2014.

[14] A. Osseiran et al., ”Scenarios for 5G mobile and wireless communications: the vision of the METIS project”, *IEEE Communications Magazine*, vol. 52, no. 5, pp. 26–35, May 2014.

[15] Reference architecture model for the industrial data space, Fraunhofer, 2017. Available at https://www.fraunhofer.de. Last access on 2017/12/01.

[16] S. Montero, J. Gozalvez, M. Sepulcre, G. Prieto, “Impact of Mobility on the Management and Performance of WirelessHART Industrial Communications”, in *Proc. 17th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, Krakw (Poland), pp. 17–21, September 2012.

[17] C. Lu et al., “Real-Time Wireless Sensor-Actuator Networks for Industrial Cyber-Physical Systems”, *Proc. of the IEEE*, vol. 104, no. 5, pp. 1013–1024, May 2016.

[18] Xiaomin Li, Di Li, Jiafu Wan, Athanasios V. Vasilakos, Chin-Feng Lai, Shiyong Wang, “A review of industrial wireless networks in the context of Industry 4.0”, *Wireless Networks*, vol. 23, pp. 23–41, 2017.

[19] J.R. Gisbert, et al., “Integrated system for control and monitoring industrial wireless networks for labor risk prevention”, *Journal of Network*

and Computer Applications, vol. 39, pp. 233–252, ISSN 1084-8045, March 2014.

[20] R. Sámano-Robles, T. Nordström, S. Santonja, W. Rom and E. Tovar, "The DEWI high-level architecture: Wireless sensor networks in industrial applications", in *Proc. of the 2016 Eleventh International Conference on Digital Information Management (ICDIM)*, Porto, pp. 274–280, 2016.

[21] I. Aktas, J. Ansari, S. Auroux, D. Parruca, M. D. P. Guirao, and -B. Holfeld: "A Coordination Architecture for Wireless Industrial Automation", in *Proc. of the European Wireless Conference*, Dresden, Germany, May 2017.

[22] DEWI Project website. Available at http://www.dewiproject.eu/. Last access on 2017/05/26.

[23] KoI Project website. Available at http://koi-projekt.de/index.html. Last access on 2017/05/26.

[24] Mehmet Yavuz, *How will 5G transform Industrial IoT?*, Qualcomm Technologies, Inc., June 2018.

[25] Qualcomm, *Private LTE networks create new opportunities for industrial IoT*, Qualcomm Technologies, Inc., October 2017.

[26] P. Gaj, J. Jasperneite and M. Felser, "Computer Communication Within Industrial Distributed Environment-a Survey", in *IEEE Transactions on Industrial Informatics*, vol. 9, no. 1, pp. 182–189, February 2013.

[27] D. De Guglielmo, S. Brienza, G. Anastasi, "IEEE 802.15.4e: A survey", *Computer Communications*, vol. 88, pp. 1–24, 2016.

[28] T. Watteyne et al., "Industrial Wireless IP-Based Cyber –Physical Systems", *Proceedings of the IEEE*, vol. 104, no. 5, pp. 1025–1038, May 2016.

[29] J. Heo, J. Hong and Y. Cho, "EARQ: Energy Aware Routing for Real-Time and Reliable Communication in Wireless Industrial Sensor Networks", *IEEE Transactions on Industrial Informatics*, vol. 5, no. 1, pp. 3–11, February 2009.

[30] D. Kim et al., "Minimum Data-Latency-Bound k-Sink Placement Problem in Wireless Sensor Networks", *IEEE/ACM Transactions on Networking*, vol. 19, no. 5, pp. 1344–1353, October 2011.

[31] C. Antonopoulos, C. Panagiotou, G. Keramidas and S. Koubias, "Network driven cache behavior in wireless sensor networks", in *Proc. of the 2012 IEEE International Conference on Industrial Technology*, Athens, pp. 567–572, 2012.

[32] C. Panagiotou, C. Antonopoulos and S. Koubias, “Performance enhancement in WSN through data cache replacement policies”, in *Proc. of 2012 IEEE 17th International Conference on Emerging Technologies & Factory Automation (ETFA 2012)*, Krakow, pp. 1–8, 2012.

[33] A. Saifullah, Y. Xu, C. Lu and Y. Chen, “End-to-End Communication Delay Analysis in Industrial Wireless Networks”, *IEEE Transactions on Computers*, vol. 64, no. 5, pp. 1361–1374, 1 May 2015.

[34] J. Li et al., “A Novel Approximation for Multi-hop Connected Clustering Problem in Wireless Sensor Networks”, in *Proc. of the 2015 IEEE 35th International Conference on Distributed Computing Systems*, Columbus, OH, pp. 696–705, 2015.

[35] 3GPP, Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); Study on latency reduction techniques for LTE, 3GPP TS 36.881, version 14.0.0, 2016.

[36] ITU-R M.2083-0, IMT Vision – Framework and overall objectives of the future development of IMT for 2020 and beyond, September 2015.

[37] Qualcomm, *Private LTE networks create new opportunities for industrial IoT*, Qualcomm Technologies, Inc., October 2017.

[38] P. Gaj, A. Malinowski, T. Sauter, A. Valenzano, “Guest Editorial Distributed Data Processing in Industrial Applications”, *IEEE Trans. Ind. Informatics*, vol. 11, no. 3, pp. 737–740, 2015.

[39] C. Wang, Z. Bi, L. Da Xu, “IoT and cloud computing in automation of assembly modeling systems”, *IEEE Trans. Ind. Informatics*, vol. 10, no. 2, pp. 1426–1434, 2014.

[40] Z. Bi, L. Da Xu, C. Wang, “Internet of things for enterprise systems of modern manufacturing”, *IEEE Trans. Ind. Informatics*, vol. 10, no. 2, pp. 1537–1546, 2014.

[41] Gary Maidment, “One slice at a time: SDN/NFV to 5G network slicing”, *Communicate (Huawei Technologies),* Issue 81, pp. 63–66, December 2016.

[42] J. Ordonez-Lucena, et al., “Network Slicing for 5G with SDN/NFV: Concepts, Architectures, and Challenges”, *IEEE Communications Magazine*, vol. 55, Issue 5, pp. 80–87, May 2017.

[43] A. Checko et al., “Cloud RAN for Mobile Networks—A Technology Overview”, *IEEE Communications Surveys & Tutorials*, vol. 17, no. 1, pp. 405–426, 2015.

[44] O. Chabbouh, S. B. Rejeb, Z. Choukair, N. Agoulmine, "A novel cloud RAN architecture for 5G HetNets and QoS evaluation", in *Proc. International Symposium on Networks, Computers and Communications (ISNCC) 2016*, pp. 1–6, Yasmine Hammamet, Tunissia, May 2016.

[45] *Cloud RAN*, Ericsson White Paper, September 2015.

[46] *5G Network Architecture. A High-Level Perspective*, Huawei White Paper, July 2016.

[47] T. P. Raptis and A. Passarella, "A distributed data management scheme for industrial IoT environments", in *Proc. of the IEEE 13th International Conference on Wireless and Mobile Computing, Networking and Communications* (WiMob), Rome, pp. 196–203, 2017.

[48] T. P. Raptis, A. Passarella and M. Conti, "Maximizing industrial IoT network lifetime under latency constraints through edge data distribution", in *Proc. of the IEEE Industrial Cyber-Physical Systems* (ICPS), Saint Petersburg, Russia, pp. 708–713, 2018.

[49] T. P. Raptis, A. Passarella and M. Conti, "Distributed Path Reconfiguration and Data Forwarding in Industrial IoT Networks", in *Proc. of the 16th IFIP International Conference on Wired/Wireless Internet Communications*, Boston, MA, (WWIC), 2018.