

A Software Defined Hierarchical Communication and Data Management Architecture for Industry 4.0

M.Carmen Lucas-Estañ¹, Theofanis P. Raptis², Miguel Sepulcre¹, Andrea Passarella², Cristina Regueiro³, Oscar Lazaro³

¹UWICORE Laboratory, Universidad Miguel Hernández de Elche (UMH), Elche, Spain,
{m.lucas, msepulcre}@umh.es

²Institute of Informatics and Telematics, National Research Council (CNR), Pisa, Italy.
{theofanis.raptis, andrea.passarella}@iit.cnr.it

³Innovalia Association, Bilbao, Spain
{cregueiro, olazaro}@innovalia.org

Abstract—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, and with particular attention on time-critical automation. To this aim, this paper proposes the use of heterogeneous communication technologies integrated in 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. Industrial applications are organized in different tiers where different management strategies are applied to satisfy their different requirements in terms of latency and reliability. 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.

Keywords— *Industry 4.0; Factories of the Future; Communication Architecture; Data Management; Industrial Wireless Networks; Hierarchical; Decentralized; 5G.*

I. INTRODUCTION

Industry is experiencing a new revolution towards the development of highly efficient, connected and flexible Factories of the Future, also known as Industry 4.0 or Factory 4.0 [1]. The introduction of Cyber-Physical-Systems (CPS) and Internet-of-Things (IoT) technology in the industrial environment fuels the evolution towards intelligent digitalized and networked factories where components and machines can talk to one another [2], and will have self-optimizing, self-configuring and self-diagnosing abilities. Moreover, in the context of the 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 (Tactile Internet) reactions of increasingly connected mobile and autonomous robots and vehicles. This is also expected to revolutionize the OT (shopfloor) and IT (Office Floor) infrastructure and interoperability supporting production environments. 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. The prevailing ISA95 function-oriented IT architecture lacks flexibility, holistic data integration and cross-hierarchical information provisioning, which limit the exploitation of big industrial data. This paper will therefore advocate for a software defined platform and edge-powered aligned with the Reference Architecture for Manufacturing Industry 4.0 (RAMI 4.0) [4] which is able to leverage a multi-layer, largely decentralized, dynamically controlled factory IT system operation, where intelligence is spread among the manufacturing execution system, controllers, and centralized digital manufacturing cloud platforms.

Two technological enablers of the 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]. These enablers respond to the Information and Communication Layers of the software architecture adopted by RAMI 4.0 reference architecture [4]. 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-efficient and ultra-low communication costs [6]. The varying and stringent communication and data availability requirements of the industrial applications pose an important

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

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 proposed in this paper is the reference communications and data management architecture for the H2020 AUTOWARE project [7], whose objective is to build an open consolidated ecosystem that lowers the barriers of Small, Medium & Micro-sized Enterprises for cognitive automation application development and application of autonomous manufacturing processes. 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 spectrum. From a data management standpoint, real-time data availability requirements, optimized utilization of IT resources (particularly for SMMEs) as well as 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.

The paper is organized as follows. Section II reviews communication architectures proposed for Industrial Wireless Networks. Section III presents the requirements imposed by the industrial environment to the communications and data management system. Section IV presents the proposed communications and data management architecture. Section V describes the technological enablers considered to build up the proposed architecture, RAN Slicing and Cloud RAN. Section VI presents an example of early adoption of communication

and data management concepts supported by the suggested architecture. Section VI summarizes and concludes the paper.

II. RELATED WORK

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 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, autonomous manufacturing system operation 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 in plug & produce scenarios. Operating in unlicensed frequency bands, WirelessHART, ISA100.11a and IEEE802.15.4e 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 Layer) layers, and share some fundamental technologies and mechanisms, e.g., a centralized network management and Time Division Multiple Access (TDMA) combined with Frequency Hopping.

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 and distributing management decisions to end devices, limits the reconfiguration and scalability capabilities of networks with centralized management, as highlighted in [8] and [9]. To overcome this drawback, [9] proposed to divide the network into multiple subnetworks. Each subnetwork has its own manager that deals with the wireless dynamics within its subnetwork. A global entity is in charge of the management of the entire network and coordinates with the subnetwork managers. A hierarchical approach is also proposed in [10] to improve the flexibility and scalability of Industrial Wireless Networks. Similarly to [9], the approach proposed in [10] considers the deployment of one or more wireless sensor networks in the lowest level of the architecture connecting sensors and actuators. The deployed devices collect data and send it towards a central control and management system, which is located at the highest level of the network architecture. Located at an intermediate level of the architecture, sink nodes collect and aggregate the sensed data, and transmit the aggregated information to the central control and management system through backhaul links.

Recent proposals for Industrial Wireless Networks rely on hierarchical management architectures and propose the integration of heterogeneous technologies; the requirement of using heterogeneous technologies in manufacturing processes was already highlighted by ETSI (ETSI TR 102 889-2). For example, the work in [11] proposes a high-level architecture that allows wireless sensor networks (WSN) that can implement different communication technologies to exchange information among them via standard interfaces. Each WSN

has its own Gateway that is in charge of the management and protocol translation. The use of resources at different WSN is coordinated by a higher-level gateway that also provides protocol translation functionalities for the WSN under its support. Interfaces, services and interoperability features of gateways are described in [11]. However, [11] is focused on IoT systems and provides connectivity to a massive number of communication devices, and it does not particularly consider applications with very stringent latency and reliability requirements. A two-tier management architecture for radio resource coordination is proposed in [12] to support mission-critical wireless communications. [12] considers an industrial environment covered by multiple small cells to guarantee the capacity and scalability requirements. Each of these small cells can implement a different wireless technology, and has a Local Radio Coordinator that is in charge of the fine-grained management of radio resources for devices in its cell. In a higher level, there is a single Global Radio Coordinator 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 [12], control management is carried out in a centralized mode at the radio coordinators. For the data plane, centralized and assisted D2D (device-to-device) modes are considered within each cell.

Architectures proposed in [11] and [12] are mainly designed to guarantee communication requirements of a given type of service (to provide connectivity to a massive number of devices in [11], and mission-critical wireless communications in [12]). By contrast, this work focusses on the design of a flexible communication architecture and data management strategies that are able to efficiently meet the varying and stringent communication requirements of applications and services coexisting within the factories of the future. In addition, this work goes a step further and analyses the requirements of the communication architecture from the point of view of the data management and distribution.

III. 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 representing a different subset of communication requirements in terms of latency, reliability, availability, throughput, etc. [6]. The identified families include from time-critical control loops with sensors transmitting with low-bitrates, ultra-low latency and ultra-high reliability requirements to non-critical quality control that presents more relaxed latency requirements but may require high-bandwidth communications. Table I shows the communication requirements of three relevant examples (extracted from [5]) that illustrate the range of diverging and stringent communications requirements imposed by Industry 4.0. ETSI divides automation systems into three classes (manufacturing cell, factory hall, and plant level) with different needs in terms of latency (from 5 ms to 20 ms), as presented in (ETSI TR 102 889-2). However, all three classes require a 10^{-9}

packet loss rate and a 99.999% application availability. The diverging communication requirements demanded by Industry 4.0 applications are also highlighted in [13]: a stringent latency bound of 1ms is imposed to transmit commands from a Programmable Logic Controller (PLC) to a robot with a data rate lower than 100 kb/s, while 1-100 Mb have to be transmitted per image from a camera to a 3D visualization system tolerating a maximum 5ms latency.

Due to the fact that the application functions should be applicable to different types of resources, they cannot rely only on specific communication functions [14], but include additional functions like smart data distribution and management. However, it should be worth noting that the ultimate Industry 4.0 application performance is the result of the concurrent operation and synergies across communication architectures and data distribution strategies. Table II 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 (extracted from [15][16]). 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

TABLE I. COMMUNICATION REQUIREMENTS FOR SOME INDUSTRIAL APPLICATIONS [5]

	<i>Motion Control</i>	<i>Condition Monitoring</i>	<i>Augmented Reality</i>
Latency/Cycle Time	250 μ s – 1ms	100ms	10ms
Reliability (PER)	1e ⁻⁸	1e ⁻⁵	1e ⁻⁵
Data Rate	kbit/s – Mbit/s	kbit/s	Mbit/s – Gbit/s

TABLE II. ADDITIONAL REQUIREMENTS FOR DIFFERENT APPLICATION SCENARIOS [15][16]

	<i>Desired value</i>	<i>Application scenario</i>
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

focus of the Industrial Data Space (IDS) initiative; one that this paper is fully aligned and supporting of [17].

IV. COMMUNICATIONS AND DATA MANAGEMENT ARCHITECTURE

The network architecture we propose is designed to provide flexible and efficient connectivity and data management in Industry 4.0. This work proposes a hierarchical and multi-tier management architecture that supports the use of heterogeneous communication technologies to efficiently meet the different requirements in terms of communications and data distribution of industrial applications.

A. Heterogeneous communication technologies

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 them. For example, unlicensed wireless technologies such as WirelessHART, ISA100.11a or IEEE802.15.4e 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 (3GPP TS 36.881) 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 Ultra Reliable Low Latency Communications (URLL) with a latency of about 1ms and reliability of 1-10⁻⁹ [18]. Also Private LTE and Private 5G networks will be relevant for the industrial environment [19], since they will allow the implementation of local networks with dedicated radio equipment using shared and unlicensed spectrum, as well as locally dedicated licensed spectrum. As a complement of wireless technologies, the use of wired communication technologies can be considered for links between static devices. For example, the IEEE Time-Sensitive Network (TSN) WG currently aims to improve the reliability and real-time capabilities of standard Ethernet, targeting guaranteed data rates and latencies even under high-load traffic conditions.

In this context, this work proposes that several subnetworks or cells (we will use the term cell throughout the rest of the paper) 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,

and can overlap in space. 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. Fig. 1 illustrates the concept of cells in the proposed heterogeneous architecture. 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 (roles of LMs and the Orchestrator are presented in next section), and it could be implemented with TSN¹.

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. Cells coordination need to be carried out dynamically based on the dynamic and changing nature of industrial environment.

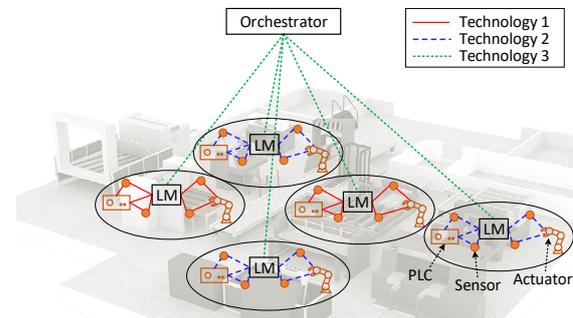


Fig. 1. Hierarchical and heterogeneous communications and data management architecture.

B. Hierarchical communications and data management

The proposed architecture considers a hierarchical structure that combines local and decentralized management with centralized decisions to efficiently use the available communication resources and carry out the data management. The management structure is depicted in Fig. 1, and the role of the Orchestrator and the LMs are next described.

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

¹ Communication links between LMs and the Orchestrator could also be implemented by a multi-hop link using heterogeneous technologies for improved flexibility and scalability (for example, IEEE 802.11 and TSN).

comply with in order to guarantee coordination and interworking of different cells, and finally guarantee the requirements of the industrial application 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, e.g. by dynamically allocating non-interfering channels to different cells based on the current demand.

LMs are implemented at each cell. A 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 Fig. 2, LMs are in charge of management functions such as Radio Resource Allocation, Power Control or Scheduling. These functions 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 and will decide to which cell a new device is attached to based on its communication capabilities of the device, the communication requirements of the application, and the current operating conditions of each cell.

We consider that control plane and user plane are separated. Therefore, although a centralized management is adopted, 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 Fig. 2).

2) Decentralized data management and distribution

On the one hand, 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 (see Fig. 2).

On the other hand, it is widely recognized that entirely centralized solutions to collect and manage data in industrial environments are not always suitable [20][21]. 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 [22][23].

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 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

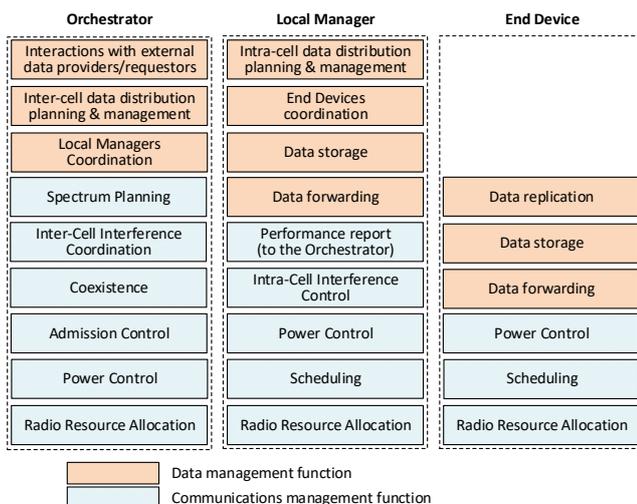


Fig. 2. Different communication and data management functions reside in different entities of the hierarchical architecture.

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 exploit devices which lie between source and destination devices, like the use of proxies for data storage and access [24], discussed in more detail in Section VI.

C. Multi-tier organization

In the proposed 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 eNB) and grant (from the 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 underutilization of radio resources. The different requirements in terms of latency and reliability of the application supported by a cell also affects 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 Fig. 3).

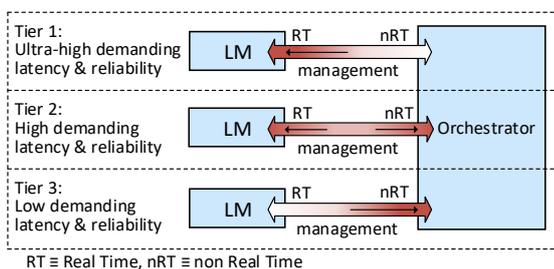


Fig. 3. LM-Orchestrator interaction at different tiers of the management architecture.

V. VIRTUALIZATION AND SOFTWAREIZATION

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) [25]. In this context, the communication and data management architecture proposed in this paper considers the use of RAN Slicing and Cloud RAN as enabling technologies to achieve the sought flexibility and efficiency.

A. 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 can be used to satisfy the ultra low-latency and high reliability communications of a time-critical automation process. The same networks could also support applications that require high throughput levels, e.g. 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.

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 proposes to split the resources and management functions of a RAN in different slices to create multiple logical (virtual) networks on top of a common network [26]. 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. One of the main objectives of RAN Slicing is to assure isolation in terms of performance [26]. 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 architectures capable to meet the stringent and diverging communication requirements of industrial applications, and in particular, those of URLLC.

Each slice of a physical cell is referred to as virtual cell in this work (Fig. 4). Virtual cells resulting from the split of the same physical cell can be located at different levels of the multi-tier architecture depending on the communication requirements of the applications. Each virtual cell implements the appropriate functions based on the requirements of the

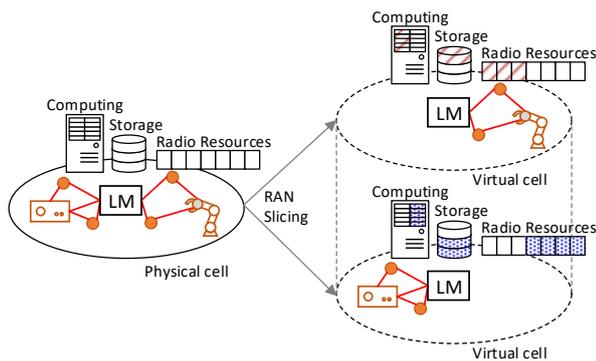


Fig.4. Virtual cells based on RAN Slicing.

application supported and must be assigned the RAN resources required to satisfy the requirements of the communication links it supports. The amount of RAN resources (e.g., data storage, computing, radio resources, etc.) allocated to each virtual cell must be dynamically adapted based on the operating conditions such as the amount of traffic, or the link quality. 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, 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 to optimize slicing itself.

B. 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 [27]. 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). 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) [27]. 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. Cloud RAN can support different functional

splits that are perfectly aligned with the foreseen needs of industrial applications [28]; 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 taken by the Orchestrator considering the communication requirements of the applications supported by each cell.

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 [27], which results 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.

VI. DECENTRALIZED DATA MANAGEMENT

The suggested architecture can be used in order to efficiently deploy 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. For example, as shown in [24], by using a subset of the data management functions coming from our architectural design, the industrial network operator is able to significantly improve the data access delay. More specifically, in large scale networks of sensing and actuating nodes, given a set of data, the sets of nodes generating and requesting them, and a maximum data access delay L_{\max} that requesting nodes can tolerate, the LM (or the Orchestrator) can efficiently identify a limited set of proxies in the network where data should be stored. Given the mentioned constraints, the computationally difficult problem of finding which network nodes to select as proxies can be solved at the LM using appropriate heuristics. Then, the proper assignment of requesting nodes to corresponding proxies can guarantee that the average access delay in the network stays below the given threshold. In fact, this kind of method can significantly outperform both entirely centralized and distributed approaches (which do not typically take into account maximum access delay thresholds), both in terms of access latency and in terms of maximum delay guarantees. For example, in Fig. 5 a simulative performance comparison is displayed. We compare (i) a decentralized data

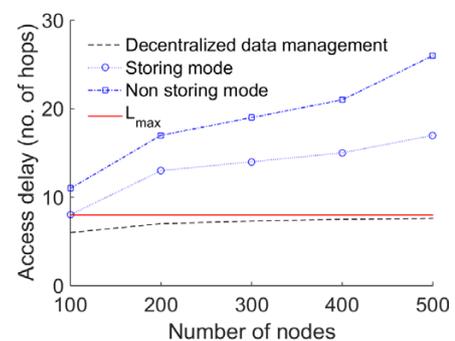


Fig. 5. End-to-end latency (number of hops) for three alternative solutions (simulation under ideal conditions [24]).

management approach which adopts the suggested architecture, (ii) a centralized approach where a single node (the LM being the natural candidate for this) stores all the data and serves all data requests coming from all nodes (referred to as non-storing mode), and (iii) a distributed approach which routes the data through some intermediate nodes (between providers and requestors) defined as the lowest common ancestors of the routing tree, assuming nodes are topologically organized according to a tree structure (referred to as storing mode). Results show that our decentralized method, where the LM decides which nodes should act as proxies for data storage, is able to guarantee the delay requirements of the applications, and significantly outperforms both a totally centralized approach, and also a distributed approach where data replication does not take into account the specific data generation and request patterns of the application.

VII. CONCLUSIONS

A software defined heterogeneous, hierarchical and multi-tier communication management architecture with edge-powered smart data distribution strategies has been proposed in this paper 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 (to optimize performance), but distinct (to achieve modularity and manageability of the architecture components) 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.

REFERENCES

- [1] European Factories of the Future Association (EFFRA) "Factories 4.0 and Beyond", September 2016.
- [2] M. Wollschlaeger et al., "The Future of Industrial Communication: Automation Networks in the Era of the Internet of Things and Industry 4.0", *IEEE Industrial Electronics Mag.*, vol. 11, pp. 17-27, March 2017.
- [3] *How Audi is changing the future of automotive manufacturing*, Feb. 2017. Available at <https://www.drivingline.com/>. Access on 2017/12/01.
- [4] *Reference Architecture Model Industrie 4.0 (RAMI4.0), Status Report*, VDI/VDE Society Measurement and Automatic Control (GMA), April 2015. Available at: <https://www.zvei.org/>. Last access on 2017/12/01.
- [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*, Oct. 2015.
- [7] H2020 AUTOWARE project website: <http://www.autoware-eu.org/>.
- [8] S. Montero et al., "Impact of Mobility on the Management and Performance of WirelessHART Industrial Communications", *Proc. 17th IEEE Int. Conf. on Emerging Technologies & Factory Automation (ETFA)*, Kraków (Poland), 17-21 Sept. 2012.
- [9] 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.
- [10] 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.
- [11] R. Sámano-Robles, et al., "The DEWI high-level architecture: Wireless sensor networks in industrial applications", in *Proc. 11th Int. Conf. on Digital Information Management (ICDIM)*, Porto, 2016, pp. 274-280.
- [12] I. Aktas, et al. "A Coordination Architecture for Wireless Industrial Automation", in *Proc. European Wireless Conference*, Dresden, Germany, May 2017.
- [13] E. Molina, et al., "The AUTOWARE Framework and Requirements for the Cognitive Digital Automation", in *Proc. 18th IFIP Working Conf. on Virtual Enterprises (PRO-VE)*, Vicenza, Italy, Sept. 2017.
- [14] M. Wollschlaeger, T. Sauter, J. Jasperneite, "The Future of Industrial Communication: Automation Networks in the Era of the Internet of Things and Industry 4.0", *IEEE Industrial Electronics Magazine*, vol. 11, no. 1, pp. 17-27, March 2017.
- [15] A. Varghese, D. Tandur, "Wireless requirements and challenges in Industry 4.0", in *Proc. 2014 International Conference on Contemporary Computing and Informatics (IC3I)*, 2014, pp. 634-638.
- [16] 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.
- [17] *Reference architecture model for the industrial data space*, Fraunhofer. Available at <https://www.fraunhofer.de>. Last access on 2017/12/01.
- [18] ITU-R M.2083-0, IMT Vision – Framework and overall objectives of the future development of IMT for 2020 and beyond, Sept. 2015.
- [19] Qualcomm, *Private LTE networks create new opportunities for industrial IoT*, Qualcomm Technologies, Inc. Oct. 2017.
- [20] 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.
- [21] 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.
- [22] 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.
- [23] P. Gaj, J. Jasperneite, M. Felser, "Computer Communication Within Industrial Distributed Environment - a Survey", *IEEE Trans. Ind. Informatics*, vol. 9, no. 1, pp. 182-189, 2013.
- [24] T. P. Raptis, A. Passarella, "A Distributed Data Management Scheme for Industrial IoT Environments", in *Proc. 4th IEEE International Workshop on Cooperative Wireless Networks (CWN)*, 2017, Rome, Italy.
- [25] Gary Maidment, "One slice at a time: SDN/NFV to 5G network slicing", *Communicate (Huawei Technologies)*, Issue 81, pp. 63-66, Dec. 2016.
- [26] 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.
- [27] A. Checko et al., "Cloud RAN for Mobile Networks—A Technology Overview", *IEEE Communications Surveys & Tutorials*, vol. 17, no. 1, pp. 405-426, Firstquarter 2015.
- [28] *Cloud RAN*, Ericsson White Paper, Sept. 2015.