A Highly Scalable IoT Architecture through Network Function Virtualization

Igor Miladinovic, Sigrid Schefer-Wenzl

IT and Telecommunication, University of Applied Sciences Campus Vienna, Favoritenstrasse 226, Vienna, Austria,
{igor.miladinovic, sigrid.schefer-wenzl}@fh-campuswien.ac.at

ABSTRACT

As the number of devices for Internet of Things (IoT) is rapidly growing, existing communication infrastructures are forced to continually evolve. The next generation network infrastructure is expected to be virtualized and able to integrate different kinds of information technology resources. Network Functions Virtualization (NFV) is one of the leading concepts facilitating the operation of network services in a scalable manner. In this paper, we present an architecture involving NFV to meet the requirements of highly scalable IoT scenarios. We highlight the benefits and challenges of our approach for IoT stakeholders. Finally, the paper illustrates our vision of how the proposed architecture can be applied in the context of a state-of-the-art high-tech operating room, which we are going to realize in future work.

TYPE OF PAPER AND KEYWORDS

Visionary paper: *IoT Architecture, Network Functions Virtualization, NFV, Software Defined Network, SDN, Large Scale Environments, 5G, Mobile Edge Computing*

1 INTRODUCTION

Over the past few years, we have witnessed an immerse success of connected devices in both consumer and business market segments and across multiple industries. In [9] Gartner Inc. predicts the number of connected devices to grow from 6,392 million in 2016 to 20,797 million in 2020. An interesting aspect of that growth is that the major part of newly connected devices is expected in the consumer market (from 4,024 million in 2016 to 13,509 million in 2020), followed by the cross industry segment (1,092 million in 2016 to 4,408 million in 2020), where - with more than four times more devices

- we can expect the highest growth. These are also called Internet of Things (IoT) devices and play an essential role for the success of IoT in general.

There are multiple other predictions on the amount of IoT devices (for example by Nokia [17], Ericsson [6] or Cisco [7]) and they all show in the same direction – a massive growth of the number of IoT devices in the next years resulting in multiple tens of billions connected devices by 2020.

However, such an amount of new IoT devices does not go without new challenges. One of them is scalability of the IoT architecture, given that we need to design a new IoT architecture tailored simultaneously for current and future devices. Another challenge is the creation of a new, converged access architecture, capable of serving people and things optimally [25]. Considering all these new devices, we need to keep in mind potential security issues. The maintainability of the architecture will play

an important role, including the possibility to easily keep the architecture up to date.

In this paper we present an approach for a scalable and maintainable IoT architecture. The architecture benefits from applying proven concepts of the telecommunication industry on IoT, in particular Network Function Virtualization (NFV) and Software Defined Networks (SDN) [12] [16]. We start with a short introduction into the related work on applying SDN and NFV technologies to IoT architectures in Section 2. Thereafter, we introduce our architecture explaining the benefits and the challenges to be considered in Sections 3 and 4. In Section 5, we present our project at the University where we are building up a modern operating room including all the equipment providing us with the opportunity to realize and evaluate our proposed IoT architecture. Finally, we conclude the paper with our major findings in Section 6.

2 NFV AND SDN FOR IoT

NFV (Network Function Virtualization) is a very prominent technology in telecommunication networks today. The basic idea of NFV is to virtualize and centralize functions in the network, such as the IP Multimedia Subsystem (IMS) or the Evolved Packet Core (EPC). These network functions are running in a data center on a standard, commercial off-the-shelf (COTS) hardware, instead of distributed, proprietary hardware. The common hardware is shared among different network functions, fully implemented in software. The main benefit of NFV is an elastic and scalable architecture, enabling not only savings in capital expenditures (CAPEX) and operational expenditure (OPEX), but also a revenue increase for the network operator [2].

Besides the virtualization of components within the network, it is also possible to virtualize components at customer premises. An example is virtual Customer Premises Equipment (vCPE) [14]. With vCPE the components residing at customer premises are very simple and all the complex application logic is moved to the network operator's data center, also called operator's cloud. This extends the life time of the equipment on customer premises and ensures that all CPE devices are running with the newest software releases.

NFV is a stand-alone technology. However, it can be combined with SDN (Software Defined Network) generating several synergy effects [5]. For example, with the centralized network intelligence provided by SDN it is possible to dynamically move a network function from one data center to another without any interruption of service. Furthermore, network resources needed to meet some critical parameters (e.g. data rate or latency) can be allocated dynamically by SDN on request of a network function or an application.

There are several approaches to apply NFV and SDN technologies to IoT. Some technical challenges and an early work towards an SDN based IoT framework with NFV have been presented in [11]. Bizanis and Kuipers give a good survey of some other concepts in [3], focusing mainly on SDN benefits to manage the expected IoT data explosion and to improve security. Another SDN based IoT architecture is presented in [22], following the distributed data and centralized control idea to overcome scalability and security challenges.

Ojo et al. show in [18] a concept of an evolution of the IoT platform, starting with a traditional IoT architecture, over an SDN-IoT architecture to an IoT architecture including both SDN and NFV. They show that the IoT architecture benefits from SDN and NFV because of reducing CAPEX and OPEX, and also by becoming a more flexible architecture, faster to introduce innovative services.

Infrastrucure agility and software-oriented innovations have also been identified as the major benefits of SDN and NFV for IoT in [19], in addition to scalability, privacy and costs savings. Finally, Mouradian et al. presents in [15] a case study of NFV based gateways for virtual wireless sensor networks to achieve an elastic and scalable deployment of gateway modules and also operational and maintenance cost savings.

Summarized, the potential of multiple aspects has been identified by applying SDN and NFV – which are currently reshaping telecommunication industry – to an IoT environment. One of the major challenges for IoT today, to design an architecture which supports the expected IoT devices explosion in the next years, can only be managed by a flexible and scalable architecture. Both of these requirements are in the focus of SDN and NFV.

3 HIGHLY SCALABLE IoT ARCHITECTURE

In Section 2 we have seen several approaches to apply NFV and SDN on an IoT architecture. The majority of them focuses on cost savings, flexibility to manage the IoT traffic increase and to enable faster service innovation. We are also going to address security benefits, given that it is less complex to apply and keep up to date modern security procedures on a centralized architecture than on each single device. Moreover, we want to emphasize benefits such as scalability and maintainability, but also – because of moving away the complexity from the IoT gateways and end devices –
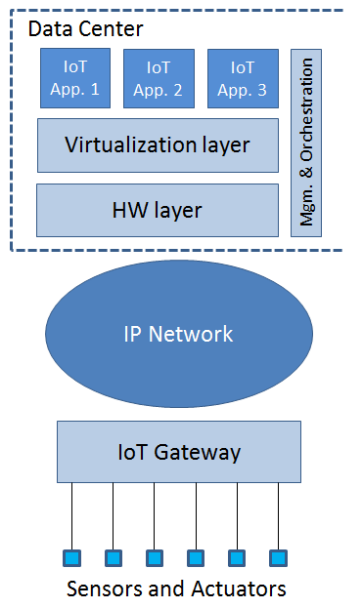
Figure 1: The proposed IoT architecture

Figure 2: Interactions between an IoT application and an IoT gateway

economies of scale.

In this section, we are presenting an architecture, which we are going to implement and apply on a concrete use case described in Section 5. Our goal is to design a scalable architecture in-depth enough to allow the mentioned implementation.

Figure 1 depicts our proposed architecture. The IoT gateway runs on a simple hardware and does not execute any application logic. The function of the IoT gateway is the translation between different protocols towards simple IoT devices on its southbound interface (e.g. ZigBee or Bluetooth) and the IP protocol on its northbound interface. All the application logic is located in a data center – following the NFV concept – and runs on standard hardware, shared among all the applications for different network functions. Consequently, the IoT application is seen as a network function and centralized in a data center. The functions of the different layers in the data center are as follows:

HW layer is composed of COTS hardware components, including storage, network interfaces and CPUs. It provides a scalable and elastic hardware platform shared among all IoT applications and other network functions.

Virtualization layer is an abstraction layer. Using the physical hardware resources it provides virtual machines towards the IoT applications.

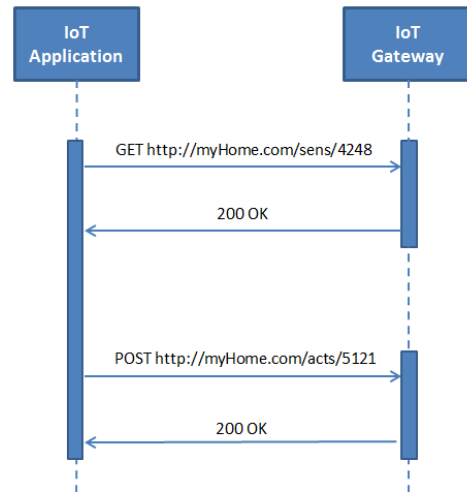**Management and orchestration layer** is responsible not only for the lifecycle management of IoT

applications, but also for the coordination of the resources and different IoT applications.

3.1 Northbound Interface of the IoT Gateway

The interface between the application logic and the IoT gateway is responsible for the exchange of monitoring and control information. It needs to be:

1. *Scalable*, in order to support a large number of IoT gateways and end devices,

2. *Flexible*, in order to be able to easily adapt to different needs of IoT applications, today and in future,

3. *Fast*, as the latency between application logic and IoT gateways can influence user experience with the complete system, and

4. *Simple*, in order to enable low complexity of IoT gateways.

We employ *Representational state transfer (REST)* or Restful Web Services [8] as the interface between the application and the IoT gateway. The main benefit of REST API for the proposed architecture is that there is no need to manage any states in the IoT gateway. Depending on the type of sensor, it either stores the latest value received from the sensor or actively queries the current value from the sensor, and forwards these data when requested by the application. Based on the received data, the application is able to place a command to the IoT gateway, which controls the actuators subsequently. With REST API we are able

Listing 1: 200 OK Response

```
HTTP/1.1 200 OK
...
{
  "id": 4248,
  "value": 22,
  "unit": "Celsius"
}
```

Listing 2: POST Request

```
POST http://myHome.com/acts/5121
...
{
  "id": 5121,
  "value": "+2"
}
```


Figure 3: The proposed IoT architecture with 5G

to keep the complexity of the IoT gateway on a low level, mainly acting as a translator between different layer 2 protocols on its southbound interfaces and the IP protocol (with REST API) on its northbound interface.

Let us consider an example shown in Figure 2. With the first *GET* request, the application is querying a value of the sensor with ID 4248. The IoT gateway either communicates periodically with the sensor and stores the latest value, or queries the current value from the sensor on request. Consequently, it is able to answer this request with the current value, provided in the body of the *200 OK* response, as shown in Listing 1.

After receiving the requested value, the application applies its logic and can decide to modify some actuators. In our example, the application identifies that the room temperature is too low, and hence it decides to modify the heater's actuator in that room. The application sends a *POST* request to advance the heater's actuator for 2 steps. Listing 2 illustrates this request. Subsequently, the IoT Gateway modifies the specified actuator and, if successful, replies with a *200 OK* response.

3.2 Towards 5G Networks

Unlike other evolution steps in mobile networks, 5G clearly goes beyond mobile Internet [10]. While the main goal of 3G and 4G was to increase data rates for end customers, research and development on 5G focusses on eight major requirements [1], all highly relevant for IoT. These requirements are: (1) up to 10 Gbps data rate in real networks, (2) round trip latency of 1 ms, (3) high bandwidth in unit area, (4) very large number of connected devices, (5) five-nines availability (99.999%),

(6) almost full coverage, (7) up to 90% reduced energy consumption, and (8) high battery live of connected devices (which is related to the 7th requirement).

Considering these requirements, in particular the 4th one, we can expect a large number of IoT devices supporting IP and be directly connected to the Internet, without the need of an IoT gateway. Our centralized approach fully supports this paradigm enabling the application to take information from and control various devices, regardless of their connectivity. In Figure 3 we demonstrate this configuration, supposing that the IoT gateway is connected via a wireline technology – for example, Fiber To The Home (FTTH) – and controls simple sensors and actuators as illustrated in Figure 1. Other IoT devices are connected directly to the Internet via 5G wireless technology. The centralized applications (App 1 to App 3 in Figure 3) play a key role in design and implementation of user centric functionalities interacting with all relevant IoT devices.

As mentioned in Section 1, the creation of a new, converged access architecture, capable of serving people and things optimally, is a challenge. A combination of modern wireline access technologies (like FTTH or G.Fast) and wireless access technologies (like 5G) is expected to fulfill this challenge. Our proposed architecture is fully supporting that converged access architecture.

To complete 5G considerations, we need to mention an emerging technology important for 5G networks - Mobile-Edge Computing (MEC) [24]. ETSI MEC ISG has specified the framework and reference architecture for MEC [13] and in [21] Sabella et al. consider the MEC
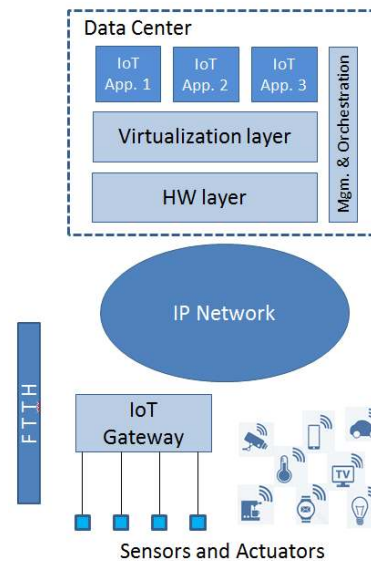
architecture with respect to IoT. The concept of MEC is to offer cloud computing functionalities within the Radio Access Network (RAN), very close to end customers. This reduces the latency for cloud applications and increases availability – as the same application can be deployed several times on different RANs. Therefore, besides on-site redundancy, geographical redundancy can be easily supported.

Our approach directly benefits from the MEC architecture. The IoT application can be installed in RAN, near to end customers, providing even better user experience by reduced latency.

4 Benefits and Challenges

In this section we consider different IoT aspects relevant for the implementation of our approach.

Scalability

Compared to a traditional IoT architecture, scalability represents one of the main benefits of the proposed approach. Keeping the IoT gateways simple and without application logic enables them to be highly scalable supporting the rising number of IoT end devices, without the need to be replaced. This also facilitates the integration of future IoT applications. At the same time, energy efficiency is improved, due to moving the computation processes from the IoT gateways towards the centralized application.

Maintainability

IoT is a rapidly changing area and the capability for updates and upgrades is important for IoT gateways. In a traditional, decentralized architecture all the gateways have to be updated separately, for example by a centralized management system. If this upgrade process is interrupted – for example by the user – irreparable damages on the IoT gateway are possible.

In our approach, updates and upgrades are centralized and decoupled from any user interaction. Moreover, it ensures that all IoT gateways are running with the latest application releases.

Security

The main focus of IoT vendors is often primarily on new functionality. Security topics gain on importance after some vulnerabilities are exploited. A prominent example is the Jeep hack from 2015 [20]. Vendors are solving these issues with software updates that have to be installed manually or by a management system on each single IoT gateway. As stated above, our approach ensures that all updates are applied centrally and immediately after their availability.

In our approach the communication between applications and gateways has to be secured. There are several mechanisms that can be combined with REST API to ensure authorisation of the application such as described by Cirani et al. in [4]. In addition, encryption protocols can be applied to protect sensitive data, such as TLS and IPSec.

Interoperability

Interoperability between different applications and vendors will gain on importance as the number of IoT devices grows. For example, in our case study (see Section 5) equipment from several vendors will be deployed. Each of them will have either an own IoT gateway or devices directly connected to the Internet. In a traditional IoT architecture, data exchange among the administration of different vendors' applications is not trivial. In our approach the interoperability is natively supported. As the application is centralized, it can monitor and control several gateways and IoT devices with an IP connection. The application combines all this information to select the best possible setup for the given environment. The centralized approach also reduces the error rate.

Availability

One difference to a traditional IoT architecture is that the availability of our approach depends on the connectivity. As the application is running in a data center and the IoT gateway is very simple, it is not possible to provide any advanced functionality without the connectivity.

However, the connectivity counts even today as a critical infrastructure. By providing several types of connection (wireless and wireline) this risk can be mitigated. As we have seen in Subsection 3.2, in future networks the availability will be very high, comparable with the availability of traditional telephone networks today.

Network Traffic

In our approach each interaction between the application and the IoT gateway generates network traffic. This leads to an increased load on the access network. Nevertheless, compared to some other applications, like 4k IPTV, this traffic is still significantly lower. The modern access networks, for example based on FTTH, G.fast, or 5G, provide data rates of 1 Gbps and beyond. In combination with SDN (see Section 2) it is possible to dynamically allocate the required data rate (and other parameters) for the IoT application.

Latency

Latency is also a parameter which needs to be considered critically regarding our approach. Due to the separation of the application logic from the gateway, we introduced a network latency between them. This could be an issue for some real-time IoT applications. On the other side, the execution of the application is faster than in a traditional IoT architecture, because it is running on high-performance hardware components. Moving the data center closer to the end customers (see Subsection 3.2) minimizes the network latency and we will investigate whether it can be completely compensated by faster application performance.

5 FUTURE WORK - PROJECT OPIC

In order to evaluate the feasibility of our approach, we are planning to apply the proposed IoT architecture in an innovative operating environment. As presented in Section 1, the prediction of Gartner Inc. indicates the highest growth of IoT devices in the cross industry segment. In this operating environment multiple industries congregate, including Medical Supplies, Health, Electronics, Telecommunication, Information Technology and Education. Therefore, it is a well-suited place to evaluate the high scalability of the proposed IoT architecture.

Today, operating rooms have to meet growing demands for technical infrastructure and related procedures, and are therefore increasingly complex and cost-intensive. Together with several partners from the health care sector, the Vienna Hospital Association (KAV) and the Ostbayerische Technische Hochschule Amberg-Weiden (OTH Technical University of Applied Sciences), our university is currently setting up a state-of-the-art high-tech operating room on the university campus (see Figure 4), called OPerating room Innovation Center (OPIC). It serves for educational and research purposes. In a second phase, the operating room is expanded by an intensive care unit, as illustrated in the floor plan in Figure 5. Afterwards, it is possible to view and analyze the clinical patient path from diagnostics to surgery and intensive medical care. This interdisciplinary project is funded by the City of Vienna through the "Wirtschaftsagentur Wien".

Our research infrastructure enables testing, evaluation and further development of innovative medical, ventilation/cooling, and information technology systems. Computer-assisted planning and execution of surgeries are as relevant to research as the optimization of the high energy requirements in the operating room. Innovative ventilation and cooling technology will be analyzed and further developed to minimize infection

Figure 4: The Innovation Center of FH Campus Vienna

risks. Optimization of lighting technology and high-quality video conferencing systems are further research fields.

Figure 6 shows our proposed IoT architecture in the OPIC environment. In an initial step there are two IoT gateways, responsible for lighting and ventilation/cooling, respectively. Each of them controls a number of sensors and actuators. Medical devices and video conferencing equipment, in contrast, are connected directly to the IP network. In the proposed IoT architecture, a common application (Innovation Center IoT Application) is capable of simultaneously controlling both IoT gateways, as well as other medical and video conferencing devices. Under consideration of all available information, the application is creating optimal operating room conditions in real-time. New IoT gateways and devices can be easily integrated in the proposed IoT architecture by an extension of the application.

With the introduction of SDN in the future, we also support optimized telemedicine scenarios. An example is a remote operation, where we can dynamically allocate high data rates and low latency connections, necessary for 4K video conferencing.

The findings of our research will be also integrated into our teaching activities, enhanced with innovative didactic concepts, for example as described in [23].

6 CONCLUSION

In this paper, we address the boundaries of traditional IoT architectures when facing a highly scalable network environment. We propose an architecture that separates the application logic from the IoT gateway and thereby provides a scalable and flexible IoT environment. We introduce REST API as the IoT gateway's fast and simple northbound interface. In addition, we leverage the effects of mobile edge computing to ensure low latency in the proposed architecture.
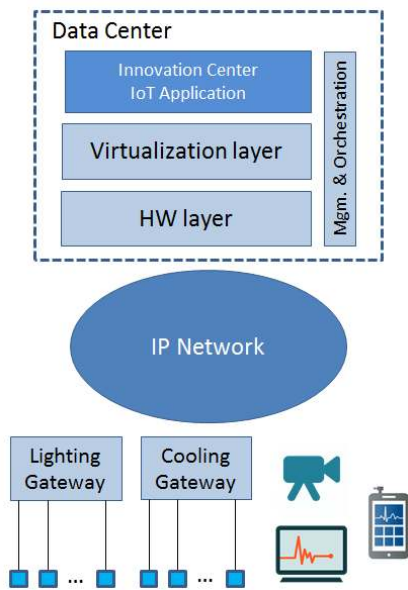
We are going to adopt our approach in a modern

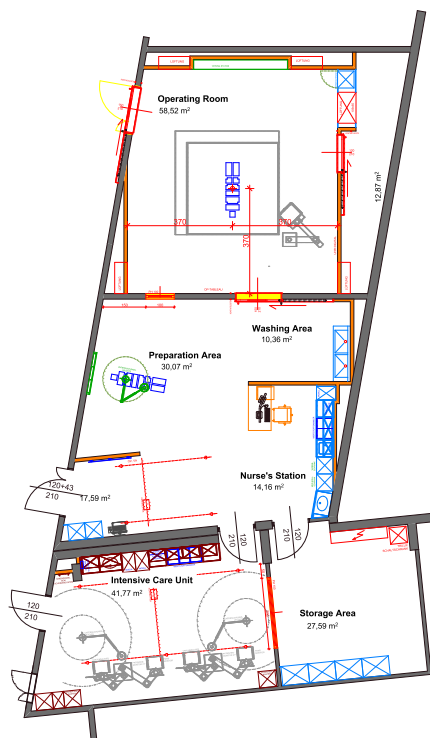**Figure 6: The proposed IoT architecture for OPIC**

Figure 5: The floor plan of the Innovation Center

operating room that is currently being built and installed at our university for teaching and research purposes.

In this use case, IoT devices from different vendors and industries will be integrated and have to work on a reliable, secure and rapid basis. We will present the results of our implementation in future work.

REFERENCES

[1] M. Agiwal, A. Roy, and N. Saxena, "Next Generation 5G Wireless Networks: A Comprehensive Survey," *IEEE Communications Surveys Tutorials*, vol. 18, no. 3, pp. 1617–1655, 2016.

[2] S. Aleksic and I. Miladinovic, "Network Virtualization: Paving the Way to Carrier Clouds," in *International Telecommunications Network Strategy and Planning Symposium*, Sept 2014, pp. 1–6.

[3] N. Bizanis and F. A. Kuipers, "SDN and Virtualization Solutions for the Internet of Things: A Survey," *IEEE Access*, vol. 4, pp. 5591–5606, 2016.

[4] S. Cirani, M. Picone, P. Gonizzi, L. Veltri, and G. Ferrari, "IoT-OAS: An OAuth-Based Authorization Service Architecture for Secure Services in IoT Scenarios," *IEEE Sensors Journal*, vol. 15, no. 2, pp. 1224–1234, 2015.

[5] J. Costa-Requena, J. L. Santos, V. F. Guasch, K. Ahokas, G. Premsankar, S. Luukkainen, O. L. Prez, M. U. Itzazelaia, I. Ahmad, M. Liyanage, M. Ylianttila, and E. M. de Oca, "SDN and NFV Integration in Generalized Mobile Network Architecture," in *European Conference on Networks and Communications (EuCNC)*, June 2015, pp. 154–158.

[6] Ericsson, "Ericsson Mobility Report, On the Pulse of the Networked Socienty," Report, Jun 2016.

[7] D. Evans, "The Internet of Things, How the Next Evolution of the Internet Is Changing Everything," Whitepaper, Cisco Internet Business Solutions Group (IBSG), Apr 2011.

[8] R. T. Fielding, "REST: Architectural Styles and the Design of Network-based Software Architectures," Doctoral dissertation, University of California, Irvine, 2000. [Online]. Available: http://www.ics.uci.edu/~fielding/pubs/dissertation/top.htm

[9] Gartner Inc., "Gartner Says 6.4 Billion Connected "Things" Will Be in Use in 2016, Up 30 Percent From 2015," http://www.gartner.com/newsroom/id/3165317, Nov. 2015.

[10] GSMA Intelligence, "Understanding 5G: Perspectives on Future Technological

Advancements in Mobile," White Paper, Dec. 2015.

[11] J. Li, E. Altman, and C. Touati, "A General SDN-based IoT Framework with NVF Implementation," *ZTE Communications*, vol. 13, no. 3, pp. 42–45, 2015.

[12] Y. Li and M. Chen, "Software-Defined Network Function Virtualization: A Survey," *IEEE Access*, vol. 3, pp. 2542–2553, 2015.

[13] MEC ETSI ISG, "Mobile Edge Computing (MEC); Framework and Reference Architecture," ETSI GS MEC 003 V1.1.1 (2016-03), 2016.

[14] P. Minoves, O. Frendved, B. Peng, A. Mackarel, and D. Wilson, "Virtual CPE: Enhancing CPE's Deployment and Operations through Virtualization," in *IEEE International Conference on Cloud Computing Technology and Science Proceedings*, Dec 2012, pp. 687–692.

[15] C. Mouradian, T. Saha, J. Sahoo, R. Glitho, M. Morrow, and P. Polakos, "NFV Based Gateways for Virtualized Wireless Sensor Networks: A Case Study," in *IEEE International Conference on Communication Workshop*, June 2015, pp. 1883–1888.

[16] T. D. Nadeau and K. Gray, *SDN: Software Defined Networks: An Authoritative Review of Network Programmability Technologies*. O'Reilly Media, 2013.

[17] Nokia, "A Buyers and Influencers Guide to Connected Device Management, for All Things Connected in Mobile, Home and IoT," Strategic Whitepaper, Jul 2016.

[18] M. Ojo, D. Adami, and S. Giordano, "A SDN-IoT Architecture with NFV Implementation," in *IEEE Globecom Workshops*, Dec 2016, pp. 1–6.

[19] N. Omnes, M. Bouillon, G. Fromentoux, and O. L. Grand, "A Programmable and Virtualized Network IT Infrastructure for the Internet of Things: How Can NFV SDN Help for Facing the Upcoming Challenges," in *International Conference on Intelligence in Next Generation Networks*, Feb 2015, pp. 64–69.

[20] P. E. Ross, "Hackers Commandeer a Moving Jeep," http://spectrum.ieee.org/cars-that-think/transportation/self-driving/hackers-take-control-of-a-moving-jeep, Jul. 2015.

[21] D. Sabella, A. Vaillant, P. Kuure, U. Rauschenbach, and F. Giust, "Mobile-Edge Computing Architecture: The Role of MEC in the Internet of Things," *IEEE Consumer Electronics Magazine*, vol. 5, no. 4, pp. 84–91, 2016.

[22] O. Salman, I. Elhajj, A. Kayssi, and A. Chehab, "An Architecture for the Internet of Things with Decentralized Data and Centralized Control," in *IEEE/ACS 12th International Conference of Computer Systems and Applications (AICCSA)*, Nov 2015, pp. 1–8.

[23] S. Schefer-Wenzl and I. Miladinovic, "A Best-Practice Mobile E-Learning Approach for Application Prototyping," in *The International Conference on E-Learning in the Workplace*, 2017.

[24] S. Wang, X. Zhang, Y. Zhang, L. Wang, J. Yang, and W. Wang, "A Survey on Mobile Edge Networks: Convergence of Computing, Caching and Communications," *IEEE Access*, vol. 5, no. 99, pp. 6757 – 6779, 2017.

[25] M. K. Weldon, *The Future X Network, A Bell Labs Perspective*. CRC Press, 2016.

AUTHOR BIOGRAPHIES

Igor Miladinovic is the head of the degree program Information Technologies and Telecommunication at the University of Applied Science Campus Vienna. He received the Ph.D. degree (with honors) in electrical engineering from Vienna University of Technology in 2003. He worked for more than 10 years on leading positions at Alcatel-Lucent (later Nokia) in area of telecommunication software and in parallel as a lecturer at two universities. His research interests cover telecommunication networks, software engineering and IoT, with over 30 publications in international journals, conferences and as book chapters.

Sigrid Schefer-Wenzl is a senior researcher and lecturer at the University of Applied Sciences Campus Vienna, WU Vienna, and the University of Salzburg. She has worked as a software analyst and developer in several companies and received the Ph.D. degree (with honors) in Information Systems from WU Vienna. Her current research and teaching activities focus on the fields of software engineering and IT-security. Sigrid has published the results of her work in top ranked journals and presented her work at various international conferences.