

Going Back to the Roots— the Evolution of Edge Computing, an IoT Perspective

Marjan Gusev
Sts. Cyril and Methodius
University

Schahram Dustdar
TU Wien Distributed
Systems Group

When the requirements and processing appetites grew sufficiently in the era of personal computing, users began to think about cloud servers. This initiated the growth of cloud computing and set up the era of Everything-as-a-Service. When the IoT idea spread, lots of sensors and smaller devices were

introduced for ubiquitous and pervasive computing. The simplest way to collect and process all this data was to connect the devices to the cloud. This required high-speed networks and interconnections. However, no matter how fast they could go, engineers understood that there's no need to transfer all the data through links to the central data server, and that it would be much more efficient to distribute smaller servers in front of the central data server and closer to the location of the user. This began the evolution of pushing data collection and processing back to end-user devices, as a new computing approach called edge computing, which appears in different forms such as fog computing, cloudlets, and mobile edge computing.

The Internet of Things (IoT) is a network of various devices that communicate with each other via the Internet. The proliferation of these devices in a communication-actuating network creates the IoT, wherein sensors and actuators blend seamlessly with the environment around us, and the information is shared across platforms.¹

Various hardware, software, data, and services are interconnected and the application domain is huge, including smart environments (smart homes, smart cities, etc.), healthcare (triage, patient monitoring, health status, etc.), emergency services (remote monitoring, resource management,

etc.), transport control (traffic management, infrastructure monitoring, etc.), and environment-related services (resource supply, environment monitoring, etc.).

The interconnected "things" have various characteristics according to their location, type, nature, power supply, and various other aspects, including, for example, characteristics of associated data that represent their state. Analyzing the location, they can be static or movable devices that need a mobile and wireless connection to enable freedom of movement—for example, wearable sensors are those that are worn on a human body and are battery operated using mobile and wireless to transmit data.

The location of a "thing" also dictates the power resource, whether it be a standard electrical power supply or battery. Battery-operated devices require small processing capabilities due to limited battery life, which raise the need to offload storage and processing to a more powerful server, such as cloud and fog computing. The data collected and processed is huge, which raises the need for offloading to a server. The offloading idea entails new architecture issues, such as where the server should be located, how to avoid long communication delays, and how to ensure low energy consumption and enable greater mobility and independence of IoT devices. This paper will analyze the evolution of communication, storage, processing, and energy consumption requirements in the IoT world. The requirements analysis is elaborated with appropriate architectural designs of computer infrastructure.

BACKGROUND AND BASIC DEFINITIONS

Sensors and actuators are essential IoT devices for production (collection) of data or control devices. Our goal is to realize small IoT devices independent of the platform, location, and environment. This can be achieved by offloading the storage and computation from the IoT devices to servers or edge devices. It will allow the IoT devices to be mobile and wirelessly connected to the environment.

In this context, edge refers to end-user devices located close to the IoT devices. Servers that communicate to edge devices, often called edge servers, are those set in proximity of the edge devices, which avoids communication latency. Examples include smaller servers at base stations or sometimes, mobile phones.

Clouds are one way to realize the offloading idea, where edge computing is an extension to clouds closer to the end-devices. Edge computing refers to the enabling technologies that allow computation to be performed at the edge of the network on downstream data on behalf of cloud services and upstream data on behalf of IoT services.²

The emergence of fog computing aims to take advantage of positioning servers close to base stations and cope with increased data traffic, and therefore, hosting services, workloads, applications, and large amounts of data at the edge of the network.^{3,4} This makes fog computing a synonym for edge computing because it facilitates the operation of computing, storage and networking services between end devices and cloud computing data centers. This approach is the communication community's answer to the rise of IT business to deliver IT-based services.

One well-known challenge of using the cloud as a server is the long latency between the mobile device and the cloud server in comparison to localized computing and small-scale distributed computing called cloudlets.⁵ In collaboration with the cloud, cloudlets make recognition and afterward send results back to the mobile device. In a cloudlet solution, mobile devices directly exchange data with the local server on the cloudlet using WiFi, instead of connecting to the cloud server.⁶

Mobile edge computing is the realization of edge computing at the base stations of mobile operators.⁷ Thus, it is a special kind of edge computing with precisely located intermediate servers.⁸ Cloudlets are also smaller servers between end-user devices and servers located on the premises of Internet providers in their LAN.

ICT PROCESSING REQUIREMENTS AND SOLUTIONS FOR IOT

Modern technology trends and available technology are pushing the commercialization and mass exploitation of various IoT sensors and remote controller devices.

Basic requirements

Let's begin the analysis with a description of a typical IoT sensor or remote controller device. An IoT sensor is a device capable of sensing data such as environmental information or digital data from social networks. It detects changes from any material (physical, chemical, or biological) and some non-material information. The sensor includes the following functions: sensing, transforming information into an electrical signal, and transmitting the electrical signal to the neighboring environment.

Some IoT sensors are connected to an IoT controller device where sufficient information that will trigger an action on the controller device. It can be any other IoT device that is not sensing and is responsible for controlling (moving) a mechanism or system, usually called an actuator.

The communication between the IoT sensor and controller device may be realized by a personal area network connection, usually realized via a direct cable, Bluetooth, or other wireless connection. This forms an independent autonomous sensing and controller system. Unless this system memorizes the previous states, there is only a limited need for processing using small amounts of memory storage.

Requirements to offload computations and storage

During the time, the requirements and processing appetites have increased and engineers have tried to realize more complex systems and connect them to the outside world. The increased complexity also requires connectivity to the outside world and more processing power and storage demands for the collected data.

The logical escape was to use the power of cloud data centers, so the next demand was to connect IoT sensors and controller devices to the Internet and cloud servers, as Figure 1 shows.

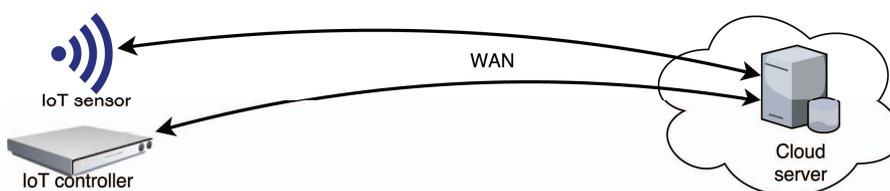


Figure 1. Requirements to connect IoT sensor and/or controller device to the cloud.

Connection to the cloud is solved rather efficiently in two ways: by Internet providers and mobile operators. The first scenario is realized by use of the Internet and local networks. End-users of IoT devices can use their Internet connection and the LAN installed at their premises. To connect to the IoT sensors, they set up a WiFi router or use a direct LAN cable connection. Figure 2 presents this idea of establishing a connection to the cloud server.

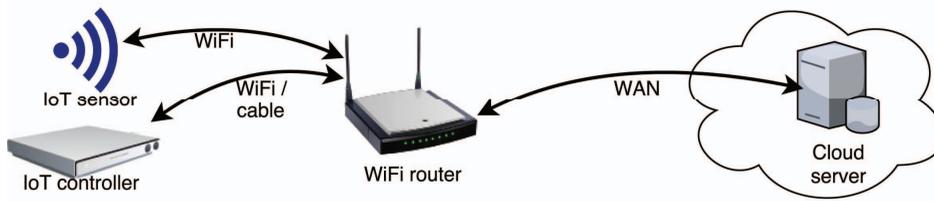


Figure 2. Requirements to set up a WiFi to connect the IoT sensor and controller device to the cloud server

This solution works efficiently while the amount of transmitted data is not too high and does not cause communication bottleneck. However, the design of IoT sensors and controller devices may use streaming data with big volume and velocity.

Another feature in the development of IoT sensors and devices was to make them mobile and wireless, free of any obstacles while moving. This means that they cannot be connected to a static electrical power supply, but, instead, that they require small batteries. So, a problem arises due to limited resources because the processing should be small enough to conserve energy and, at the same time, enable longer battery life prior to recharging.

Limited resources do not only constrain the processing, but also the use of wireless connections. The solution is to use small personal area network connections to conserve the resources of the IoT sensor and devices for as long as possible. Typical examples include Bluetooth or other related wireless network technologies. This changed the overall concept because WiFi requires a more powerful wireless connection and is not appropriate for battery-operated devices.

The solution to this problem was to create new devices with a special purpose to accompany IoT devices and function as digital signal repeaters. These devices are connected to the IoT sensor or controller device via a personal area network (such as a direct cable or Bluetooth) and can communicate with the WiFi router, as Figure 3 shows. Their only function is to receive small power signals, buffer them and then transmit to the destination over WiFi or other LAN connection.



Figure 3. IoT sensor and controller devices connected to a cloud server via a digital repeater and WiFi connections.

Some of these devices are no longer required to be mobile or wirelessly connected, but they can have a regular electrical power supply and be connected to the LAN installed at the premises where they are hosted. The other will need to recharge but their size can allow a higher battery capacity and decrease the need for frequent recharging. A typical example of a device in this category is a smartphone, which can communicate to the personal wearable sensors from one side and the WiFi from the other side.

The solution presented in Figure 3 is a typical solution of an IoT sensor or controller device to the cloud using a standard Internet connection.

Cloud-based solutions

We next explain the alternative to a direct Internet connection. Mobile operators offered a solution to this problem—instead of setting a WiFi router and Internet connection, they offered a mobile network solution. To come closer to IoT sensors, they use 3G/4G or 5G connections from their operated services. This solution requires a new device capable of accepting signals from the personal area network and connecting to the mobile operator's network. The user can use a classical mobile device and connect to the IoT devices by Bluetooth. The market also offered solutions with special dedicated devices that can communicate with Bluetooth and also accept a SIM card to connect to the mobile operator's network. Figure 4 presents a typical solution provided by the mobile operator for the purpose of IoT sensors and remote controller devices.

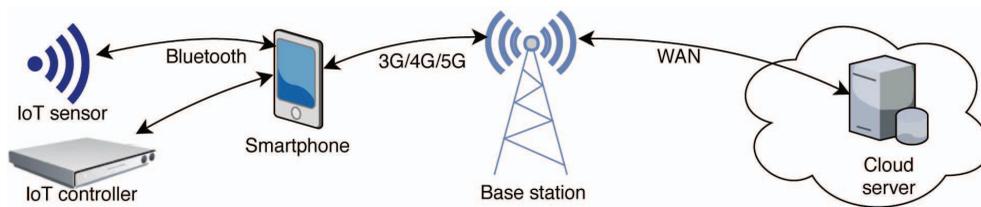


Figure 4. IoT sensor and controller devices connected to a cloud server via smartphone and mobile operator's network.

These designs are good enough if the IoT sensors do not require a lot of communication and processing. For example, in the case of measuring environmental temperature, there is no need to transmit data permanently, but at regular intervals, such as one value per minute, or less.

New technologies enabled more sophisticated sensors that permanently sense data with a high frequency. For example, a camera can sense the traffic density or an ECG sensor can sense heartbeats. The birth of big data is based on such streaming sources that provide data with high volume, velocity, variety, variability, and veracity.

The big data evolution introduced new functions in the IoT world and characteristics for IoT devices, and also addressed the need for increased storage and processing requirements. The previous design considerations, now introduce new problems such as where to store the streaming data and who will process it to trigger relevant activities on the controller device.

The existing designs of cloud-based IoT (Figure 3) or mobile operated cloud-based IoT solution (Figure 4) are not good enough. The problems occur because streaming data transmitted over the WAN will soon occupy the network throughput as the number of these connections increases and provokes server congestion. Of course, there is a solution for the increased WAN bandwidth, but the number of connected IoT sensors rises with a higher rate than the network providers can cope with.

Cloudlet and mobile edge computing

The solution for avoiding communication bottlenecks is to distribute the processing closer to the user. This is the basis of a cloudlet solution. A smaller server is hosted closer to the user, and all the storage and processing requirements are now distributed to these smaller cloudlet servers. They provide the same functions as the cloud data center, but can accept only a limited number of IoT devices and provide relevant service. The cloud data center is now relieved with high communication demands and collaborates with the cloudlet to exchange relevant information and provide extended services with added value.

Figure 5 presents the architecture of the cloudlet solution for streaming IoT devices. The IoT sensor or controller devices communicate with the digital repeater via a personal area network, using wireless Bluetooth technology or a direct cable connection. The digital repeater buffers the received signal and then transmits it to the cloudlet server via a WiFi router. The cloudlet server

collects the streaming data and processes it, providing relevant services. Only selected and processed information is sent to the cloud data center, which provides a more comprehensive analysis of the input data. This solution is a successful upgrade of the direct cloud solution due to the smaller latencies and the distribution of high throughput streaming data and processing to localized servers. It has been proven that the latencies generated by the digital repeater and WiFi router, the Bluetooth/WiFi and LAN connections, and slower cloudlet servers are smaller than the overall latencies to access the cloud data center even if it processes faster than the cloudlet server. The highest impact of this design is the decrease of high throughput demands.



Figure 5. Cloudlet solution for streaming IoT sensor and controller device.

Mobile operators have offered a similar concept to solve the high throughput demands of IoT devices (Figure 6). They host smaller servers at the premises of their base stations, and bring the storage and processing closer to the users at the edge of their network. This is known as the mobile edge computing solution.

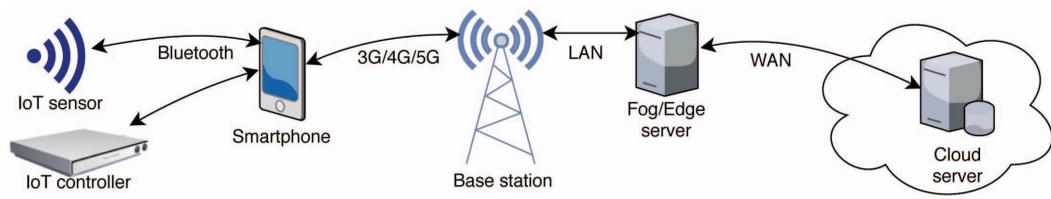


Figure 6. Mobile edge computing solution for streaming IoT sensor and controller device.

Edge computing solutions

So far, we have presented four solutions for the IoT world's requirements for computation and storage offload:

- Cloud directly connected via Internet (Figure 3).
- Mobile operated cloud solution (Figure 4).
- Cloudlet solution (Figure 5).
- Mobile edge computing solution (Figure 6).

The first and third use a kind of open environment since the user can select the Internet provider and cloud provider. The second and last solutions are proprietary for mobile operators and the user is usually in a “lock-in” situation with no possibility of choosing or changing anything. In the cloudlet solution, the user needs to install his or her own cloudlet server and pay a monthly subscription fee. In the case of mobile operator solutions, the monthly-prepaid subscription is the only alternative.

The last two designs, which solve the problem of streaming IoT data, encompass the movement of centralized services closer to the user (an implementation of the edge computing concept to push applications, data, and computing power by bringing the processing closer to the logical extremes). The architecture of cloudlets or the mobile edge computing solution assumes the establishment of an intermediate layer between the cloud server and the IoT device in the form of an edge server.

This leads back to the concept of moving the data collection and initial processing even closer to the user and closer to the place where the data is produced and used. This concept means that there will be another layer, an additional edge device, between the end-user IoT device and the intermediate layer.

The idea presented in Figure 7 changes the use of the repeater with a smaller server in a form of an edge device (sometimes called dew⁹). It can be a desktop computer, laptop, tablet, smartphone or another mobile device capable of processing the incoming data stream locally. In addition, it can store data for a limited time, and unless the data is unloaded, the old data will be deleted. It can communicate both to the lower hierarchy level (IoT devices) and higher levels (cloudlet or mobile edge computing server and cloud data center).



Figure 7. Edge computing solution using an IoT and edge devices.

The communication with IoT devices consists of reception of data and transmission of triggers. The higher levels enable interconnection and information exchange with the outside world, for example, to unload the stored data and send relevant information to the cloud data center. This edge computing design brings the data collection and processing closer to the source and end-users.

DISCUSSION

Dolui and Datta give a good comparison of related terms when they compare terms based on their implementations.¹⁰ They analyze fog computing node devices by routers, switches, or gateways, with cloudlets as data centers in the nearby LAN, and mobile edge computing as servers at the base stations of mobile operators.

Edge computing has been analyzed from various architectural approaches. Nastic et al. introduced serverless architecture for edge computing, so the idea of going closer to the roots of data producers is even more extracted.¹¹

Analysis of the best edge computing solution

Our idea was to choose the best edge computing solution to distribute processing closer to the user and data source. One would ask which solution is the best, cloudlet, fog or mobile edge computing, or a combination of these concepts. There is no straight answer to this question, so we will analyze the conditions that help solution providers choose the best.

The direct cloud solution opposed to the edge computing concept is the best alternative in the case of a static IoT sensor that senses environmental data that does not change a lot over time, such as temperature. This sensor does not stream data but rather sends information on regular time intervals and, in this case, the communication, storage and processing requirements are very small. This is especially important in the case of a regular power supply of electricity.

The edge computing solution with a cloudlet server and cloud data center is the best in the case of a static IoT sensor that streams a lot of data, such as video signal. The sensor transmits a lot of data that is stored for a limited time period on the edge device and selected data is sent to the cloudlet for further processing in collaboration with the cloud data center.

The mobile edge computing solution is a good alternative if the sensor is not static (it changes its geographic position) and there is no available Internet provider coverage. In this case, the sensor

directly communicates with the mobile operator's network where the fog/edge server will collect and process incoming data, collaborating with their cloud data center.

The mobile edge computing and cloudlet solutions are the best alternative for a movable and streaming IoT sensor and edge device. The mobile edge device may not have a continuous Internet connection. Since this edge device collects, processes and stores data, it is not necessary to have a permanent Internet connection. In this case, whenever the IoT sensor and edge device are within the range of an Internet connection, the data will be unloaded and the cloudlet/cloud server will update the current situation. It is also possible to use a 3G/4G connection to exchange data with the cloudlet/cloud server.

Table 1 gives a summary of the best solutions in various scenarios for the IoT sensor and its corresponding edge device. Based on their position, the following categories are determined:

- static location, when the IoT sensor and edge device do not move over time and have permanent energy supply;
- movable location in the case of a movable IoT sensor; and
- edge device, charged by a battery.

In our classification, the sensor streams data if the sampling frequency is higher than 1Hz at one data item per second. We understand that this value is not determined exactly, but usual values for streaming data are those with sampling frequencies higher than 100 Hz and those that transmit data less than 0.1 Hz are assumed not to be streaming.

Table 1. The Best solution in the case of static and movable IoT sensors based on data streaming and Internet availability.

Position/ Data	Streaming Data	Sending Files
Static	Edge device and cloudlet/cloud	Cloud only
Movable	Edge device and MEC	Fog only
Movable with WiFi	Edge device and cloudlet/cloud	Cloud only

In the presented categorization, we also differ between a permanent Internet connection and availability of mobile-operated network access.

A CASE STUDY: A WEARABLE ECG SENSOR

Let's analyze the ICT requirements of an existing health-related IoT solution that uses an edge device. The case study consists of a wearable ECG sensor attached to the human body and a mobile phone that acts as an edge device. The communication between the wearable sensor and the edge device is via Bluetooth and via WiFi or 3G/4G for the Internet.

The users prefer to wear their ECG sensors and monitor heart functions on the edge device (smart-phone) without any constraints that prevent their normal physical activities. They prefer mobile phones whose batteries last more than 20 hours without recharging since it is common practice to recharge the mobile phone at night while the user is sleeping.

Storage requirements

The edge device (smartphone) stores data with a limited capacity. The assumption is that the attached wearable ECG sensor samples data with a 500 Hz frequency and uses a 2-byte sample. So the edge device (smartphone) needs to store 1KB per second, or 86.4 MB per day. Unless the edge device (smartphone) does not unload and delete old data, the requirement to store the incoming data will reach 2.6 GB in one month.

Annual storage requirements will exceed 949 GB closing to the TB limits.

Communication requirements

If all data is streamed over the WiFi to higher level servers, then the demands of this system require 8Kbps transfer rate if streaming data is transmitted without identification, time, and space stamps. However, application software may involve a big overhead, for example, complex .Net methods transmit textual data instead of binary numbers.

A typical TCP/IP connection usually adds overhead due to the headers of the implemented TCP/IP levels. Considering the acknowledgements and other controlling mechanisms, we can expect this transfer rate to double.

Even a doubled transfer rate value is acceptable since the current standard cable operators offer uplink speed of at least 1Mbps. So, the communication with the cloud will not cause a substantial problem from the user's perspective.

However, a communication bottleneck may appear at the cloud data center if hundreds or thousands of users log into the cloud and concurrently stream data. The problem arises mostly at the server side, at the premises of the cloud data center.

Processing requirements

The processing requirements depend mostly on the implemented algorithms. Small pipelined ECG analysis and feature extraction algorithms may require up to 1K computations per sample, but comprehensive analysis may require more than 1M operations per sample.

Assuming that 500 samples arrive each second then the requirement will be 500K operations per second for streaming pipelined algorithms and 500M operations per second for comprehensive analysis. Theoretically, all modern smartphones can perform sufficient operations per second, for example, iPhone 6 reaches 2 Gflops/s, and Samsung Galaxy S6 even higher.

However, if hundreds or thousands of ECG sensors stream data to a cloud center, then a bottleneck should be expected at the premises of the cloud data center.

Energy consumption

So far, we have analyzed the ICT requirements and concluded that the edge device (smartphone) can provide sufficient capacity to cope with them. However, the real problems start with an analysis of energy consumption.

Smartphone producers claim that the battery of their models from 2016 have a capacity of at least 1.800 mAh and can play video for at least 10 hours until they run out of battery. However, in terms of 3D gaming, they can only survive about two-and-half hours. Both these examples show that in case of increased processor activity, the battery life is not sufficient for a careless 24 hour use. A careful analysis shows that most of the energy is consumed by the screen, which is not required in the case of a monitoring application. The screen will be used only if the user likes to monitor the activities, or in the case of an emergency. Several tests show that the transmission of the ECG signal via WiFi or 3G/4G can be realized with available battery capacity. However, the problem is the software solution and the implemented algorithm to detect the heartbeat and diagnose the heart function. An algorithm using a comprehensive analysis of the ECG signal will require too much processor activity that will drain the battery life in at least 4 hours (less than 3D video gaming). So there must be a compromise, and the developers have to code software with highly efficient streaming algorithms.

Recommendations

Wearable IoT (ECG) sensors and edge devices (smart-phones) can be built in such a way that enables sufficient capacity for the communication, storage and processing requirements. However, the real problem arises when analyzing energy consumption of the wearable edge device.

The compromise is to decrease some of the analyzed communication, storage, and processing requirements. The first solution is to select what is to be transmitted to the cloud, the second is to select what data to store (not all data), and the last is to realize less precise monitoring algorithms.

Usually, the solution providers compromise with a combination of all these requirements or select the last one: to develop smaller monitoring capacities and transfer all data to the cloud, where a comprehensive analysis can be realized. This solution is designed to remain on the safe side, keep and transmit all data to the server, and enable essential monitoring information.

In the case of the ECG, some other relevant issues can also arise, since not all data is needed unless the heart is detected to be out of normal function. This is why the edge device will play a huge role in further development of this computing concept. Only selected data will be transmitted and stored on the cloud. This is mostly comprised of the problematic issues since data that represents normal cardiac activity is not intended for further analysis. Of course, a limited storage capacity of the cloud can be offered to store a certain time interval, such as one or past two months. The old data will be deleted unless it is marked to be saved.

CONCLUSION

In this paper, we have presented the evolution of the edge computing concept from an IoT perspective. In the past, IoT sensors and controller devices required processing that is done by connection to a cloud data center, either via WiFi or via 3G/4G or 5G. This concept successfully completes centralized processing in the cloud environment.

However, the evolution of IoT sensors with streaming data, changed the course of events. At the moment of writing communication and processing are brought closer to the user, either by establishing a cloudlet server or by a mobile edge computing server on the edge of the mobile operator network. The new edge computing concepts go a step beyond this concept, bringing the communication and processing even closer to the user, and hosting an edge device close to the IoT devices and end users. This evolution showcases how computing architecture has changed over time, first we had a centralized solution, and now we are going back to distributed end-user devices.

Finally, edge computing enables independence of constant networking, reducing processing latency and ,at the same time, allows autonomous functioning and collaboration with distant servers.

REFERENCES

1. J. Gubbi et al., "Internet of things (IoT): A vision, architectural elements, and future directions," *Future Generation Computer Systems*, vol. 29, no. 7, 2013, pp. 1645–1669.
2. W. Shi and S. Dustdar, "The Promise of Edge Computing," *Computer*, vol. 49, no. 5, 2016, pp. 78–81.
3. I. Stojmenovic and S. Wen, "The fog computing paradigm: Scenarios and security issues," *2014 Federated Conference on Computer Science and Information Systems (FedCSIS 14)*, 2014, pp. 1–8.
4. F. Bonomi et al., "Fog computing and its role in the internet of things," *Proceedings of the first edition of the MCC workshop on Mobile cloud computing*, 2012, pp. 13–16.
5. M. Satyanarayanan et al., "The case for VM-based cloudlets in mobile computing," *IEEE Pervasive Computing*, vol. 8, no. 4, 2009, pp. 14–23.

6. A. Bahtovski and M. Gusev, "Cloudlet challenges," *Procedia Engineering*, vol. 69, 2014, pp. 704–711.
7. M.T. Beck et al., "Mobile edge computing: A taxonomy," *The 6th International Conference on Advances in Future Internet (IARIA 14)*, vol. IARIA 14, 2014, pp. 48–54.
8. P. Mach and Z. Becvar, "Mobile edge computing: A survey on architecture and computation offloading," *IEEE Communications Surveys & Tutorials*, 2017.
9. M. Gusev, "A dew computing solution for IoT streaming devices," *40th IEEE International Convention on Information and Communication Technology, Electronics and Microelectronics*, vol. MIPRO 17, 2017, pp. 387–392.
10. K. Dolui and S.K. Datta, "Comparison of edge computing implementations: Fog computing, cloudlet and mobile edge computing," *Global Internet of Things Summit (GloTS 17)*, 2017, pp. 1–6.
11. S. Nastic et al., "A serverless real-time data analytics platform for edge computing," *IEEE Internet Computing*, vol. 21, no. 4, 2017, pp. 64–71.

ABOUT THE AUTHORS

Marjan Gushev is a computer scientist and professor at University Sts. Cyril and Methodius Faculty of Information Sciences and Computer Engineering. His research interests include parallel processing, computer networks, and Internet technologies. Gushev received a doctoral degree in computer science from the University of Ljubljana, Slovenia. Contact him at marjan.gushev@finki.ukim.mk.

Schahram Dustdar is a full professor of computer science, and he heads the DSG at TU Wien. His work focuses on distributed systems. Dustdar is an IEEE Fellow, a member of the Academia Europaea, an ACM Distinguished Scientist, and recipient of the IBM Faculty Award. He is on the editorial boards of *IEEE Internet Computing and Computer*. He's an associate editor of *IEEE Transactions on Services Computing*, *ACM Transactions on the Web*, and *ACM Transactions on Internet Technology*. He's the editor-in-chief of *Springer Computing*. Contact him at dustdar@dsg.tuwien.ac.at; <http://dsg.tuwien.ac.at/staff/sd/>.