# AI AS A MICROSERVICE (AIMS) OVER 5G NETWORKS

*Gyu Myoung Lee[1], Tai-Won Um[2], Jun Kyun Choi[3]*

[1]Liverpool John Moores University (LJMU), g.m.lee@ljmu.ac.uk
[2]Chosun University, Corresponding author, twum@chosun.ac.kr
[3]Korea Advanced Institute of Science & Technology (KAIST), jkchoi59@kaist.edu

**ABSTRACT**

*As data-driven decision-making services are being infused into Internet of Things (IoT) applications, especially at the 5G networks, Artificial Intelligence (AI) algorithms such as deep learning, reinforcement learning, etc. are being deployed as monolithic application services for autonomous decision processes based on data from IoT devices. However, for latency sensitive IoT applications such as health-monitoring or emergency-response applications, it is inefficient to transmit data to the Cloud data centers for storage and AI based processing. In this article, 5G integrated architecture for intelligent IoT based on the concepts of AI as a microservice (AIMS) is presented. The architecture has been conceived to support the design and development of AI microservices, which can be deployed on federated and integrated 5G networks slices to provide autonomous units of intelligence at the Edge of Things, as opposed to the current monolithic IoT-Cloud services. The proposed 5G based AI system is envisioned as a platform for effective deployment of scalable, robust, and intelligent cross-border IoT applications to provide improved quality of experience in scenarios where real-time processing, ultra-low latency and intelligence are key requirements. Finally, we highlight some challenges to give future research directions.*

**Keywords** – AIMS, Microservice, AI, 5G

## 1. INTRODUCTION

Over the last decade, the number of Internet-enabled devices has run into billions, in fact exceeding the number of people on the planet. The emergence of these devices and the large volume of data they generate have a significant impact on our day-to-day existence in diverse ways. Thus, with this development and the success of Cloud computing, a whole new networking paradigm has emerged known as the fifth generation wireless networks (5G), leading to the development of completely new platform for large scale distributed applications and mobile services in variety of application domains such as the Smart City, Smart Grid, Healthcare Monitoring, intelligent transport, Smart manufacturing, etc., exploiting the humongous data generated by the “smart” objects [1].

A microservice is a software development technique that structures an application as a collection of loosely coupled services. Decomposing an application into different smaller services can improve modularity and make the application easier to develop and deploy [2].

In this article, we propose a new concept of AI as a microservice (AIMS) as an IoT data-driven intelligence-provisioning infrastructure with the 5G capability to provide intelligent connectivity as services closer to the Things by leveraging resources of emerging computing technologies like Real-time Onsite Operations Facilitation (ROOF) [3], Fog [4] and Edge computing. The AI services will be provided as microservices, implementing lightweight functions that have been factored from the AI algorithms or processes. We envision the proposed system to allow AI functionality to be infused into 5G networks as distributed, composable microservices consisting of independent virtual components that can be deployed on the federated ROOF-Fog-Cloud continuum to improve scalability, interoperability and cutting down latency for real-time 5G applications. Thus, the proposed 5G integrated platform allows AI services to be provided seamlessly not only at the centralized data centers but essentially, also at the edge, closer to the IoT devices. The system and its concepts advocate an architectural principle based on the abstraction that provides end-to-end fabrics for composing, provisioning, deploying, managing and monitoring AI services regardless of whether such services are composed from the ROOF, Fog or cloud microservices. The proposed 5G architecture allows hierarchical and horizontal federation of smaller data centers from the Edge to the Cloud data centers continuum so that AI features and functions can be incrementally composed from microservices in such hierarchy as dictated by the required and available networking and computing resources at the ROOF and Fog levels to independently execute such AI functions. In fact, this architectural design would allow cross-border microservices to be composed from different local ROOF and Fog microservices over 5G network slices.

The remainder of this article is organized as follows. Background information on 5G networks and distributed Cloud is provided in Section 2. The proposed AIMS infrastructure is described in Section 3. Section 4 shows AIMS use cases. In Section 5, we highlight some challenges to give future research directions. Finally, we summarize our work in Section 6.

## 2. BACKGROUND

One major enabler of AIMS is the integration of 5G and Cloud computing, enabling the 5G applications to leverage the abundant compute and storage power of the geo-distributed Cloud data centers.

One of the widely researched challenges relates to how heterogeneous services and their operating platforms can interoperate on the same network. However, various research enthusiasts are aggressively addressing these “5G Vertical” challenges to enable the development of network slicing, multi tenancy, network programmability [5]. One of the main weaknesses of the solutions in this regard is that they still rely on transporting humongous IoT data across the 5G networks to various cloud data centers for storage or processing. We have identified the negative impacts and challenges of this model as follows:

- Latency sensitive nature of the Edge based application services necessitates that real-time decisions based on the acquired data from the Edge devices requires mechanisms for real-time processing of data for real-time intelligence [6]. How do we design, model and expose these intelligent services for decision making at the Edge of Things to address latency related problems of AI services across the 5G networks?

- Intelligent decision making at the Edge of Things introduces new AI dimension to IoT services such as real-time local processing of IoT data for quick intelligent decision making without necessarily transporting the heavy data through the expensive 5G networks. The challenge here is how do we develop data-centric IoT Services in which AI is a first-class design element?

Indeed, to take advantage of interoperable IoT platforms over 5G networks, IoT applications should be driven by AI deployed as autonomous microservices, essentially implementing the DIKW (Data, Information, Knowledge and Wisdom) at the edge of IoT [7]. Additionally, as interoperable IoT based platforms are being deployed through various use cases such as Smart City applications, Smart Manufacturing, etc., transporting huge volume of data from the IoT edge to the geo-distributed centralized Cloud data centers for processing is not only efficient in terms of communication bandwidth and energy consumption but also cannot support ultra-low latency applications [8].

With these ultra-low-delay sensitive applications, the current solutions are obviously not practicable. For example, a security surveillance application requires real-time processing of huge live video data, which is transmitted to the Cloud data centers for processing before intelligent decisions can be made [9]. This approach will not only be impossible to meet the latency requirements as such application may have to identify object in real-time but also such delay could lead to disastrous consequences. Thus, it is necessary to devise alternative solutions to the current store-and-process later systems such that processing and intelligent decision-making based on such data can be done close to the data sources in real-time.

Additionally, various solutions and concepts have recently been proposed to address this problem, from federated clouds, to Edge computing [8]. The federated clouds are a collection of heterogeneous infrastructures that may span the entire globe and requires that data from the IoT devices be transmitted to the cloud data centers. Thus, this architecture still depends on the Internet and public telecommunication infrastructure with very high latency and bandwidths requirements. Fog, on the other hand, aims to provide a system level horizontal architecture that distributes computing, storage, control and networking functions closer to the users in the Thing-Cloud continuum [10]. ROOF computing is closely related to Fog computing in that it provides highly distributed pervasive and virtualized platform data/processing to a central cloud data center. However, ROOF computing has been proposed to provide highly functional, secure and scalable IoT. It promises interoperable connectivity for variety of Things under the ROOF, context information and decisions for taking actions in real-time, information management and efficient connectivity to the Cloud and Service as well as efficient network design [4], [11]. This reduces communication delays and the size of data that needs to be migrated across the 5G to the cloud data centers.

## 3. AI AS MICROSERVICES (AIMS) AT THE EDGE OF THINGS

To deploy data-driven intelligent capability at the 5G networks, AI in various forms of machine learning algorithms, such as the deep learning, must be infused into the Edge-Cloud platform components. Thus, the 5G capabilities should be equipped with tools that allow intelligent services to be composed as data-driven microservices [12], [13]. The rationale is to address the weaknesses of the current monolithic Cloud based AI services, which cannot meet the requirements of real-time and ultra-low delay sensitive 5G applications. Rather than shipping the data to the cloud data centers where AI algorithms are applied to incorporate intelligent decision-making capabilities into 5G applications, these AI algorithms can be implemented and deployed closer to the sources of the IoT data and users by factoring the AI functionality into smaller functions that can be implemented as distributed microservices [14]. We proposes a hierarchically integrated infrastructure spanning the ROOF, Fog and Cloud computing platforms (Figure 1), to exploit resources at the Edge of Things (ROOF and Fog Computing resources) and Cloud data centers, as well as microservice concepts to incorporate AI capabilities into IoT applications. The microservice concept allows the decomposition or factoring of the current monolithic AI services (which are deployed only on the centralized Cloud data centers) into smaller functions deployed as AI microservices. The microservices can then be deployed closer to the data sources and users allowing seamless composition of AI services across the ROOF-Fog-Cloud layer. The AI features and functions are composed from the distributed microservices AIMS integrated platform, allowing AI functionality as intelligence services to be implemented and deployed close to the data sources and users despite the limited resources available at these layers and the huge computing resources required by AI algorithms.

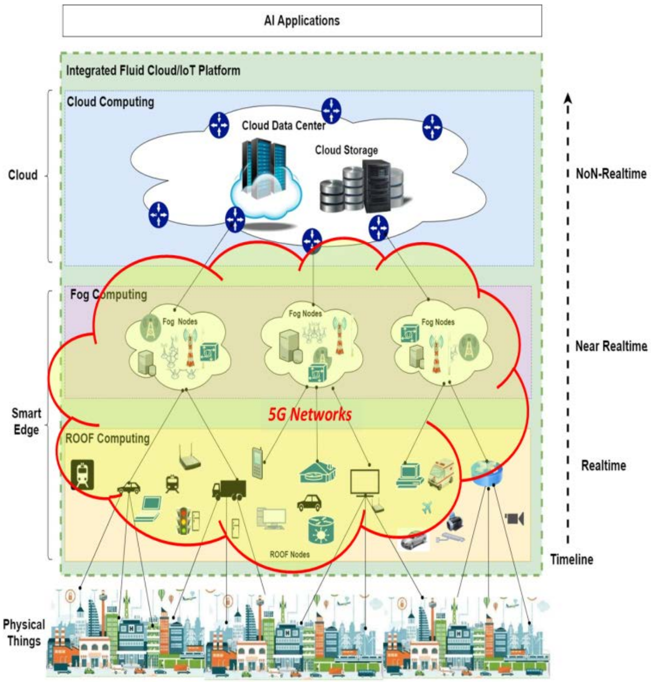


**Figure 1** – Edge-Cloud high-level integrated architecture over 5G networks for various AI enabled IoT applications

In software engineering, the concept of service-oriented architecture is not new. This is based on the idea that an application can be designed such that the functionality it provides is divided into smaller functions and implemented as services that can interact via well-defined programming interfaces and thus allowing scalability, robustness and interoperability. Network slicing is recognized as a game changer in the remarkable paradigm shift from 4G to 5G era because it can maximize the sharing of network resources and flexibility for dedicated logical networks [15]. AIMS based application involves composing interoperable microservices from the ecosystem of microservices distributed across the ROOF-Fog-Cloud systems as illustrated in Figure 2, showing how microservice at each 5G network slice can be composed. At ROOF level, for example, an AIMS service could execute a decision process based on the data obtained from the IoT devices after some other microservices at this layer have executed data gathering, and pre-processing tasks on the collected data. In fact, the pre-processed data can be temporarily stored on some of the nodes at this layer. One advantage of service composition at this layer is that some more important decision process with low latency can be executed without offloading such process and data to the upper layer such as the Fog and Cloud layers. Note that these microservices are developed and deployed independently of each other and are composable at runtime. The composition of the microservices can be realized sequentially based on Network Function Virtualization (NFV) [2] management and inter-slice resource brokering process. The dynamic adopting microservice architecture for the deployment of AI services means that we can now engineer data-driven IoT based applications that are composed of multiple hierarchical self-contained, lightweight, portable runtime and modular components deployed across a federation of network slices. This means that AI algorithms can be factored into modular functional entities that can be implemented as data-driven reusable algorithmic primitives. For example, the core functionality of a particular AIMS service could be a service providing regression analysis, classification, clustering, IoT data pre-processing functions such as feature extraction, feature reduction, dealing with missing data values etc. Each microservice is responsible for the execution of a smaller portion of an AI task with its own data, processing and notification points accessible to other microservices.

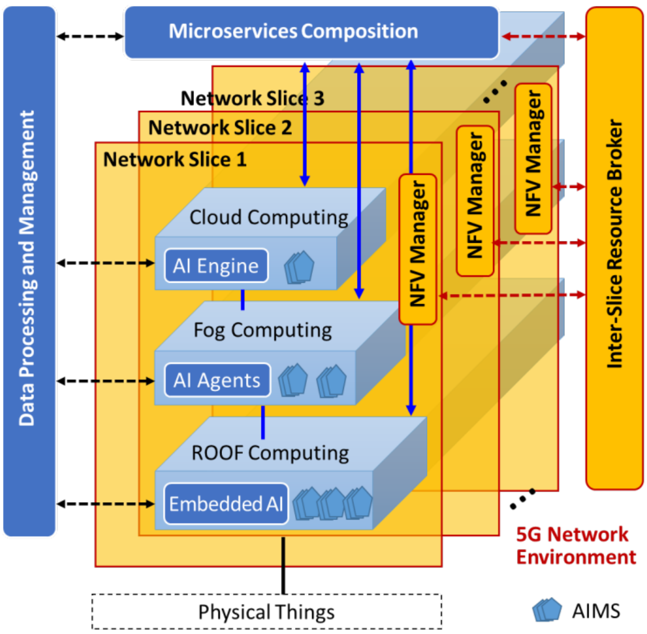


**Figure 2** – The ecosystem of microservices distributed across the ROOF-Fog-Cloud systems over 5G networks

In Figure 2, the first layer is the physical layer consisting of the IoT devices and connectivity protocols. The IoT devices can be categorized into 3 types. The first type consists of the edge sensors and actuators. These devices are capable of capturing data with little or no processing capability of operating system. They are equipped with low power 5G radio connectivity with which they can communicate with the edge devices. The second type of IoT devices at the edge are the edge devices. These are devices with the capability to run operating systems such as Android, IOS, Windows, Linux, etc. They possess the capability to not only aggregate data but also to execute pre-processing on the collected data. Some of these devices can also run some embedded AI algorithms as microservices to provide simple intelligent decision and insights on the data they have aggregated.

The ROOF layer consists of devices and nodes such as 5G gNodeBs, home routers, smartphones that provide the resources for always-available services, security, privacy in real-time as the next hop for the Things. It can be implemented on these devices that serve as Things' proxies for connectivity to the network and Cloud. In our proposed architecture, ROOF serves as a proxy for the physical Things for connectivity to the Fog and to the Cloud Computing data centers. At the layer of the architecture, AI agents and other related distributed applications can be deployed as microservices.

The third layer is the Fog layer, which is a virtualized layer providing compute, storage and networking services between the ROOF and the traditional Cloud data centers. This layer can deliver more powerful 5G application services that can be supported by the ROOF layer. This layer consists of Fog nodes, which are facilities and infrastructure that can provide resources for distributed 5G application services. In our architecture, base stations and other core network gateways serve as Fog nodes.

The fourth layer is the Cloud layer, which is located in the core network and support interoperability and wide-usage as AIMS modules independent to the data. In addition, it provides long-term decision making in the smart city services.

## 4. AIMS: USE CASES

### 4.1 Smart City Surveillance Application

One of the key areas where the architecture proposed in this article is most useful is in security surveillance in a Smart City platform. In a Smart City, there are numerous smart cameras installed in various parts of the city for different purposes ranging from traffic monitoring, security surveillance at train stations, bus stations, airports, shopping malls, streets, etc. Imagine that there is intelligence about an intending terror attacks and the pictures of possible suspects have been shared among various security monitoring systems in the city. The security monitoring system is linked with the smart cameras. To report the sighting of a suspect, the smart cameras should be empowered to carry our real-time analysis of live streams of video data and decide if an individual with a suspicious bag is one of the wanted terror suspects. To realize that, different analytical AI algorithms, such as anomaly detection using deep learning, can be deployed as microservices to support the surveillance cameras installed in the 5G based virtualized service infrastructure.

AI algorithm analyzes the video data for autonomous local decision-making. The smart camera can then communicate its decision to the appropriate authority for action while the cameras keep on monitoring the suspect and if need be passing control information to nearby cameras should the suspect move away from the current camera. Thus, the system can locally process the streams of live video data among themselves and thus to reducing traffic overhead, latency in 5G networks.

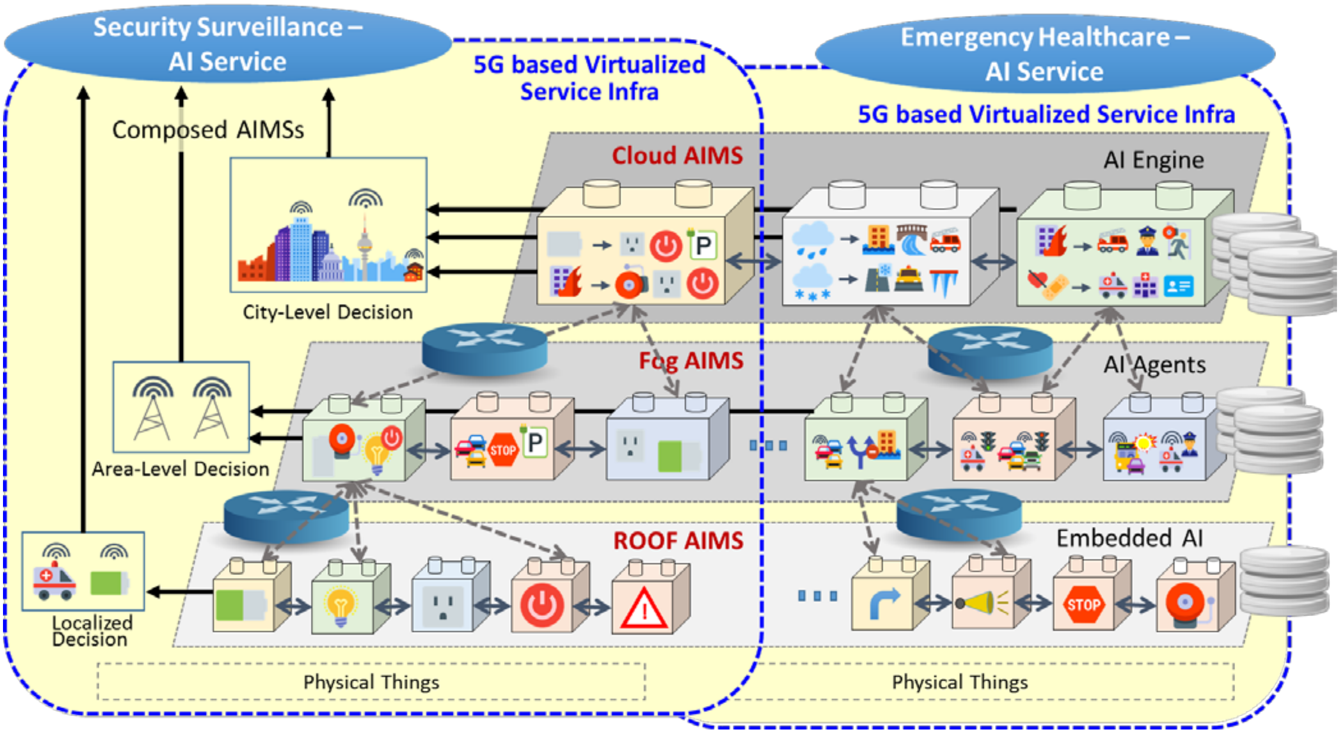


**Figure 3 –** 5G based Virtualized Service Infrastructure

### 4.2 Emergency Healthcare System

AI Microservices in the Fog-Clouds could be used to provide smart healthcare services where real-time vital sign and other data must be processed, and an instant decision has to be communicated to healthcare providers such as emergency health services. A microservice, might be responsible for collecting the data, another for processing the data and other for predicting, and another for deciding on action to take based on the prediction. The AI algorithms that these microservices implement are deployed on cross-border Fog-Cloud systems. This kind of system would provide low-latency, privacy, trust and secured mobility and location-aware supports to the individuals in the 5G environment.

To facilitate AI-powered 5G applications such as the security surveillance and emergency healthcare as shown in Figure 3, AIMS infrastructure hierarchically incorporates Cloud and Edge computing with AI and 5G technologies. AIMS provides multi-level AI components located from the Smart Edges (ROOF/Fog) of things to the Cloud centers. Thus, the AIMS enables various levels of intelligence, which are deployed at ROOF/Fog/Cloud layers, to be developed as independently deployable microservices (AIMS components). These AIMS components can then be incorporated based on message driven communications provided by the platform, allowing easier extensibility, interoperability, evolution, integration and composition of high-level, complex AI-powered 5G services.

## 5. CHALLENGES

The 5G integrated AIMS platform is envisioned to address important challenges of an advanced and efficient federated cloud platform with IoT for AI applications. It will be designed to offer distributed AI services (as a microservice) over 5G networks, leveraging multi-cloud computing, IoT and Big Data technologies.

### 5.1 Defining an integrated platform architecture for Cloud, AI and 5G

One of the key challenges is how to define an integrated reference architecture for multi-cloud IoT based microservices, enabling intelligent data acquisition and analysis through integrated protocols and standards with uniform access while supporting different interactions between various IoT services deployed on federated cloud systems at the 5G networks. Presently, there are frameworks providing solutions in this direction. A good example is the EdgeXFoundry open source platform developed for the edge of the network [16]. It interacts with the physical everyday working world of devices, sensors, actuators and other IoT objects. It has been designed as a framework for industrial IoT edge computing, enabling rapidly growing community of IoT solution providers to work together in an interoperable ecosystem of components to reduce uncertainty, accelerate time to market and facilitate scale. This platform brings the much-needed interoperability that makes it easier to monitor the physical world, send instructions to them and collect data, move the data across the Fog up to the Cloud where it can be stored, aggregated, analyzed and turned into information that can be acted upon. One important aspect of this platform that relates to AIMS is its capability that allows data to travel northwards and laterally to other edge gateways, or back to devices, sensors and actuators. However, the edge gateways only function as data collectors or aggregators for the IoT devices from which such data is transmitted to the cloud data centers. The EdgeXFoundry platform does not provide integration with other cross-border Cloud-IoT platforms and also does not incorporate intelligence at the edge of the network to allow application of AI algorithms for data processing and analyzing IoT data for intelligent decision making. Another is the MUSA project sponsored by the European Union [17]. MUSA is a distributed multi-cloud application platform over heterogeneous cloud resources. Its components are deployed in different cloud service providers and work in an integrated way and transparently for the end users. BigClouT [18] is another similar ongoing project sponsored by the European Union that leverages the power of Cloud computing, IoT and Big data analytics to provide distributed intelligence in a smart city network. The AIMS aims to define and develop an integrated platform architecture for the incorporation of multi-clouds systems and IoT for AI based services.

### 5.2 Specifying essential components and interfaces to support data-driven AI services

The AIMS infrastructure consists of broad variety of heterogeneous nodes, devices, protocols, etc. That interacts in diverse operating conditions from ROOF to the Cloud. This heterogeneity raises important question of how microservices deployed across this ecosystem of the federated AIMS platform would be able to communicate to exchange information and data that are in different formats. The popular solution would be to design a unified middleware framework, providing the abstractions of various layers on top of AIMS to hide this complexity from the microservices and allow them to fluidly exchange not only heterogeneous data and information but also intelligence seamlessly. Thus, various components and interfaces for communication across a federation of ROOF, Fog and Cloud platform would be specified. This middleware and its associated interfaces should be designed to guarantee interoperability between the federated ROOF, Fog and Cloud elements, coordinating the life cycle of the whole tasks of various microservices taking part in delivering intelligence as a service. Components for communication, configuration, microservice and resources discoveries, composition via orchestration or choreography and other related service interfaces would be specified and designed.

### 5.3 Supporting the harmonious management of computing resources

In the proposed AIMS integrated platform, large heterogeneous and distributed IoT devices will produce huge volume of data at rapid velocity. Gleaning meaningful information and insights from this data using distributed AI services will require a shift from the traditional architectural style to a more agile approach that allows more robust scalability, evolvability and maintainability of large-scale distributed multi-cloud IoT systems. Microservice architecture, as one of the recent trends in the design and development of agile distributed systems, defines a new approach to designing and developing a single application as a suite of smaller services, each running in its own process and communicating with lightweight mechanism to execute just one task. Such services are small, highly decoupled, independently deployable, focusing on doing a small specific and interdependent task that can provide some level of intelligence and yet when combined with other tasks provide higher or deeper intelligence depending on available and required resources. In order to achieve a much bigger task, these services can be combined to realize such functionality. One of the key challenges therefore that need to be addressed is that of dynamic allocation and orchestration of resources for the distributed microservices in the AIMS federated platform depending on what compute resources are currently available and how much of resources are required by the current microservices for the task execution. Although, there has been existing resource allocation in 5G networks, however, there is no yet concrete solution for resource allocation for integrated ROOF, Fog and Cloud platform. Even for the more mature Fog, resource and service orchestration remains a challenging research problem. For the AIMS platform, there would be several microservices sharing resources and this might result in resource contention and interference. Thus, new mechanisms and strategies for dynamic and fluid resource allocation and scheduling would be investigated to reduce response time for task execution across the 5G integrated AIMS platform.

### 5.4 Applying new mechanisms using intelligence in data lifecycle

To provide distributed intelligence at the edge of things, a critical factor for deploying AI services on the ROOF-Fog-Cloud integrated infrastructure is more related to application partitioning or factoring, real-time service composition, data mobility and aggregation. To address these issues, there is need for new mechanisms for factoring or decomposing AI services into functions that can be delivered as re-usable microservices for executing specific smaller tasks. These new mechanisms for service composition must be developed to achieve a fluid decision making process exploiting raw data from the physical devices, extracting meaning and insights in order to achieve the DIKW at the ROOF, Fog and Cloud layers of the infrastructure's hierarchy. Such mechanisms should support the dynamic discovery, composition and relocation of AIMS according to the required and available resources across the integrated nodes on the AIMS platform. How to determine based on the available resources, what level of intelligence should be provided by a microservice, what tasks or functions should be executed and at what layer of the infrastructure are important technical challenges.

### 5.5 Supporting trusted AI services

The AIMS based applications will process large volume of data using distributed microservices from ROOF to Cloud continuum of the platform. Thus, as distributed and interoperating microservices execute intelligence based on data from the physical devices, such data can be compromised as they may be exposed to malicious third parties. In fact, malicious microservices can be injected into the system to wreak havoc or to provide false or misleading decisions. This is a crosscutting challenge since it does not only affect a layer but all layers and aspects of the AIMS ecosystem, from radio communications to the microservices across the 5G networks. Additionally, the interfaces between Cloud, Fog and ROOF computing are potential sources of vulnerability and consequently may lead to corruption of IoT data and services. To ensure trust, privacy and security, capabilities for end-to-end encryptions, intrusion detection and prevention of unauthorized microservices or services will be required. Trust management should be investigated as a useful technology for providing such required security services. How can trust management be used to provide security, dependability and reliability for AIMS and associated data at various layers of the ROOF, Fog and Cloud integrated platform? For users' needs and rights to be enforced as autonomous microservices exploit IoT data to infuse intelligence into IoT applications, there is need to investigate integrated and federated ROOF, Fog and Cloud platform to propose the best and unique trust and security mechanisms for enforcing integrity, dependability and reliability of the platform and its services.

## 6. CONCLUSION

This article proposes an IoT data-driven intelligence-provisioning infrastructure with the 5G capabilities to provide intelligent connectivity as services closer to the Things by leveraging the compute resources of a hierarchically integrated computing environment (ROOF-Fog-Cloud). The proposed AIMS aims to provide a lightweight platform for effective deployment of scalable, robust, and intelligent cross-border 5G applications. We have envisioned the proposed architectural approaches in terms of system perspectives to allow AI functionality to be infused into 5G networks as distributed, composable microservices consisting of independent virtual components that can be deployed on the federated Roof-Fog-Cloud continuum to improve scalability, interoperability and cutting down latency for real-time 5G applications. In this article, we have also highlighted some challenges to give future research directions.

**ACKNOWLEDGEMENT**

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT (2018R1A2B2003774).

**REFERENCES**

[1] A. Krylovskiy, M. Jahn and E. Patti, "Designing a Smart City Internet of Things Platform with Microservice Architecture," 2015 3rd International Conference on Future Internet of Things and Cloud, pp. 25-30, Aug. 2015.

[2] "Microservices," https://en.wikipedia.org/wiki/ Microservices.

[3] P1931.1, "Standard for an Architectural Framework for Real-time Onsite Operations Facilitation (ROOF) for the Internet of Things," https://standards.ieee. org/project/1931_1.html.

[4] N. Mohan and J. Kangasharju, "Edge-Fog cloud: A distributed cloud for Internet of Things computations," 2016 Cloudification of the Internet of Things (CIoT), pp.1-6. 2016.

[5] Marco Gramaglia, Albert Banchs, "Definition and specification of connectivity and QoE/QoS management mechanisms," 5G NORMA Deliverable D5.2, Jul. 2017.

[6] P. Patel, M. Intizar Ali and A. Sheth, "On Using the Intelligent Edge for IoT Analytics," in IEEE Intelligent Systems, vol. 32, no. 5, pp. 64-69, Sep. 2017.

[7] M. Fazio, A. Celesti, R. Ranjan, C. Liu, L. Chen and M. Villari, "Open Issues in Scheduling Microservices in the Cloud," in IEEE Cloud Computing, vol. 3, no. 5, pp. 81-88, Sep. 2016.

[8] H. Liu, F. Eldarrat, H. Alqahtani, A. Reznik, X. de Foy and Y. Zhang, "Mobile Edge Cloud System: Architectures, Challenges, and Approaches," IEEE Systems Journal, vol. 12, no. 99, pp. 1-14, Feb. 2017.

[9] A. Celesti, M. Fazio, M. Giacobbe, A. Puliafito and M. Villari, "Characterizing Cloud Federation in IoT," 2016 30th International Conference on Advanced Information Networking and Applications Workshops (WAINA), pp. 93-98, Mar. 2016.

[10] C. Long, Y. Cao, T. Jiang and Q. Zhang, "Edge Computing Framework for Cooperative Video Processing in Multimedia IoT Systems," in IEEE Transactions on Multimedia, vol. 20, no. 99, pp. 1126-1139, May 2018.

[11] Y. Ai, M. Peng, and K. Zhang, "Edge cloud computing technologies for internet of things: A primer," Digital Communications Networks, vol. 4, iss. 2, pp. 77-86, Apr. 2017.

[12] A. Taherkordi and F. Eliassen, "Poster Abstract: Data-centric IoT Services Provisioning in Fog-Cloud Computing Systems," 2017 IEEE/ACM Second International Conference on Internet-of-Things Design and Implementation (IoTDI), pp. 317-318. 2017.

[13] James Kobiolus, "Building AI Microservices for Cloud-Native Deployment," https://wikibon.com /building-ai-microservices-for-cloud-native-deployments/.

[14] M. Villari, M. Fazio, S. Dustdar, O. Rana, and R. Ranjan. "Osmotic Computing: A New Paradigm for Edge/Cloud Integration," IEEE Cloud Computing, pp. 76–83, 2016.

[15] Qi Wang, et. al., "SliceNet: End-to-End Cognitive Network Slicing and Slice Management Framework in Virtualised Multi-Domain, Multi-Tenant 5G Networks," 2018 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB), pp. 1-5, 2018.

[16] "Edge X Foundry," https://wiki.edgexfoundry.org/ display/FA/EdgeX+Foundry+Project+Wiki.

[17] A. Sill, "The Design and Architecture of Microservices," IEEE Cloud Computing, vol. 3, no. 5, pp. 76-80, Sept. 2016.

[18] "BigClouT: Big data meeting Cloud and IoT for empowering the citizen clout in smart cities," http://bigclout.eu/.