

A Paradigm-Shift for the IoT-Edge-Cloud Continuum

Panagiotis Kokkinos, Aristotelis Kretsis, Emmanouel Varvarigos

SERRANO EU Project (ict-serrano.eu)

Institute of Communication and Computer Systems (ICCS)

National Technical University of Athens (NTUA), Greece

Internet of Things (IoT) devices including but not limited to robots, cameras, smart meters, operating both indoors and outdoors, are increasing in number and serve as important data sources for analytics and machine learning (ML) tasks. In particular, the number of IoT devices in a wide array of domains operating at the network edge is forecast to grow to almost 6.5 billion by 2030, an increase of over 4 billion compared to 2020¹. At the same time, these devices are subject to energy, processing and connectivity constraints and are vulnerable to security threats². Centralized cloud computing infrastructures and respective services currently handle most of the processing and storage requirements of from respective applications, making decisions and taking actions based on the data received from IoT devices. Nevertheless, it has been recognized that only a fragment of the data generated will be truly useful, while their size will exceed the storage capabilities of today's cloud data centres³. At the same time the applications' performance requirements become increasingly strict, e.g., in terms of latency, geographic density, energy consumption. From the above, it is clear that bringing computation to the edge is a necessity and a key to the performance of the new IoT-based applications^{4,5}. Edge resources serve as aggregators for cloud services and are responsible for rapid and secure storage and processing of IoT data, extending cloud resources to the edge. The edge ranges from the on-device and on-premise edge, to "near edge", to "far edge", and to regional data centres, which together with the traditional cloud resources form an **IoT-edge-cloud continuum**. This offers a corresponding cost and performance continuum that **promises, if properly used, better user experience and lower monetary and energy costs** for the ecosystem of application, service, and infrastructure providers. Fully exploiting the IoT, edge and cloud resources requires building **hyper-distributed systems** that **intelligently** adapt to dynamic application requirements and resource availability across these heterogeneous layers. Furthermore, this eminently distributed architecture brings new challenges beyond resource management, since each of the layers may represent one or more distrusting administrative domains, usually spanning multiple physical locations.

Today, there are a **several efforts to build this IoT-Edge-Cloud continuum**, initiated from the industry, the academia, the EU and other private and public organizations in international level. These efforts develop the required software and algorithms for resource orchestration, application and data management, development and deployment, using **open source software** and standards. In addition, there is a plethora of **hardware-accelerated devices** from FPGA, GPUs and DPUs that can be deployed in the cloud, in the edge and on-device to accelerate computing intensive workloads. **High Performance Computing (HPC)** technologies and infrastructures are also part of the computing and data equation. The above developments and technologies are also **validated in different domains** and discussions are ongoing regarding their applicability and benefits. Also, in European level the goal is to **strengthen the European strategic autonomy** in the evolving data and AI-economies, foster competition in the Cloud/Edge services market for the European cloud/edge and software industry, while facilitating European access to foreign markets.

In parallel and beyond the above there are still key challenges to be addressed. Some of these are technology related but we believe that the most important ones have to do with the **need of a paradigm shift** in the way we

¹ <https://www.statista.com/statistics/1259878/edge-enabled-iot-device-market-worldwide/>

² <https://www.iotsecurityfoundation.org/survey-less-than-10-of-iot-devices-keep-data-secure/>

³ <https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html>

⁴ <https://aioti.eu/wp-content/uploads/2020/10/IoT-and-Edge-Computing-Published.pdf>

⁵ <https://atos.net/wp-content/uploads/2021/08/atos-2021-perspective-on-edge-computing-white-paper.pdf>

utilize these technologies and in particular the ones that relate to edge computing, so as to increase the **deployment, densification and interoperability**, while still satisfying the requirements of **security, trust and energy efficiency**.

The realization of any edge infrastructure depends highly on its deployability: deployment needs to be massive for the infrastructure to provide the anticipated edge services. The implementation of Europe's Digital Decade⁶ policy programme for 2030, aims at deploying 10,000 climate-neutral secure edge nodes, enabling secure and sustainable digital infrastructures, showing that the European market has strong potential for growth in the use of edge technology. In order, to achieve the required edge densification and capacity expansion to support competitive services, it is **necessary to enable more affordable deployments of novel edge platforms**. This can be assisted through computing and storage edge **disaggregated white boxes, open specifications and reference implementations** based on off-the-shelf hardware (e.g., hardware accelerators) and modular **open-source software and open APIs**, leading to large cost savings and deployment flexibility. The maximum use of open-source software will reinforce competitiveness, portability and interoperability. In addition, commoditization of edge hardware, starting with disaggregation, has the potential to create an open and competitive market for interchangeable parts that will help cloud-edge achieve economies of scale.

Edge computing also **business perspective must be carefully thought** and fulfilled, since IoT devices and edge resource **deployability, in turn, depends on the economic benefits expected** and the desired Return on Investment (RoI). In practice today, several EU companies and self-employed individuals buy commodity or high-end equipment for local storage and processing, in manufacturing, warehousing, agriculture, medical services and other domains. However, beyond the self-targeted and immediate benefits, there are, no real incentives, mainly in terms of uptake of investment, for deploying and supporting new edge infrastructures. In this way, **today the existing edge resources are limited and isolated**. We believe that **edge resources and specialized IoT devices can be viewed as new marketable entities, through sharing and federations**. This will enable small infrastructure owners to reach the critical mass required to participate in the respective market according to an open market model and have better negotiating power in "selling"/trading their free resources and thus favouring in practice edge and IoT deployments. In this way, owners of infrastructure will become prosumers (i.e., producers and consumers simultaneously) of processing and storage capacity.

Another, important challenge is the applicability of **edge processing and storage in various domains**. No, particular, "killer application" for edge has been identified. In practice, however this may not be necessary, instead it is reasonable that **future edge infrastructures should be able to accommodate any kind of processing and storage tasks/workload that today are served by cloud resources**. This is necessary so that edge infrastructures are ready for the very realist scenario when the cloud resources are not available at all or not efficient (in terms of cost, performance, security etc) to use. This will also strengthen the EU independence from external (cloud) providers.

In addition, **interoperability between data producers and data consumers should be enhanced enabling** the development of innovative application in various domains by **third-party developers** utilizing: i) shared data from IoT devices (e.g., drones, cars, city sensors, robots), shared resources (processing, storage, specialized hardware) and algorithms (e.g., for AI/ML analysis). These can be combined in application workflows that are setup in an infrastructure-agnostic manner and then transparently deployed and autonomously served in the IoT-Edge-Cloud continuum, using AI-based orchestration mechanisms.

Addressing these challenges can revolutionize the European IoT/edge/cloud sectors, offering a credible pathway to strengthen European leadership, autonomy and innovation in the global data economy.

⁶ <https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/europes-digital-decade-digital-targets-2030>