

Edge Computing: The Computing Infrastructure for the Smart Megacities of the Future


Naser Hossein Motlagh , University of Helsinki

Lauri Lovén , University of Oulu and TU Wien

Jacky Cao , University of Oulu and Norwegian University of Science and Technology

Xiaoli Liu  and **Petteri Nurmi** , University of Helsinki

Schahram Dustdar , TU Wien

Sasu Tarkoma , University of Helsinki and University of Oulu

Xiang Su , University of Oulu and Norwegian University of Science and Technology

This article contributes a research vision for using edge computing to deliver the computing infrastructure for emerging smart megacities, with use cases, key requirements, and reflections on the state of the art. We also address edge server placements, a key challenge for edge computing adoption.

Urbanization is continuing at unprecedented rates, with estimates suggesting that, by 2030, the global population of major cities with 10 million or more inhabitants will grow from

3.2 billion to almost 5 billion.¹ Parallel to this growth, there is an increasing trend to make cities smart by integrating sensors, networks, and artificial intelligence (AI) to optimize city functions and offer services that support inhabitants' lives. For example, environmental sensors are increasingly being deployed within cities to monitor various pollutants at fine spatial and temporal resolutions,

Digital Object Identifier 10.1109/MC.2022.3203613
Date of current version: 15 November 2022

and AI algorithms are being developed to investigate the links between pollutants and adverse health conditions. This information can then be used to offer services that support the understanding and mitigation of adverse health effects and assess the effects of diverse environmental initiatives.²

INTRODUCTION

The continuing growth of cities poses challenges in ensuring their continual smartness while taking advantage of technological advances. As illustrated in Figure 1, future megacities are expected to integrate a wide range of sensors, such as wearables, diverse cameras (for example, surveillance, thermal, and hyperspectral cameras), and environmental sensors. Indeed, according to forecasts, the number of Internet of Things (IoT) and sensor devices is expected to reach 30 billion per city by 2030.³ While this data growth provides opportunities for optimizing the city and the inhabitants' lives, significant challenges for

computing infrastructure are introduced for processing and analyzing measurements.

Currently, the most common smart city solution is to externalize the computing needs to a cloud provider. While this offers scalability, the costs of using the cloud for data storage and processing will become prohibitive as the city and the number of sensors within it grow. This is particularly prevalent as we transition from smart cities to smart megacities with millions of inhabitants and potentially billions of sensors that require persistent storage and constant processing. Another limitation of the cloud-centered approach is the requirement of strong centralization without being able to exploit the spatial nature of the data (for example, processing data for individual neighborhoods), which can result in imbalanced processing, significant overhead, and, ultimately, high cost. As a result, there is a clear need for alternative processing and storage solutions that address the needs of megacities.⁴

This article contributes a vision for using edge computing as smart megacities' computing and storage infrastructure. Edge computing is well suited for smart megacities as it supports distributing the processing and allows the exploitation of spatial characteristics in the data. Edge computing also integrates support for minimizing latency, conserving bandwidth consumption, and providing resilience against security and privacy threats.⁵ Ongoing 5G—and future 6G—deployments are further expected to support edge computing by providing the infrastructure for scalable deployment of edge computing nodes and to offer additional capabilities that benefit megacities.⁶ We derive use cases, identify key requirements, and reflect on current state-of-the-art solutions to analyze the potential of using the edge to deliver a scalable and cost-effective infrastructure for the computing needs of emerging smart megacities. We also address infrastructure scalability by addressing the placement of edge nodes.

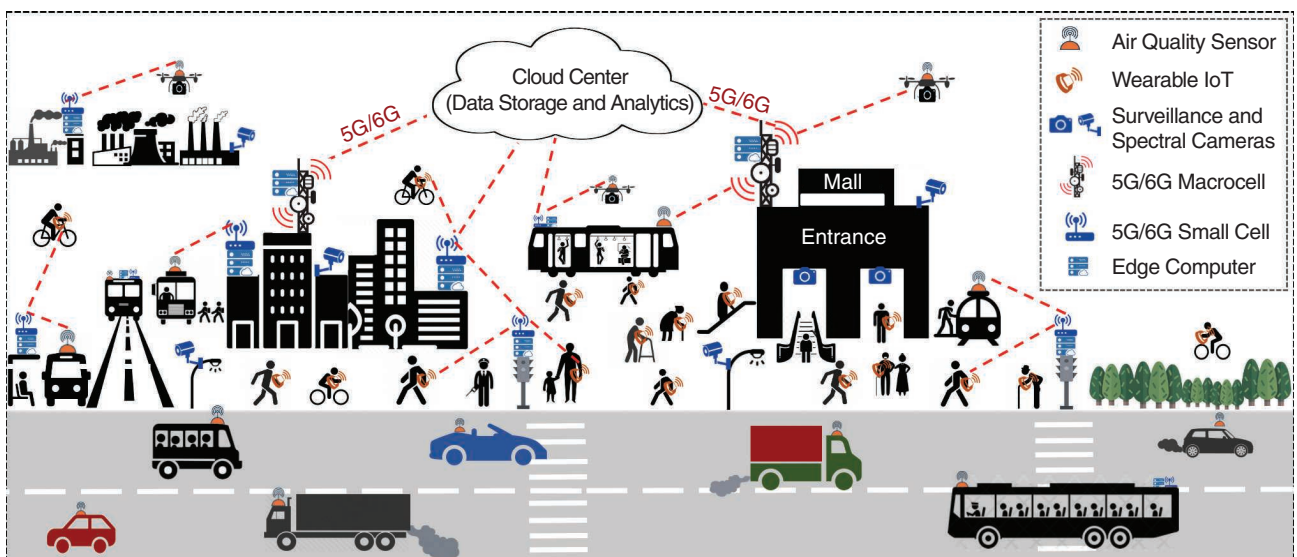


FIGURE 1. The vision of massive IoT sensor deployment in future smart megacities. IoT: Internet of Things.

We use air quality data drawn from a real-world deployment to demonstrate how optimizing deployment locations helps to enhance and optimize the computing resources available for the city while meeting practical constraints (such as budget and energy).

APPLICATION USE CASES

Future smart megacities are expected to contain large-scale pervasive deployments of sensor-enabled devices (see Figure 1). These deployments are expected to increase by tens of thousands or even billions of sensors, generating massive data streams and requiring flexible processing and storage support from the underlying computing infrastructure. These deployments offer significant opportunities for enhancing the functionality of cities and delivering services that support inhabitants. However, at the same time, megacities require sufficient (computing) infrastructure to support these applications. We next discuss representative examples of typical application use cases to highlight the range of opportunities and the diversity of the applications' processing needs for edge computing. We note that many of these use cases are already being explored in smart cities, but as we transition to smart megacities, the scale of sensor deployments increases by several orders of magnitude, demanding more flexibility and scalability from the underlying infrastructure.

Surveillance and hyperspectral cameras

Smart megacities are expected to integrate massive numbers of surveillance, hyperspectral, and other cameras, which can offer a wide range of services. For example, processing video streams at the network edge from densely deployed

surveillance cameras in cities supports counting people and vehicles to model people's mobility and estimate health risks.^{7,8} Thermal cameras can be installed at points of interest (for example, at the entrances of shopping malls) to scan for potential signs of illness and to mitigate health risks for others. Surveillance cameras can also offer security by detecting suspicious activities and criminals. Surveillance cameras' resolution and frame rate are continually improving, resulting in each camera producing larger volumes of data at a higher velocity than before.

At the same time, alternative imaging modalities, such as thermal and hyperspectral cameras, are also integrated, further increasing data processing needs. For example, hyperspectral cameras generate images of 30–300 MB per image.⁶ Processing the data centrally becomes highly difficult as the scale and nature of the deployments increase. As envisioned in this article, edge computing offers flexibility to scale the infrastructure together with the deployments and harness the spatial nature of the data when they are being processed. Naturally, these scenarios also raise security and privacy issues, but taking advantage of the edge instead of the cloud mitigates these issues by providing local storage and allowing privacy protection to be handled close to the location where the data are generated. Edge servers also reduce communication costs and the overall latency.

High-resolution air pollution monitoring

Air pollution is considered one of the grand challenges of our time, and it is particularly important in megacities where pollutant concentrations tend to be the highest. Low-cost air quality sensors are emerging as a solution

to increase the resolution of air quality monitoring, but they require massive-scale deployments that link with high-quality monitoring stations. Transmission and processing of mobile and geographically distributed data at the network edge enable new applications that can provide real-time and hyperlocal exposure data.² Edge servers are also crucial for improving data quality as low-cost sensors are prone to measurement drift and require periodic recalibration.^{9,10} Edge servers can support recalibration and ensure that accurate and real-time air quality information can be offered to the public.⁶

Unmanned aerial vehicles

Unmanned aerial vehicles (UAVs) are expected to be a significant source of data generation and an important tool for enabling smart megacities. Among others, UAVs can enhance network connectivity, deliver emergency services in disaster situations, support pollution or power line monitoring, and offer security and surveillance services. In smart megacities, the number of UAV flights will increase, which necessitates controlling UAVs in real time to avoid collisions between them and other obstacles. The amount of data that UAVs generate can also be very high (up to terabytes¹¹), and many applications would require near real-time processing (for example, collision detection) of these data masses.² Edge servers are expected to play a crucial role in providing the platform for coordinating and controlling UAV operations and delivering services that utilize the data the UAVs collect.

Health care

Health-care delivery is a central concern for the governance of future smart megacities. Future health-care services

are expected to take advantage of wearables and in-home medical (IoT) devices, increasing the complexity of health-care services and the volume of data being generated (thousands of terabytes per year). These data can offer personalized health services and optimize health-care delivery. Enabling personalized health care requires intelligent and privacy-preserving processing of the collected data. Edge servers provide a natural point for offering this functionality while meeting service-level quality requirements, including high reliability, low latency, and mobility support.¹²

Extended reality

Extended reality (XR) categorizes immersive technologies, including augmented reality (AR), virtual reality (VR), and mixed reality. AR enhances the physical world with virtual annotations where environments are typically analyzed with computer vision algorithms.¹³ In the context of smart megacities, XR technology is essential for analyzing the sensor streams in context, thus providing visualization support for different applications. The algorithms needed for processing are often not deployable on the AR hardware (for example, smartphones or head-mounted displays) as they lack the necessary computational resources. Therefore, the most resource-hungry processes should be offloaded. Efficient edge servers are preferred to the cloud as the edge is closer to end users, providing lower latency communication, privacy, and security benefits. One of the most pressing challenges for VR is the necessity of wired VR headsets to transmit high-resolution video at high frame rates as current wireless technologies (LTE and Wi-Fi) are not capable of doing so.¹⁴ Thus, improved communication technology combined

with efficient edge resources is needed to route and analyze such data.

REQUIREMENTS AND ENABLERS

Providing communication and computation capacities to support massive-scale IoT applications in smart megacities is challenging because of the limitations of the existing technologies and infrastructure. This section highlights the key requirements for massive IoT deployments in smart megacities. The requirements are summarized in Table 1.

Ubiquitous connectivity and massive IoT connections

Reliable network connections and extensive computational resources provide the foundation for massive-scale, diverse, and highly dense IoT deployments in megacities. To offer smart services at these locations and accommodate the expected massive number of 30 billion connected IoT devices by 2030,³ ubiquitous connections, optimizations of existing network infrastructure, the deployment of upcoming 5G and beyond (6G) networks, and IoT gateway solutions are required. Currently, IoT connectivity in cities is accomplished through wide area network (WAN) technologies, such as LoRa, Sigfox, LTE-4G, and narrow-band (NB)-IoT, but the current technologies contain several limitations. For example, they are incapable of supporting massive numbers of connections in dense and hyperlocal (for example, indoor) environments. Upcoming 5G (and beyond) networks are expected to support extremely massive connections and provide high-speed connectivity almost anywhere—for example, on the ground and sea, in the sky, and even in spaces not covered by existing technologies.

Extreme high-capacity communications

Many megacity use cases require a considerably larger network capacity than current smart city deployments offer to transmit large amounts of data at a fast rate. Examples include continuous image streaming from surveillance cameras with 16K video resolution in 360° with a refresh rate of 240 Hz, transmitting hyperspectral camera images (such as images of size 30–300 MB taken in less than a second for estimating air quality over open areas), and data transmission in multisensory applications (for example, autonomous driving requires data rates and communication capacity exceeding 1 Tbps). Supporting these data rates requires enhanced communication technologies, which in turn require denser deployments of communication infrastructure, providing the necessary infrastructure for integrating edge computing support. The necessary improvements are foreseen to be offered by 5G millimeter-wave (mmWave) communications, which offer much higher bandwidths. While the bandwidth provided by mmWave is promising, the overall loss resulting from path loss, rain attenuation, and atmospheric absorption is more significant for a point-to-point link than in other communication standards. In the near future, terahertz (THz) communications are expected to evolve with 6G communication systems to overcome these performance limitations of mmWave systems.

Ultrareliable and low-latency communications

Real-time and mission-critical IoT use cases, such as autonomous cars and UAVs, require extremely low-latency communications and effective data

processing. These use cases are characterized by fast processing, analytics, and intelligent decision making at the network edge and include scenarios such as avoiding accidents and hazards. The current network technologies, including 4G and NB-IoT, used in IoT communications are generally incapable of fulfilling the extremely low-latency requirements of mission-critical IoT use cases, even when combined with edge computing. Fortunately, 5G networks satisfy most of the requirements, including offering ultrahigh network reliability (>99.99%) and ultrareliable low-latency communications of 1-ms delay per packet transmission for IoT applications. 6G networks are further expected to

decrease latency (eventually to <100 μ s), offering to balance reliability, latency, and scalability and providing the extremely low latency that massive IoT connections require.¹⁵

Thus, emerging communication standards already cover most of the communication requirements, leaving the provisioning of computing infrastructure the main challenge for supporting megacity-scale deployments. Edge computing can offer this infrastructure while offering flexibility in scaling the deployments to reflect differences in population density across the city.

Localization

Offering accurate and reliable service in megacities also requires knowing the

precise position and elevation of where data are collected. The location of the IoT devices also helps to identify the best and nearest edge server to connect with. For example, a sensor mounted on a UAV or a sensor used on higher floors of a building would produce location information in 3D space. Current techniques, such as GPS and experimental low-power WAN (LPWAN) localization methods, provide satisfactory solutions for today's IoT deployments, but they do not meet the demands of future smart megacities when massive numbers of IoT devices are deployed. 5G technologies integrate mmWave communications, and future 6G networks propose THz standards (such as massive deployments of infrastructure, that is, edge

TABLE 1. The state of the art, key requirements, and possible solutions.

Concern	State of the art	Key requirements	Expected solutions
Ubiquitous connectivity and massive IoT connections	LPWAN standards, such as LTE-M, NB-IoT, and LoRaWAN, and other wireless technologies are used.	Enabling massive connections, optimizing existing network infrastructure, and achieving ubiquitous connectivity in dense and hyperlocal environments	Future communication networks, such as 5G and 6G, promise extreme massive connections, such as 10 million connections/km ² .
Extreme high capacity communications	mmWave communications present solutions for data-intensive applications.	mmWave has overall loss, such as path loss and atmospheric absorption, and extremely high data rates (>1,000 Gbps) are needed.	THz communications are expected to evolve with 6G communication systems.
Ultralow latency communications	Current mobile communication and other short-range wireless technologies are used in IoT communications.	Current technologies are incapable of meeting the extremely low-latency requirements of mission-critical IoT use cases.	5G will offer extremely low latencies of 1-ms delay per packet transmission, and 6G will provide latencies of less than 100 μ s.
Localization	GPS and experimental LPWAN localization techniques, such as LoRa, are examples of existing solutions.	Current technologies do not meet the needs of future smart megacities when massive numbers of IoT devices are deployed.	THz communication standards based on massive infrastructure deployments would decrease inaccuracy errors to a cm level.
Data management	Existing solutions use layered structure (storing and processing data on devices, at the edge, and on the centralized cloud).	The need for agile data management architectures, platforms, and data governance policies and practices	Upcoming 5G and 6G are expected to enhance processing and storage capacities by real-time analytics at the edge.

LPWAN: low-power wide area network; NB-IoT: narrow-band IoT; LoRaWAN: long-range WAN; mmWave: millimeter-wave.

nodes, extended use of beamforming, and utilizing dedicated algorithms) that enhance the localization accuracy by providing extremely high-precision positioning, which is expected to achieve centimeter-level accuracy.¹⁶

Data management and analytics

In future smart megacities, massive deployments of IoT devices intended for diverse applications will produce massive volumes of heterogeneous and geographically dispersed real-time data of different varieties. Managing these data will require agile data management architectures, platforms, and data governance policies and practices. Existing smart city IoT data management platforms utilize a layered structure through storing and processing of the data locally on the device, at the network edge, and on centralized clouds. Upcoming 5G and 6G networks will provide enhanced edge computing platforms that overcome the limitations of existing IoT management platforms. Indeed, in 5G, real-time analytics improve the latency of IoT applications by enhancing the processing and storage capacities at the edge, and 6G is expected to further improve the analytics by making AI an integral part of the communication infrastructure.

EXPERIMENTS

The previous sections addressed the technical capabilities of edge servers to support emerging IoT application scenarios tailored to smart megacities. In practice, taking advantage of edge computing also requires a sufficient proliferation of edge servers within the megacity. Next, we present experiments to demonstrate how optimizing the placement can ensure the necessary processing while minimizing resource costs.

Experimentation setup

City-District sensing and data collection. We draw on an air quality dataset collected from a measurement campaign in the Pakila district of Helsinki, Finland, from 30 October 2019 to 15 January 2020. The Pakila district is a crowded residential living area containing many detached houses where air quality in the area is affected mainly by wood burning, fireplace usage, and street dust. For measurements, research participants used 40 portable air quality sensor devices. After the provision of instructions on how to use the devices, they were delivered to citizens in the area. The citizens were expected to carry these sensors with them in their daily lives. The collected data present the measurements for an area of circa 6 km² that approximately maps to the Pakila district. The details of the measurement campaign and air pollution profile of the district are explained in Kortoçi et al.¹⁷ Naturally, we expect smart megacities to deploy significantly more sensors, thus resulting in larger volumes and faster velocity for the data streams. Nevertheless, our experiments are based on data from real users and reflect actual movements in a city district. Thus, our experiments indicate the geographical needs of the processing infrastructure.

Portable sensor devices. The portable sensor devices used for the measurements are based on a BMD-340 module and connect to a smartphone over Bluetooth Low Energy. The accuracy of the air quality sensors has been verified both in a laboratory and in the field.¹⁷ The sensor device is designed as a smart city solution for monitoring air quality at a low cost with an

approximate price of US\$250 per unit, which decreases to < US\$100 when produced in large quantities. The sensor devices are equipped with a Sensirion SPS30 sensor that measures masses of particulate matter, such as PM_{2.5} and PM₁₀; meteorology sensors for temperature, relative humidity, and air pressure; gas sensors, including carbon monoxide (CO₂), nitrogen dioxide, and ozone; and an ambient light sensor.¹⁸ After the measured variables are transmitted to the smartphone, time stamps and coordinate (positioning) information are combined with the sensor measurements, and the data are transferred to nearby cell towers.

Cell towers. We collect the coordinates of cell towers deployed in the Pakila district using CellMapper, a crowd-sourced website for cellular towers and coverage mapping. We consider the coordinates of the LTE-4G and 5G cell towers for the three main operators in Finland and collect 25 coordinates for the Pakila area. All three operators offer high-speed connectivity at low prices, even in remote regions, and they also provide 5G New Radio (NR) (on 3,500 MHz) services in cities. Based on OpenSignal, 5G NR services are available to 11% of the population in Finland, whereas LTE-4G service availability is around 93% by the three mobile operators.

Simulations. We conduct a series of simulations to optimize edge servers' number and placement. As the optimization criteria, we consider a utility function that balances (linearly) workload balance and latency and allows the assigning of weights to these two criteria. We assume the servers are placed at local cell towers, ensuring connectivity. We use the EDISON

method for finding the optimal edge server placement as it has been shown to provide improved results over other state-of-the-art methods.¹⁹ We first divide the area into a grid of 100×100 cells and then count the number of observations in each cell.

We assign the grid cells to their nearest base station, summing up the number of observations as the workload passes through the base station. Finally, we place the edge nodes using the PACK capacitated clustering algorithm, which is capable of balancing the cluster allocation.²⁰ The PACK method minimizes an objective function that sums the distances between base stations, weighted by their workload, and the locations of the edge servers. The minimization is conducted with constraints for workload balance (that is, the sum of the workloads of the base stations allocated to each edge node) and server homogeneity (ensured by a minimum number of workloads per edge node). The workload balance relates to the processing

needs of the deployment, whereas homogeneity determines the latency of the deployment. These two parameters are key to smart megacities as they determine the cost (including energy) of running the edge servers and their capability to react to data generated by sensors. We consider three scenarios with different tradeoffs between latency and workload balance: 1) workload balance prioritized over latency, 2) equal importance for workload balance and latency, and 3) latency prioritized over workload balance.

Results

Figure 2(a) shows the values of the PACK objective function for different numbers of edge servers. A lower objective function value indicates a better placement in terms of distance from sensors to edge servers and translates to lower latency for edge-based analytics on the sensor data. The knee point of the curve, that is, the point where the objective function stabilizes, is at 14 servers; additional

servers provide only marginal improvements. Figure 2(b) shows the final deployment of the 14 edge servers when latency and workload balance are weighted equally. In the figure, the base station to the immediate left of each edge server serves as the deployment location. The colored dots in Figure 2(b) show the locations where data are generated. The edge servers are placed along the major streets in the resulting deployment. In total, 11 edge servers are placed on the main street intersecting the area and running from north to southeast. The remaining three are located along the main horizontal street running from east to southwest. The sensor data are concentrated along these streets, demonstrating hotspots that would benefit from edge servers.

The three scenarios are further detailed in Table 2. Latency is calculated as the mean distance between edge servers and the allocated base stations, and balance is the standard deviation of the workload for the edge

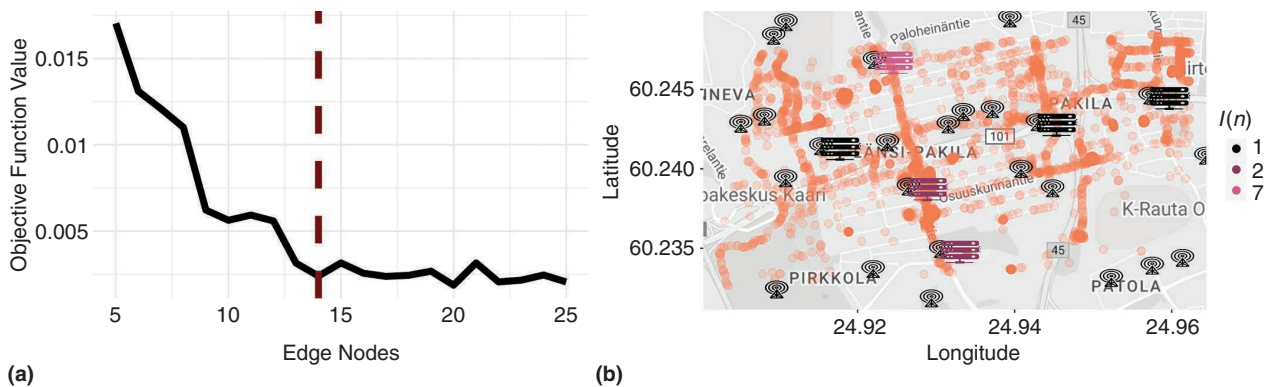


FIGURE 2. Deployment of edge servers. (a) Edge node placement objective function values; the lower the value, the better the placement. The dashed line at 14 edge nodes indicates the “elbow,” after which the objective function value does not significantly improve. (b) Edge node placement. To ensure connectivity, edge nodes are placed at cell towers (on the immediate left of the node icons). Some locations need more than one edge node to ensure sufficient computational capacity. The different colored circles represent different numbers of edge servers: the color ● refers to one edge server, the color ● indicates two edge servers, and the color ● means seven edge servers.

nodes in each scenario. The tradeoff between latency and balance is clearly visible, with improvements in latency decreasing workload balance and vice versa. Indeed, compared to the scenario where latency and workload are weighted equally, the scenario focusing on balance has a very homogeneous node distribution, reducing the standard deviation by 96% while increasing the latency by 31%. Conversely, the scenario emphasizing latency more than halves (-58%) the mean distance between base stations and edge servers, but it comes with a cost of a very high standard deviation in the resulting edge server workload (+689%).

The 40 citizens carrying the portable sensors were recruited through random sampling, and the resulting mobility patterns reflect the movements of the people in the sample. Generally, the patterns reflect mobility along the main roads in areas that contain main concentrations of shops, transportation options, and other functionality and hence can be argued to be representative of the mobility patterns within the district. Nevertheless, the server deployment is always relative to the data availability. Therefore, the deployment should be reoptimized in case there are differences in the mobility patterns of people within the district (for example, new constructions). In practice, the dominant mobility patterns are likely to remain; hence, scaling up deployments to changes would likely require placing new servers in expanding areas. In case more significant changes occur—for example, as in the case of the COVID-19 pandemic—the deployment would need to be reoptimized, and any application components running on the edge servers would need to migrate to other servers.

We also note that the optimization was solely based on data from mobile air quality sensors, and, in practice, the smart megacity would contain a mixture of mobile and static sensors, as envisioned in Figure 1. Adding sensors into (or removing them from) the district would naturally also require recalculating the optimal deployment, but, in practice, the resulting model would likely prioritize areas that are close to main intersections as they are natural points of focus for activity in urban areas.

DISCUSSION

Circular smart cities

Computing for environmental sustainability is currently an active area of research, and many sensing technologies for detecting potential ecological impacts of human behavior are emerging—ranging from climate to litter and overall use of resources. This indicates that the transition to circular economies is likely an important source of sensor data for emerging megacities. Carefully designing edge server placement is thus important for supporting sensor data analysis while ensuring the resulting analysis can be carried out cost-effectively and with minimal resources. The edge servers can further take advantage of public

infrastructure, for example, by being deployed at transportation hubs or public transport vehicles to provide ubiquitous computing support for citizens.

Benefits to city stakeholders

Municipal authorities, organizations, and companies have specific use cases and use the collected data to develop applications that serve the improvement of the city and its inhabitants. For example, an organization that operates massive air quality sensors can utilize the collected data for monitoring green travel routes and even offer pollution exposure analytics for citizens. Similarly, while UAVs are delivering parcels to their destinations, they can avoid generating CO₂ as the operating organization can use onboard air quality sensors or even perform video streaming during flight for real-time analysis of ground activities for the benefit of the smart city as a whole.

Intelligence on the edge

AI is bound to become an integral part of future 6G networks. Edge servers provide opportunities to deploy AI and machine learning models and enable analysis and predictions based on sensor data. This can increase the smartness of future megacities, for example, by predicting traffic on a specific route

TABLE 2. Three edge node placement scenarios exploring the tradeoff between latency and workload balance.

Focus	Latency (mean distance)	Difference	Balance (standard deviation of workload)	Difference
Balance	818	+31%	14	-96%
Middle	623	(baseline)	361	(baseline)
Latency	260	-58%	2,849	+689%

ABOUT THE AUTHORS

NASER HOSSEIN MOTLAGH is a docent and a researcher in the Department of Computer Science, University of Helsinki, Helsinki 00014, Finland. His research interests include the Internet of Things, wireless sensor networks, unmanned aerial vehicles, and autonomous underwater vehicles. Motlagh received a D.Sc. in networking technology from Aalto University, Finland. Contact him at naser.motlagh@helsinki.fi.

LAURI LOVÉN is a faculty university assistant at the Distributed Systems Group, TU Wien, Vienna, Austria, and a post-doctoral researcher at the Center for Ubiquitous Computing, University of Oulu, Oulu 90014, Finland. His current research concentrates on edge intelligence. Lovén received a D.Sc. (Tech.) from the University of Oulu. He is a Senior Member of IEEE. Contact him at lauri.lovén@oulu.fi.

JACKY CAO is a Ph.D. student at the University of Oulu, Oulu 90014, Finland. His research interests include mobile augmented reality, edge computing, and 5G networks. Cao received an MPhys. degree from Durham University. Contact him at jacky.cao@oulu.fi.

XIAOLI LIU is a postdoctoral researcher in computer science at the University of Helsinki, Helsinki 00014, Finland. Her research interests include big data analysis, artificial intelligence, the Internet of Things, and wearable computing. Liu received a Ph.D. in mathematics and statistics from the University of Helsinki. Contact her at xiaoli.liu@helsinki.fi.

PETTERI NURMI is an associate professor at the University of Helsinki, Helsinki 00014, Finland. His research interests include

distributed systems, pervasive data science, and sensing systems. Nurmi received a Ph.D. in computer science from the University of Helsinki. Contact him at petteri.nurmi@helsinki.fi.

SCHAHRAM DUSTDAR is a professor of computer science with the Distributed Systems Group, TU Wien, 1040 Vienna, Austria. His research interests include cloud computing, service-oriented computing, the Internet of Things, smart cities, and mobile and ubiquitous computing. Dustdar received a Ph.D. in business informatics (Wirtschaftsinformatik) from the University of Linz, Austria. Dustdar serves on the editorial boards of *IEEE Internet Computing* and *Computer*. He is a Fellow of IEEE. Contact him at dustdar@dsg.tuwien.ac.at.

SASU TARKOMA is a full professor in the Department of Computer Science, University of Helsinki, Helsinki 00014, Finland. His research interests include mobile computing, Internet technologies, and artificial intelligence. Tarkoma received a Ph.D. in computer science from the University of Helsinki. He is a Senior Member of IEEE. Contact him at sasu.tarkoma@helsinki.fi.

XIANG SU is an associate professor in the Department of Computer Science, Norwegian University of Science and Technology, Trondheim 7491, Norway and Center of Ubiquitous Computing, University of Oulu, Oulu 90014, Finland. His research interests include extensive expertise in the Internet of Things, edge computing, mobile augmented reality, knowledge representations, and context modeling and reasoning. Su received a Ph.D. in technology from the University of Oulu. Contact him at xiang.su@ntnu.no.

or even the exposure risk to a particular disease in an environment. Edge computing platforms using AI and learning can also optimize network resources and enable self-control and self-optimization of the network infrastructure to improve service quality and experience for end users.

Regulations

Massive-scale IoT deployments introduce challenges for regulations. For example, UAVs and other types of autonomous vehicles can disrupt a city's operations, disrupt citizens, and cause damage if they fail or are hacked. Regulations and security mechanisms are

needed to determine the specific operational routes, and the relevant authorities must create and enforce those rules and laws. Regulations are also needed to specify the integration of different sensors into smart megacities, for example, to avoid nefarious or obstructive operations and to preserve citizens' privacy.

Extensions

Our experiments focused on air quality monitoring as a representative example of smart city applications that require access to computing resources. Naturally, other application domains should also be investigated to better understand the requirements posed on smart city infrastructure. We plan to further explore this dimension and consider multimodal data combining traffic, weather, mobility, and surveillance data. Another potential extension is to investigate more complex utility functions that also consider parameters other than cost and latency, such as interference, bandwidth, average connectivity, and overall network load. This would not only benefit city planners but also provide network operators insights on how to further optimize network and edge placements.

The development of future smart megacities is far from straightforward as cities continue their unprecedented growth—both in the number of inhabitants and the number of networked devices available within the city. Current cloud-centered data processing models are insufficient as they rapidly become too costly for the needs of cities, even if they can offer sufficient elasticity. We presented a vision for using edge computing to satisfy the computing needs of emerging smart megacities while at the same time better supporting scalability and cost-effectiveness compared to cloud-centered models. We also described how optimizing the edge server placement is necessary to ensure the best possible computing resource availability. Overall, our research demonstrates how edge computing deployments are essential for

fulfilling smart megacities' increasing data processing requirements and ensuring the business models remain feasible for offering the broadest range of services to citizens and the cities themselves. ■

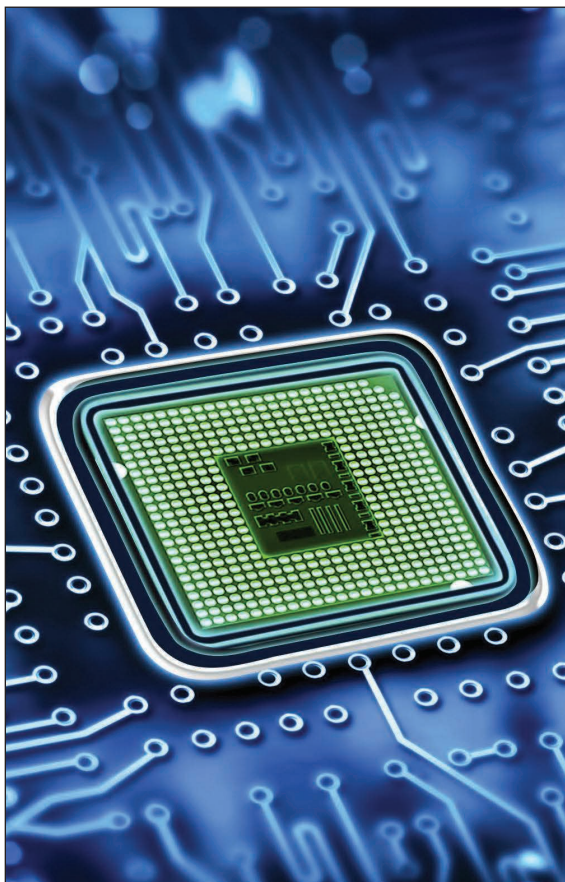
ACKNOWLEDGMENT

This work is supported by the Nokia Center for Advanced Research and the Academy of Finland under grants 335934, 345008, 339614, 319670, 325774, and 326305.

REFERENCES

1. "World's population increasingly urban with more than half living in urban areas," United Nations, New York, NY, USA, 2014. Accessed: Apr. 12, 2022. [Online]. Available: <https://www.un.org/development/desa/en/news/population/world-urbanization-prospects.html>
2. N. H. Motlagh et al., "Toward massive scale air quality monitoring," *IEEE Commun. Mag.*, vol. 58, no. 2, pp. 54–59, Feb. 2020, doi: 10.1109/MCOM.001.1900515.
3. N. H. Mahmood et al., "White paper on critical and massive machine type communication towards 6G," 2020, *arXiv:2004.14146*.
4. L. U. Khan, I. Yaqoob, N. H. Tran, S. A. Kazmi, T. N. Dang, and C. S. Hong, "Edge-computing-enabled smart cities: A comprehensive survey," *IEEE Internet Things J.*, vol. 7, no. 10, pp. 10,200–10,232, Oct. 2020, doi: 10.1109/JIOT.2020.2987070.
5. D. Xu et al., "Edge intelligence: Empowering intelligence to the edge of network," *Proc. IEEE*, vol. 109, no. 11, pp. 1778–1837, Nov. 2021, doi: 10.1109/JPROC.2021.3119950.
6. X. Su et al., "Intelligent and scalable air quality monitoring with 5G edge," *IEEE Internet Comput.*, vol. 25, no. 2, pp. 35–44, Mar./Apr. 2021, doi: 10.1109/MIC.2021.3059189.
7. M. Satyanarayanan, G. Klas, M. Silva, and S. Mangiante, "The seminal role of edge-native applications," in *Proc. 2019 IEEE Int. Conf. Edge Comput. (EDGE)*, pp. 33–40, doi: 10.1109/EDGE.2019.00022.
8. R. Y. Brzezinski et al., "Automated processing of thermal imaging to detect COVID-19," *Sci. Rep.*, vol. 11, no. 1, pp. 1–10, Sep. 2021, doi: 10.1038/s41598-021-96900-9.
9. F. Concas et al., "Low-cost outdoor air quality monitoring and sensor calibration: A survey and critical analysis," *ACM Trans. Sens. Netw.*, vol. 17, no. 2, pp. 1–44, May 2021, doi: 10.1145/3446005.
10. M. A. Zaidan et al., "Intelligent calibration and virtual sensing for integrated low-cost air quality sensors," *IEEE Sensors J.*, vol. 20, no. 22, pp. 13,638–13,652, Nov. 2020, doi: 10.1109/JSEN.2020.3010316.
11. F. Greenwood, E. L. Nelson, and P. G. Greenough, "Flying into the hurricane: A case study of UAV use in damage assessment during the 2017 hurricanes in Texas and Florida," *PLoS One*, vol. 15, no. 2, p. e0227808, Feb. 2020, doi: 10.1371/journal.pone.0227808.
12. M. Giordani, M. Polese, M. Mezzavilla, S. Rangan, and M. Zorzi, "Toward 6G networks: Use cases and technologies," *IEEE Commun. Mag.*, vol. 58, no. 3, pp. 55–61, Mar. 2020, doi: 10.1109/MCOM.001.1900411.
13. R. T. Azuma, "A survey of augmented reality," *Presence, Teleoper. Virtual Environ.*, vol. 6, no. 4, pp. 355–385, 1997, doi: 10.1162/pres.1997.6.4.355.
14. E. Bastug, M. Bennis, M. Médard, and M. Debbah, "Toward interconnected virtual reality: Opportunities, challenges, and enablers," *IEEE Commun. Mag.*, vol. 55, no. 6,

- pp. 110–117, Jun. 2017, doi: 10.1109/MCOM.2017.1601089.
15. W. Saad, M. Bennis, and M. Chen, “A vision of 6G wireless systems: Applications, trends, technologies, and open research problems,” *IEEE Netw.*, vol. 34, no. 3, pp. 134–142, May/June 2019, doi: 10.1109/MNET.001.1900287.
 16. T. S. Rappaport et al., “Wireless communications and applications above 100 GHz: Opportunities and challenges for 6G and beyond,” *IEEE Access*, vol. 7, pp. 78,729–78,757, Apr. 2019, doi: 10.1109/ACCESS.2019.2921522.
 17. P. Kortoçi et al., “Air pollution exposure monitoring using portable low-cost air quality sensors,” *Smart Health*, vol. 23, p. 100,241, Mar. 2022, doi: 10.1016/j.smhl.2021.100241.
 18. A. Rebeiro-Hargrave, N. H. Motlagh, S. Varjonen, E. Lagerspetz, P. Nurmi, and S. Tarkoma, “MegaSense: Cyber-physical system for real-time urban air quality monitoring,” in *Proc. 2020 15th IEEE Conf. Ind. Electron. Appl. (ICIEA)*, pp. 1–6, doi: 10.1109/ICIEA48937.2020.9248143.
 19. L. Lovén et al., “EDISON: An edge-native method and architecture for distributed interpolation,” *Sensors*, vol. 21, no. 7, p. 2279, Mar. 2021, doi: 10.3390/s21072279.
 20. T. Lähderanta et al., “Edge computing server placement with capacitated location allocation,” *J. Parallel Distrib. Comput.*, vol. 153, pp. 130–149, Jul. 2021, doi: 10.1016/j.jpdc.2021.03.007.



IEEE TRANSACTIONS ON

COMPUTERS

Call for Papers: *IEEE Transactions on Computers*

Publish your work in the IEEE Computer Society's flagship journal, *IEEE Transactions on Computers*. The journal seeks papers on everything from computer architecture and software systems to machine learning and quantum computing.

Learn about calls for papers
and submission details at
www.computer.org/tc.

