Towards Large-Scale IoT Deployments in Smart Cities: Requirements and Challenges

Naser Hossein Motlagh, Martha Arbayani Zaidan, Roberto Morabito, Petteri Nurmi, and Sasu Tarkoma

Department of Computer Science, University of Helsinki,
Pietari Kalmin katu 5, 00560 Helsinki, Finland
{naser.motlagh,martha.zaidan,roberto.morabito,petteri.nurmi,sasu.tarkoma}@helsinki.fi

Abstract. The Internet of Things (IoT) plays a significant role in the development and future evolution of smart cities by connecting physical devices and systems to the Internet to collect and exchange data, automate processes, and improve overall urban management, and quality of life. This chapter presents the requirements and challenges to realize IoT deployments in smart cities, including sensing infrastructure, Artificial Intelligence (AI), computing platforms, and enabling communications technologies such as 5G beyond networks. To highlight these challenges in practice, the chapter also presents a real-world case study of a city-scale deployment of IoT air quality monitoring within the city of Helsinki. The results demonstrate the role that IoT plays in future smart cities, illustrating how deployments of air quality monitoring devices can benefit decision-making by supporting local air pollution monitoring, traffic management, and urban planning. Lastly, the chapter discusses the role of AI and other emerging technologies in the future of smart cities.

Keywords: Internet of Things, Smart Cities, Artificial Intelligence, Sensor Deployment, 6G Networks

1 Introduction

The number of Internet of Things (IoT) devices has long since surpassed the number of people on Earth and is expected to continue growing with estimates suggesting nearly 30 billion devices will be deployed by 2030 [1]. Cities and urban areas are one of the main areas for these devices with examples ranging from smart home sensors to driverless cars, portable IoT devices, smart wearables, and different types of drones. Examples of these devices in operation within a smart city are shown in Figure 1.

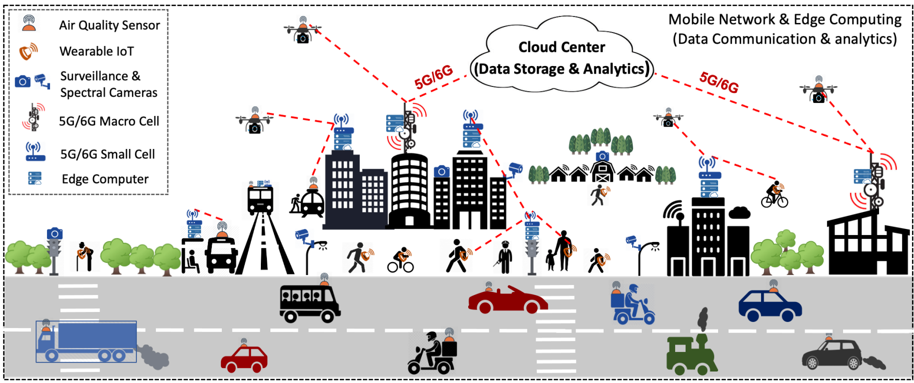


Fig. 1: Illustration of large-scale sensor deployment in a smart city.

The characteristics of the IoT devices vary depending on the device designs and their intended applications, which in turn poses requirements for the infrastructure that is available in the city. For example, driverless cars require continuous and persistent network connections, whereas wearables typically require discontinuous and transient connections. Similarly, applications that target the immediate needs of citizens tend to require support for real-time computation and processing, whereas analytics and other more long term services can operate without support for real-time processing. Besides the need for real-time responsiveness of the networks, some of these applications would be computationally demanding. Providing the necessary networking and computational support in an affordable, efficient and scalable manner is highly challenging [2]. Besides these overall infrastructure challenges, deploying the sensors can also be demanding. IoT devices that benefit the city mostly can be categorized into fixed sensors and mobile sensors. Fixed sensors require strategic planning for deployment and to ensure the necessary electricity, networking, computations, and security support are in place. Mobile sensors, in turn, need to have sufficiently dense coverage and data quality may be an issue as certain locations or demographic groups may be overrepresented.

Taking all of the above into account, massive-scale deployments of IoT sensors in smart cities that meet the needs of citizens and applications is a highly challenging task. This chapter details these challenges, beginning from requirements (Section 2) and continuing to key challenges (Section 3). To highlight some of the potential benefits that can be obtained from IoT deployments in smart cities, in Section 2 we present a case study and results from a deployment of air quality sensors in the city of Helsinki. We further provide a discussion about the role of AI and emerging technologies in future smart cities in Section 5. Finally, we conclude the chapter in Section 6.

2 Requirements for IoT Deployment in Smart Cities

2.1 Reliable Network Connection

To ensure successful operations of deployed IoT devices in smart cities, it is mandatory to have robust and seamless network services in the cities. While some IoT applications would require ultra-low latency services from the network, other applications may demand high bandwidth or may need to obtain massive connections [3]. The following are examples of applications requiring different forms of services from the network.

Ultra-low Latency: The driverless cars and drones are outstanding example applications that would need fast data processing in order to make precise decision-making, e.g., for avoiding obstacles and changing directions. To ensure the safe operations of the applications, thus, driverless cars and drones are expected to have stringent latency requirements [2, 4].

High Bandwidth: The surveillance cameras have been widely used in urban areas to monitor human activities as well as face recognition [5]. To perform real-time image processing streamed from the cameras therefore there is a need for enhanced bandwidth from the network such that it can support the transmission of tens of video frames every second while each frame requires a few Megabits of bandwidth from the network. Assuming a frame size of 20KB and 30fps be the standard frame rate, then the required bandwidth for a single frame would be 4.8Mb/s (20KB x 30fps x 8 bits/byte). This bandwidth requirement is further enhanced with the frame transmission rate of the surveillance camera. Hyperspectral cameras that are widely used for environmental and pollution monitoring are the other prominent examples of IoT applications as they can produce images of 30–300 MB in less than a second [6, 7]. Therefore, for frame transmission, compared to the surveillance cameras they require even higher bandwidth from the network.

Massive Connection: In addition to the example IoT devices mentioned earlier, the number of other types of IoT devices and applications rapidly increase which mandates obtaining ubiquitous and responsive network services in cities. Among many, examples of such applications include portable low-cost air quality sensors, smart homes, smart grids, smart metering, and different forms of wearables such as smartwatches and smart rings. The increasing number of IoT devices either mobile (carried by people or vehicles) or installed at fixed locations requires providing massive connections by the networks [4].

2.2 Infrastructure Deployment

To provide effective network services and ensure successful operations of IoT it is necessary to first, place the network infrastructure such as base stations, IoT gateways, and edge computers in strategic locations (PoIs) to enable providing full network coverage in cities [8]. This is needed to ensure support for mobile IoT devices (either vehicles, drones, or people) moving at various speeds.

Second, when deploying fixed sensors in urban areas, it is essential to install them in places that can easily connect to the network and maintain its connection. In addition, as IoT devices require power supplies as well as continuous maintenance, it is important to install the devices in locations with energy sources and easy access for inspection.

3 Key Aspects of Sensor Deployment and Data Management in Smart Cities

3.1 Sensor Deployment and Placement

Urban environments are complex systems as they consist of different urban elements such as residential areas, shopping centers, parks and green areas, and highways and streets (with high and low levels of traffic). These urban environments do not only span horizontally, they grow vertically (such as tall buildings and skyscrapers) with the population growth in cities. Therefore, to optimally provide IoT services [9] and also better monitor the health of city infrastructures [10], there is a need for optimal sensor deployments and placement methods in order to cover the whole city environment.

In the existing methods, the solutions include "citizen-centric" sensor placement approach by i) installing sensors near public places, e.g., schools and hospitals, ii) providing local information by minimizing the distance between the sensors and the people, and iii) placing and optimizing sensors on critical urban infrastructure, e.g., monitoring traffic emissions on roads with high traffic levels [11].

In addition, current sensor deployment and placement the most areas of a city are not covered. The areas that fall under a certain radius of a sensor are considered covered by sensing systems. Therefore, to cover the missing areas, the current approaches rely on interpolating data using the measurements of other sensor nodes in the same area. Indeed, the city environments because of their complex features and dynamics make sensor deployment challenging. Thus, sensor deployment and placement require new models that take into account the dynamics of the city blocks, urban infrastructure, building shapes, demographics, and the micro-environmental features of the regions.

In light of the challenges associated with sensor deployment and placement outlined in this section, it is crucial to consider the broader ecosystem in which these sensors operate. Effective sensor deployment is but the first step in a multi-faceted process that ultimately leads to the delivery of valuable services and applications within smart cities.

Figure 2 provides an illustrative overview of this ecosystem, segmented into four primary layers: Data Collection, Data Transmission, Data Services, and Applications. Each layer represents a critical stage in the data lifecycle, with its unique challenges and requirements.

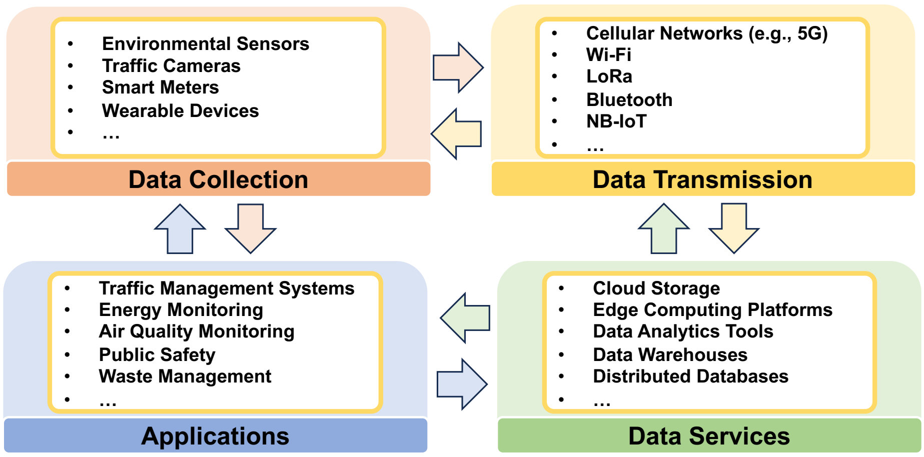


Fig. 2: Illustration of the four primary layers in smart city data management: Data Collection, Data Transmission, Data Services, and Applications, each with representative examples.

Each of these layers is interconnected, collaboratively ensuring that data is effectively collected, transmitted, managed, and utilized to provide intelligent and responsive smart city applications.

Within this complex framework, security and ethical considerations permeate every layer. The process of data handling often involves sensitive or personally identifiable information, necessitating stringent ethical considerations and robust security measures. Techniques like data anonymization are implemented to protect privacy, while adherence to international and local legal frameworks, like the GDPR in Europe, guide the ethical collection and handling of data [12]. Security considerations are equally crucial, involving the deployment of encryption technologies and access control mechanisms to safeguard data at rest and in transit, providing a secure environment for data storage and processing [13].

The following sections will delve deeper into the challenges and considerations associated with Data Collection, Data Transmission, and Data Services within this secured and ethically compliant framework. Then, in Section 4, we will explore a practical application of this layered framework through a case study on *Air Quality Monitoring* with IoT for Smart Cities, offering real-world insights into how these layers function consistently to support smart city initiatives while upholding the highest standards of security and ethics.

3.2 Data Collection

Data collection is the foundational component in the IoT lifecycle within smart city applications, requiring robust and efficient processes to ensure the efficacy of subsequent analytics and decision-making. In the realm of IoT, data collection entails gathering various types of data from devices like environmental sensors, traffic cameras, smart meters, wearable devices, and RFID tags, as illustrated in Figure 2.

Each device plays a specific role in collecting different data types, which are essential for various applications in smart cities. For instance, environmental sensors gather crucial data on air quality, temperature, and humidity, providing real-time information necessary for monitoring and responding to changes in the urban environment.

To facilitate reliable and efficient data collection, adherence to established protocols and standards is crucial. Protocols like MQTT and CoAP [14], while also playing a role in the transmission, are fundamental at the collection stage for ensuring data is gathered and packaged correctly for transmission. MQTT is notable for its lightweight characteristics, making it ideal for scenarios with limited bandwidth, high latency, or unreliable networks. CoAP, used for devices in constrained environments, simplifies data transmission at the initial collection point.

Interoperability is another crucial factor at the data collection stage [15], ensuring that various devices can communicate and share data effectively. Interoperability not only considers the compatibility between different device types but also the protocols and standards they use, fostering a seamless and efficient data collection process [16]. Initiatives and efforts, such as those led by the Internet Engineering Task Force (IETF) and many other standardization bodies (e.g., 3GPP, IEEE, etc.), actively work towards standardization to ensure that different protocols, data formats, and devices can effectively interoperate with one another [15, 17].

3.3 Data Transmission

Efficient data transmission is critical in the deployment of IoT systems within smart cities, as it acts as the bridge between data collection and data services. The significance of effective data transmission lies in the necessity for real-time (or near real-time), accurate, and secure transmission of data from myriad IoT devices to their respective end-points.

The challenges in data transmission are multiple. Applications within smart cities necessitate the transmission of a wide and varied volume of data, requiring robust and adaptable networks [18]. The latency in data transmission, or the delay in data transfer, becomes particularly significant for applications that mandate immediate or real-time responses. Limited bandwidth is another substantial challenge, often stressed in areas densely populated with devices simultaneously transmitting data.

Additionally, the heterogeneity of transmission technologies introduces complexity. Various technologies, including LoRa, Wi-Fi, Bluetooth, LTE-M, NB-IoT, and 5G, offer different advantages and challenges [19]. For instance, while LoRa provides long-range connectivity and low power consumption, it might not offer the high data rates required for some applications. Conversely, 5G provides high data rates and low latency, supporting applications with demanding throughput and responsiveness requirements.

Smart city applications can also be characterized by different requirements [20], aligning with the categorizations provided by 5G networks. Ultra-Reliable Low Latency Communications (URLLC) is crucial for applications that require immediate responses with minimal delay. Enhanced Mobile Broadband (eMBB) caters to applications that need high data rates and bandwidth. Finally, massive Machine Type Communications (mMTC) is essential for supporting a massive number of connected devices, typically seen in densely populated urban areas.

To address these challenges, it is fundamental to deploy and use optimized data transmission protocols and technologies, ensuring each application's unique requirements are met. Techniques like data compression can be utilized to reduce the amount of data transmitted, saving bandwidth and improving transmission efficiency.

3.4 Data Services

Data services play an fundamental role in the framework of IoT within smart cities, offering a wide set of functionality that are essential for effectively managing and utilizing the data gathered. Within this landscape, we identify four main components belonging to data services: *Data Storage*, *Data Processing*, *Data Analytics*, and *Data Sharing and Access*. These components are interconnected, each playing a critical role while collaboratively working to ensure that data flows seamlessly through the system from collection to actionable insight, ultimately serving as the backbone for various smart city applications.

Data Storage and **Data Processing** are pivotal in the IoT lifecycle within smart cities [21], serving as the repository and analysis mechanism for the vast data generated. Efficient and secure data storage solutions are essential due to the immense volume of data continuously produced by various IoT devices. These solutions must guarantee data integrity, swift retrieval times for real-time applications and robust security to protect sensitive information from unauthorized access and potential breaches. On the processing end, transforming the raw data into actionable insights presents its challenges. First, there is a demand for substantial computational power to analyze and process the collected data efficiently. Quality control of the data is also paramount; ensuring accuracy is crucial for reliable analysis and insights. Strategies and technologies must be in place to handle incomplete or "noisy" data, requiring sophisticated data cleaning and validation processes. Additionally, for real-time applications, minimizing latency—from data collection to insight generation—is critical.

Several technologies and strategies have emerged to address the challenges associated with data storage and processing. Cloud computing [22] offers a viable solution, providing scalable storage and computing resources. This technology is particularly well-suited for applications without stringent latency requirements. For applications demanding real-time data processing, edge computing [23] offers a solution by processing data closer to its generation point, thereby reducing latency and conserving bandwidth. Data warehouses and distributed databases also play a crucial role [24]. Data warehouses serve as centralized repositories that store integrated data from various sources, designed mainly for query and analysis. In contrast, distributed databases provide a framework for storing and processing large data volumes across a network of computers, offering scalability and fault tolerance.

Data Analytics takes the processed data to the next level by employing advanced tools and algorithms to interpret and analyze it for patterns, trends, and hidden insights. While data processing prepares and refines the data, data analytics is concerned with drawing meaningful conclusions and providing foresight and understanding that inform decision-making processes. Within this framework, technologies like AI and ML play a significant role in providing deeper insights, offering predictive analytics and facilitating more informed and proactive decision-making and planning in the urban context. This process encompasses three main analytics types: descriptive, predictive, and prescriptive [19, 25]. *Descriptive analytics*, commonly utilized in business, measures, and contextualizes past performance to aid decision-making. It brings out hidden patterns and insights from historical data but isn't primarily used for forecasting. *Predictive analytics*, on the other hand, goes beyond description, extracting information from raw data to identify patterns and relationships, thereby facilitating forecasts of behaviors and events. Using both historical and current data, predictive analytics provides valuable foresights. *Prescriptive analytics* advances further, quantifying the potential effects of future decisions to provide recommendations and insights on possible outcomes. This advanced analytics type supports decision-making by offering choices and suggestions based on data analysis, making it a crucial tool for planning and strategy in smart cities.

However, the integration of big data analytics necessitates a clear understanding of specific functional and non-functional requirements [26, 27], given the diverse and dynamic nature of data sources and applications within smart cities. Functional requirements encompass aspects like interoperability, real-time monitoring, access to historical data, mobility, service composition, and integrated urban management. On the other hand, non-functional requirements include sustainability, availability, privacy considerations, social impact, and scalability. Addressing these requirements is imperative for developing robust and resilient smart city architectures that can seamlessly integrate and analyze data from heterogeneous sources, including IoT sensors, social media networks, and electronic medical records. Furthermore, the dynamic urban environment of smart cities demands attention to stream data analytics, enabling real-time services while also accommodating planning and decision-making processes through historical or batch data analytics. Essential characteristics that a big data analytics platform should embody to navigate the challenges of big data include scalability, fault tolerance, I/O performance, real-time processing capabilities, and support for iterative tasks.

Effective and secure **Data Sharing and Access** is key to maximizing the utility of data in smart cities. This involves making collected data available to authorized entities, departments, or individuals who require it for various applications and analytics, always with robust data access policies and mechanisms in place to ensure both data sharing and privacy protection. Data sharing in the context of smart cities encompasses a set of technologies, practices, and frameworks aimed at facilitating secure and efficient data access amongst multiple stakeholders without compromising data integrity [28]. This process is integral to improving efficiency and fostering collaboration not only within city departments but also with external partners, vendors, and the community at large, all while being aware of and mitigating associated risks. There are at least two main factors that strengthen the importance of data sharing in smart cities. The first relates to the possibility of integrating data from different sources, which can possibly enhances the value and performance of dedicated services [29]. For instance, data sharing enables improved urban planning and transportation management by combining information from traffic cameras, sensors, and public feedback, leading to more effective and responsive city services. The second is linked to a more effective *Data-Driven Decision-Making.* Transparent information sharing facilitates improved analytics, enabling city officials and stakeholders to make informed and effective long-term decisions [30]. For example, integrating data from environmental sensors, healthcare institutions, and public service departments can provide a holistic view of city health and environmental conditions, aiding in timely decision-making and policy formulation.

However, the process of data sharing is not without challenges. Risks include potential privacy disclosure, where organizations must navigate legal and ethical obligations to protect customer data while sharing information responsibly. The process also opens up possibilities of data misinterpretation and issues related to data quality, including hidden biases in datasets [28].

In mitigating risks and facilitating data sharing in smart cities, several technologies are essential. Among these, Data Warehousing is crucial for internal data sharing, serving as a repository for data from various departments and allowing isolated access to shared information [28]. Next, APIs play a key role by enabling fine-grained communication and controlled data sharing between software components. They precisely dictate accessible data and usage rules, ensuring structured and secure data exchange [31]. Lastly, Federated Learning is transformative, allowing collaborative AI and ML development while maintaining data control and privacy for each contributor. This approach not only enhances data-driven insights but also ensures confidentiality, supporting robust and intelligent smart city applications [32].

While Data Services provide the foundational support for various applications in smart cities, the effectiveness of these applications is highly dependent upon the quality of the data being collected, transmitted, and analyzed. The following section, therefore, will delve into the topic of Data Quality, exploring the challenges and considerations related to ensuring the accuracy, reliability, and validity of data in smart city ecosystems.

3.5 Data Quality

An IoT application may comprise hundreds or thousands of sensor devices that produce vast amounts of data. This data is rendered useless if it is riddled with errors as poor sensor data quality caused by the errors may lead to wrong decision-making results. In order to enable massive deployment, most IoT applications use low-cost sensor techniques, though at the expense of data quality. As a result, IoT often encounters soft faults (i.e., error) which are associated to outliers, bias, drifts, missing values, and uncertainty, which should be detected or quantified and removed or corrected in order to improve sensor data quality [33]. Due to the diverse nature of IoT deployments and the likelihood of sensor failures in the wild, a key challenge in the design of IoT systems is ensuring the integrity, accuracy, and fidelity of sensor data [34]

The error within an IoT application may take place for different reasons. For example, in a sensor network serving an IoT application, poor data quality may arise from congested and unstable wireless communication links and can cause data loss and corruption [35]. The other example pertains to the damage or exhaustion of battery in sensor devices that would cause the data quality to degrade, as towards the end of its battery life, sensors tend to produce unstable readings [36].

In addition, the role of external factors such as the hostile environment is not negligible on sensor readings and data quality. For example, air quality IoT devices that include aerosol, trace gases, and meteorological sensors are often placed outdoors and are subjected to extreme local weather conditions such as strong winds and snow, which might affect the operation of the sensor [37].

In IoT datasets, one of the most common data quality problems is called missing data (incomplete data) which indicates a portion of data that is missing from a time series data [38]. In principle, the missing data may be caused by different factors such as unstable wireless connection due to network congestion, sensor device outages due to its limited battery life, environmental interferences e.g. human blockage, walls, and weather conditions, and malicious attacks [39].

To cover missing data, one solution can be to re-transmit the data. However, since most IoT applications are in real-time, therefore, the data re-transmission would not be effective as i) rendering the data is not beneficial if there is a delay, and ii) the re-transmission adds to the computation and energy costs. The latter is due to the fact that the sensor devices are usually limited in terms of battery, memory, and computational resources. However, to fill in the missing data an alternative would be applying imputation based on Akima Cubic Hermite [40] and multiple segmented gap iteration [41] methods.

Another common problem that involves data quality is called outlier which can be in the forms of anomalies [37, 42] and spikes [43, 44]. An outlier takes place when sensor measurement values exceed thresholds or largely deviate from the normal behavior provided by the model. In other words, the outlier occurs when the sensor measurement value is significantly different from its previous and next observations or observations from neighboring sensor nodes [45, 46]. In practice, outliers can be identified by applying anomaly detection methods based on adaptive Weibull distribution [37] and Principal Component Analysis (PCA) [47, 48].

In addition to the outliers, another common problem in IoT data quality is known as bias or offset [49], which occurs when the sensor measurement value is shifted in comparison with the normal behavior of a reference sensor. A drift is a specific type of bias that takes place when the sensor measurement values deviate from their true values over time. Drifts are usually caused by IoT device degradation, faulty sensors, or transmission problems [50]. In current solutions, the drifts caused by any reasons can be detected by comparing two types of Bayesian calibration models [51] or applying ensemble classifiers where each classifier will learn a normal behavior model and compare it with the current reading [52]. In order to correct the bias and drift, calibrations are usually required [51]. For example, air-quality low-cost sensors often experience bias and drift in the field due to the sensors' device quality and variations in environmental factors. The sensors can then be calibrated using machine learning (ML) models, such as nonlinear autoregressive network with exogenous inputs (NARX) and long short-term memory (LSTM), to improve data quality and meet the data quality of reference instruments [40].

With a clearer understanding of the importance of Data Quality analysis, and having navigated through the various challenges and solutions crucial to each aspect of the data lifecycle in IoT as summarized in Table 1, we move forward to explore how the concepts and challenges discussed thus far manifest in real-world scenarios. The next section provides a practical perspective through a case study on *Air Quality Monitoring with IoT for Smart Cities*. This case study offers a valuable understanding into the application of data collection, transmission, services, and quality principles in the development and implementation of smart city applications, serving as a tangible example of theory translated into practice.

4 Case Study: Air Quality Monitoring with IoT for Smart Cities

This section presents a case study where IoT devices were used for an air quality monitoring network in Helsinki, Finland, a well-known smart city. Air pollution is known to be harmful to human health and the environment. According to the World Health Organization (WHO), air pollution causes approximately 7 million in deaths each year. Of this, an estimated 4.2 million deaths are due to outdoor exposure [53]. Official air quality monitoring stations have been established across many smart cities around the world. Unfortunately, these monitoring stations are sparsely located and consequently do not provide high-resolution spatio-temporal air quality information [54]. Thanks to advances in communication and networking technologies, and the Internet-of-Things (IoT), low-cost sensors have emerged as an alternative that can be deployed on a massive scale in

Table 1: A summary of key challenges and solutions for deploying massive IoT in smart cities.

Concern	Key Challenge	Solution
Sensor Deployment and Placement	Identifying the key locations and finding optimal places that allow the most coverage	Current solutions use interpolations (of data collected from other nodes in the sensor network). The need for enhanced methods that consider the population, urban, and environmental factors
Data Collection	Gathering various types of data that are reliable and efficient and interoperability	i) Establishing protocols, such as MQTT and CoAP, to ensure reliability and efficiency ii) Establishing standardization to ensure that different protocols, data formats, and devices can effectively interoperate with one another
Data Transmission	The latency in data transmission, limited bandwidth, limited connectivity, and heterogeneity of transmission technologies	i) 5G networks also provide URLLC, eMBB, and mMTC to respond to the minimal delay, high bandwidth, and massive connection requirements, respectively. ii) Optimizing data transmission protocols and technologies e.g., by applying data compression methods to reduce the amount of transmission data and save bandwidth
Data Services	i) Data storage and data processing to handle the immense volume of generated IoT data ii) Data analytics to draw meaningful conclusions and provide foresight for decision-making processes iii) Data sharing and access to maximize the utility of data	i) Cloud computing provides scalable storage and computing resources, and edge computing offers processing data closer to its generation point, data warehouses, and distributed databases facilitate storing and processing large data volumes ii) Technologies like AI and ML can play a significant role in providing deeper insights, offering predictive analytics and facilitating more informed and proactive decision-making and planning iii) Integrating data from different sources can enhance the value and performance of dedicated services, and transparent data sharing can improve analytics and lead to more effective decision-making
Data Quality	The occurrences and identification of poor data quality, e.g., missing data, outlier bias, and drift data	i) Identification of anomalous and poor data quality through drift detection ii) Correcting the data by applying imputation and calibration methods

cities [40]. This deployment offers a high resolution of spatio-temporal air quality information [6]. This case study demonstrates how air quality IoT devices benefit several aspects in terms of local pollution monitoring, traffic management, and urban planning.

4.1 IoT Installation

This subsection describes the experimental details including the sites, IoT devices, and the data collected from the experiments.

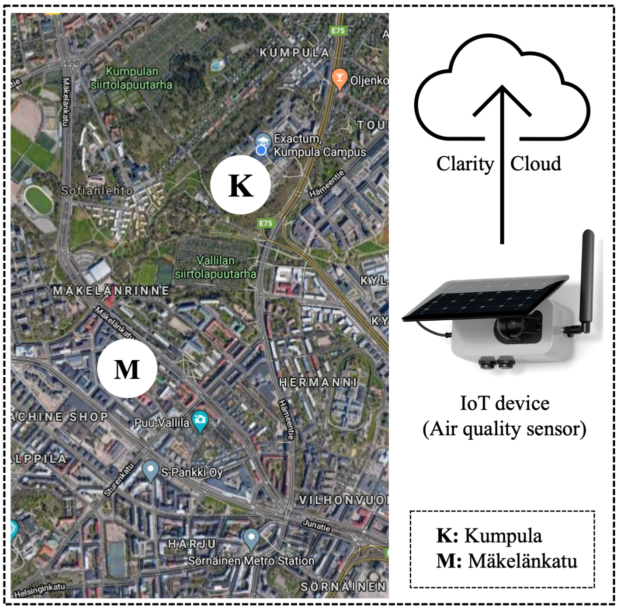


Fig. 3: The sites and the IoT devices used in the experiment.

Experimental sites

In this case study, two air quality IoT devices were installed at the following two different sites in the city of Helsinki, Finland. These sites include:

1. The Kumpula site that is located at Kumpula campus of the University of Helsinki in the front open yard and about 4 kilometers northeast of the Helsinki center. The site is also considered as an urban background, that is situated at about 150 meters from a main street in Kumpula district in Helsinki [55].

2. The Mäkelänkatu site is known as a street canyon and is located just beside Mäkelänkatu Street, which is one of the arterial roads and is lined with apartment buildings. The street consists of six lanes, two rows of trees, two tramlines, and two pavements, in a total of 42 meters of width. Every day, different types of vehicles including cars, buses, and trucks cross this street and thus cause frequent traffic congestion [56].

The map of both sites is presented on the left-hand side picture in Figure 3. The Kumpula site is notated by **K**, whereas the Mäkelänkatu site is notated by **M**. The distance between the two sites is 900 meters.

IoT devices

Air quality IoT devices used in this experiment are developed by Clarity Corporation, a company that is based in Berkeley, California, the USA. These IoT devices are shown on the right-hand side of Figure 3. The weight of the device is 450 grams. The input power of the sensor is 5 volts. The sensor device is designed to operate by battery and has a battery lifetime of 15 days of continuous measurements. If the battery operates by harvesting solar power, its operation time extends to 1 to 2 years. In our experiment, we used grid electricity for the sensor's input power. The sensors offer sensing meteorological variables including the Temperature (Temp) which uses band-gap technology and Relative Humidity (RH) which uses capacitive technology. The sensors also measure particulate matter (PM) and CO_2 with laser light scattering technology and metal oxide semiconductor technologies, respectively.

The sensors underwent a laboratory calibration process, by the manufacturer, using Federal Reference Method (FRM) instruments. The sensors are equipped with the LTE-4G communication module to transmit the measured data. The transmitted data is also stored in a cloud platform facilitated by Clarity[1]. The cloud platform allows access to the raw sensor and visualized data. The data can also be downloaded using a user interface accessible by SmartCity WebApp[2]. The measurement frequency of data varies around 16-23 minutes per data point. We installed one of these IoT devices on a container at the Kumpula site (**K**) about 2 meters from the ground level and another one at the Mäkelänkatu site (**M**) on the top of a container about 4 meters above the ground level.

The data

We collected the datasets from 1st January to 31st December 2018 from the two IoT devices. For our analysis, in this chapter, we use $PM_{2.5}$ and PM_{10}, and Air Quality Index (AQI) variables, extracted from the datasets. In our analysis, we process the data in an hour resolution. In practice, AQI is defined as the maximum of the indexes for six criterion pollutants, including PM_{10}, $PM_{2.5}$, CO, NO_2, O_3, and SO_2 [57].

[1] smartcity.clarity.io
[2] clarity.io/documents

4.2 Air quality IoT monitoring for a smart city

This subsection explains how air quality IoT devices can benefit a smart city using the analysis extracted from the IoT experiments. These benefits include local air pollution monitoring, traffic management, and urban planning.

Local air pollution monitoring

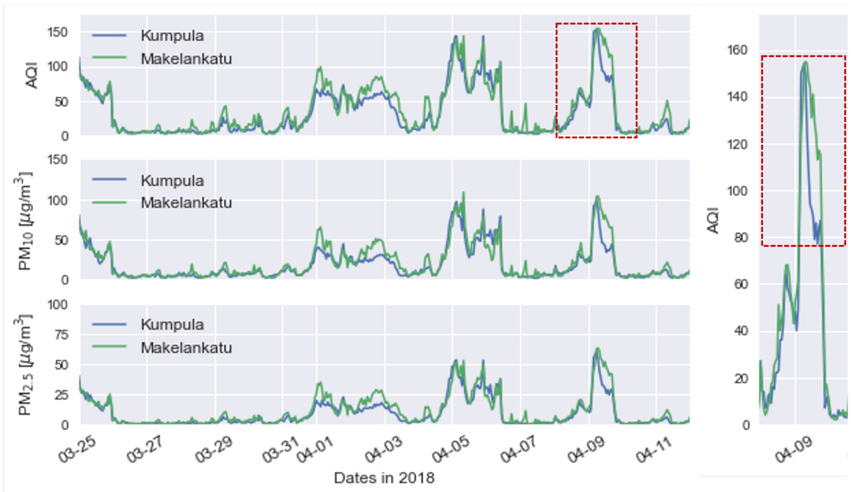


Fig. 4: Time-series data of AQI, PM_{10} and $PM_{2.5}$ concentrations (in $\mu g/m^3$) at Kumpula (K) and Mäkelänkatu (M) sites.

One of the key motivations for deploying dense air quality IoT devices in city districts is to provide local air pollution monitoring at fine-grained resolution. In principle, in urban areas the quality of air changes even at a few ten meters of distance. To show such a variation, we extract measurements of AQI, $PM_{2.5}$, and PM_{10} from our two IoT devices, between 25th March and 11th April 2018. Then, as illustrated in Figure 4, we plot the time series of these variables. In the figure, the blue color presents the measurements from the Kumpula site, and the green color portrays the air quality captured at Mäkelänkatu site. In the figure, the top subfigure shows the AQI variations, and the middle and bottom subfigures depict the PM_{10} and $PM_{2.5}$ concentrations, respectively.

As shown in the plots, in general, both measurements have similar patterns. The green curves lie slightly above the blue curves most of the time, indicating that the pollution level in Mäkelänkatu site is higher than the Kumpula site. Between 27th to 31st March, PM_{10} and $PM_{2.5}$ show relatively low pollution concentrations. These results are also confirmed by AQI which indicates overall low pollution levels for those dates. On the 1st April, all pollutant indexes fluctuate and show a slight increase and decrease. Then, we observe another fluctuation with a higher increase from 5th to 7th April. Again, we observe another rapid fluctuation between 9th to 10th April. Furthermore, by only considering the fluctuations in the air quality from 9th to 10th April (as zoomed in and shown on the right side of Figure 4), we observe a large discrepancy between the pollution levels **K** and **M** with a difference of 80 μg/m^3.

As a result, the fluctuations shown for the period of the time series plot, as well as the variations of the measurements in both sites **K** and **M**, call for the need for the deployment of air quality IoT devices separately at both sites in order to detect pollution hotspots and also monitor the air quality at fine-grained resolution in real-time. Indeed, deploying dense air quality sensors in cities could provide more accurate information leading to more robust and reliable conclusions about air quality levels at higher resolution, even at a few meter distances. A dense deployment can also assist in creating emission inventories of pollutants and detecting pollution sources, as well as allowing real-time exposure assessment for designing mitigation strategies [58].

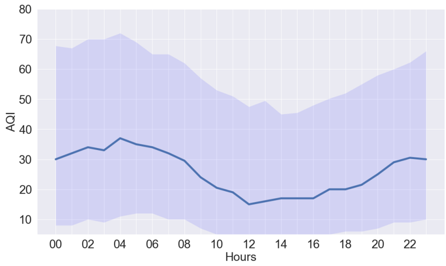
Traffic Management

Traffic is one of the main sources of outdoor air pollution in urban areas [59, 60]. The health effects of traffic-related air pollution continue to be of important public health risks [61]. In order to carry out effective traffic management driven by the level of air pollution, it is important to have air quality IoT devices installed next to roads. Therefore, the patterns of air pollution can be observed in roads allowing designing appropriate traffic management strategies.

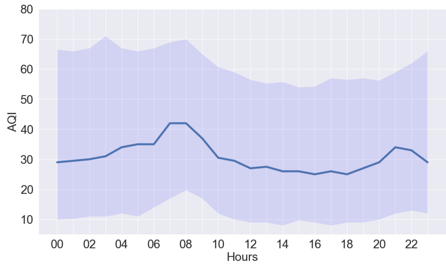
Figure 5 shows diurnal cycles of AQI, PM_{10} and $PM_{2.5}$ at the sites of Kumpula (right) and Makelankatu (left). The x-axes show the 24-h time period whereas the y-axes exhibit the levels of AQI and PM concentrations (in μg/m^3). The blue curves are the median of the data for each variable aggregated from one year of data whereas the shaded areas represent the lower quartile (25%) and upper quartile (75%) of the data for each variable aggregated from one year data (i.e., from 1st January to 31st December 2018).

As demonstrated in Figure 5, on the Kumpula site (the left subfigures), the AQI, PM_{10} and $PM_{2.5}$ do not increase during the peak hours (i.e, rush hours when people and vehicles movement is high). This is due to the fact that the Kumpula site is located in an urban background with less exposure to traffic emissions. However, on the Mäkelänkatu site (the right subfigures), the AQI, PM_{10} and $PM_{2.5}$ show an increase during peak hours, mainly between 8 AM and 10 AM. These patterns explain that Mäkelänkatu street is a busy road during the rush hours, especially in the mornings.

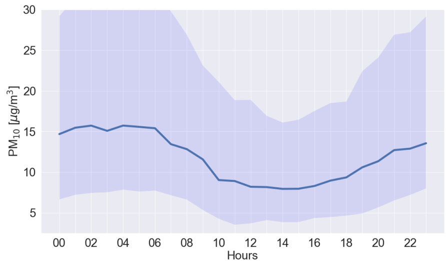
As a result, these patterns and the pollution concentration levels can be used by authorities to study for example the traffic behaviors and types of vehicles and therefore devise possible interventions to reduce the amount of pollutants in the areas where the IoT devices are installed. For instance, $PM_{2.5}$ (that are known as fine particles) are predominantly emitted from combustion sources like vehicles, diesel engines, and industrial facilities; and PM_{10} (that are known



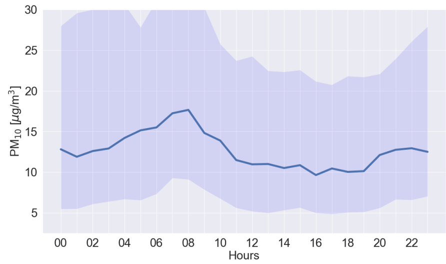
(a) AQI at Kumpula site



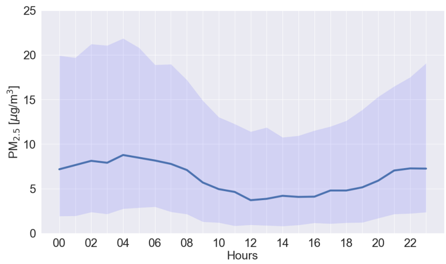
(b) AQI at Mäkelänkatu site



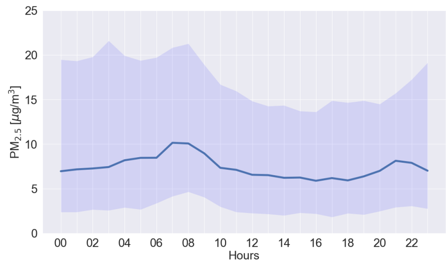
(c) PM_{10} at Kumpula site



(d) PM_{10} at Mäkelänkatu site



(e) $PM_{2.5}$ at Kumpula site



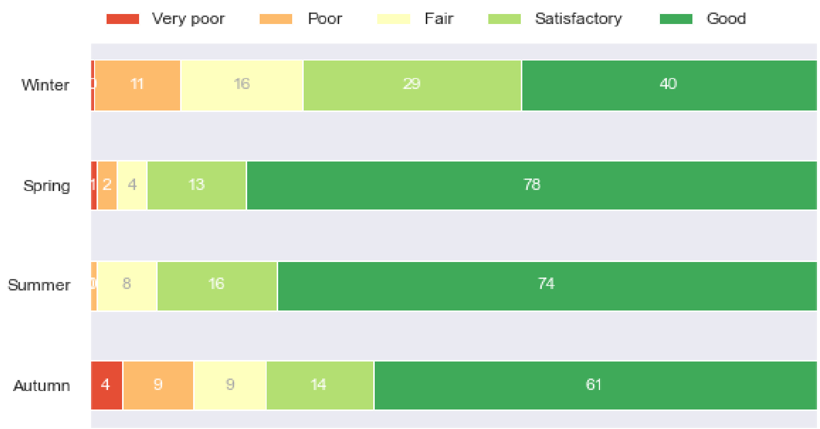
(f) $PM_{2.5}$ at Mäkelänkatu site.

Fig. 5: Diurnal cycles for AQI, PM_{10}, and $PM_{2.5}$ in Kumpula (left) and Mäkelänkatu (right) sites.

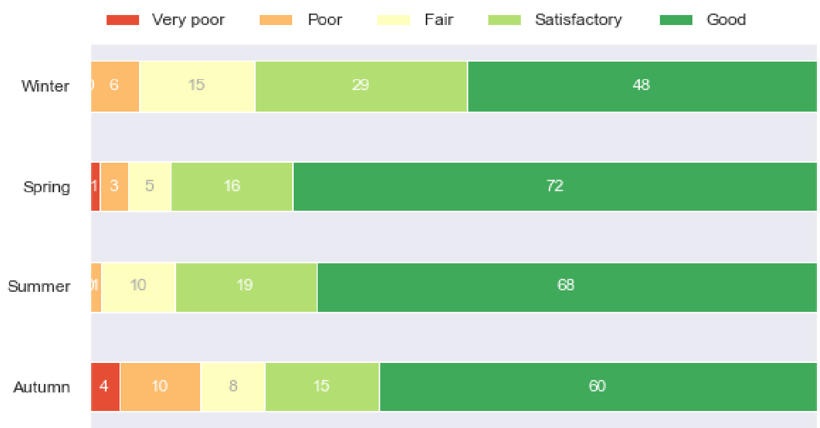
as coarse particles) are directly emitted from activities that disturb the soil including travel on roads, construction, mining, open burning or agricultural operations [62]. Hence, understanding the levels of PM_{10} and $PM_{2.5}$ concentrations at different locations enables planning appropriate interventions and designing effective traffic management strategies.

Urban Planning

Modern urban planning needs to consider environmental pollution and factors that threaten cities. Among many, AQI is known to be an important indicator that plays a vital role in urban life. Based on yearly AQI information, appropriate urban planning can be designed by considering the effects of different factors on air quality such as topography, buildings, roads, vegetation, and other external sources (e.g., traffic) [63]. Thus, poor AQI levels may indicate areas that are unsuitable for certain types of land use. For instance, sensitive land uses like schools, hospitals, and residential areas can be kept away from major pollution sources like factories or highways.



(a) AQI at Kumpula site



(b) AQI at Mäkelänkatu site

Fig. 6: Different AQI levels (%) in four different seasons at the two sites.

Figure 6 presents different percentages of AQI levels in four different seasons for the two sites. The figure shows the whole data aggregated for a year (1st January to 31st December 2018). The AQI is divided into four levels including Good (green), Satisfactory (light green), Fair (yellow), Poor (orange), and Very

Poor (red). For example, in the summer, the AQI levels in Kumpula (sub-figure 6a) are better than in Mäkelänkatu (sub-figure 6b). This is because the Kumpula site is surrounded by vegetation and trees during the summertime. In wintertime, on the other hand, the Kumpula site is slightly more polluted than Mäkelänkatu, as there is no vegetation and trees are without leaves, causing the Kumpula site to be exposed easily to air pollutants transported by nearby roads. The Kumpula area hosts residential buildings, university campuses, and a school, thus to mitigate the air pollution effects, in this area it is important for city planners to consider planting evergreen trees [64] such as Scots pine, Norway spruce, common juniper, and European yew.

In Mäkelänkatu site, on the other side, due to its proximity to the main road, the AQI levels are worse than Kumpula site. Therefore, better traffic management strategies can be devised for the Mäkelänkatu road. In general, air quality analysis based on AQI can provide information about prominent air pollution problems. Therefore, scientific assessments can be carried out in order to realize future development and planning for smart cities [65].

5 Role of AI and Emerging Technologies in Future Smart Cities

The convergence of AI and IoT—often defined as AIoT [66]—is not only expected but is already serving as a foundational element in the development of smart cities. With AI currently playing a key role in managing and interpreting the increasing volumes of data generated by a diverse array of IoT devices, it is evident that its significance will only amplify moving forward. As the data landscape continues to expand and AI methods undergo continuous refinement and innovation, there is growing potential for integrating newer, more efficient AI models and methodologies into key enabling technologies. Such integration can facilitate the creation of fully automated AI-enabled smart cities but and it also ensures that smart city ecosystems are equipped to adapt and respond to the ever-changing demands and challenges of evolving urban spaces.

Below, we outline a set of pivotal enabling technologies situated at the intersection of AI and IoT, each playing a crucial role in fostering the development of future smart cities. It is worth highlighting that the list presented is not exhaustive. Instead, it provides an illustrative snapshot of significant, emerging technological trends that are currently shaping the smart cities landscape. These identified technologies are presented as key drivers facilitating the emergence of cities that are not only smarter but also more efficient and responsive. Each technology contributes its unique strengths and capabilities, offering varied solutions. Together, they equip smart cities with functional modules necessary for addressing the myriad challenges these complex ecosystems currently face and will encounter in the future.

Digital Twin Systems

Deploying IoT and sensor networks in urban areas provides the opportunity for the creation of digital twin systems in smart cities. For example, deploying a massive number of surveillance cameras in cities can enable real-time monitoring of the people and traffic flow in cities and learning patterns from the movements and moving directions, allowing better planning for the traffic design. Similarly, using the telecom infrastructure and wireless access points deployed in cities makes it possible to estimate the number of access requests by the users (even for specific IoT applications), and therefore planning better resource management and thus improving the quality of experiences by the users. Moreover, as highlighted earlier in this chapter, deploying air pollution sensors allows for capturing air pollution in real-time and identifying hotspots in cities, leading to better planning for the cities. Using such massive deployments therefore enables the creation of digital twins, a powerful tool that provides the digital transformation of smart cities that enables real-time and remote monitoring of the physical elements (such as buildings and transportation systems) in cities, and therefore enables effective decision-making by the policy makers [67].

On-Device Machine Learning

On-device ML, also known as TinyML, is pivotal in advancing AIoT, offering substantial benefits in terms of efficiency, latency, and privacy [68]. TinyML enables devices to process and analyze data locally, reducing the need for constant connectivity and data transmission to centralized data centers, thereby decreasing latency and minimizing bandwidth usage. This approach makes AIoT applications more responsive and reliable, while also enhancing privacy and security by keeping sensitive data on the device. In the specific context of smart cities, there are several application scenarios where TinyML can play a transformative role. For example, it fosters the development of smart and autonomous entities capable of making decentralized and quick decisions in applications like traffic and pollution monitoring, thereby contributing to the collective intelligence in smart cities. Such deployment simplicity of TinyML, coupled with its independence from the power grid, facilitates the establishment of smart spaces even in remote and disadvantaged areas, promoting their economic and technological revitalization [69]. With the impending surge in urban populations, and the consequent strain on city resources and infrastructure, the introduction of TinyML in smart spaces is crucial for efficient resource optimization and energy waste reduction. This is imperative not just for managing the increasing energy demands but is integral to meeting stringent carbon neutrality goals set for sustainable urban living [70]. Furthermore, practical applications of TinyML, such as deploying LSTM Autoencoders on constrained devices for tasks like anomaly detection in urban noise sensor networks, showcase its potential and versatility in urban settings, paving the way for future explorations into on-device model training and trust management systems among sensor devices [71]. Each aspect of on-device ML represents a significant step toward decentralized, efficient, and intelligent urban planning and decision-making processes in smart cities.

6G Connectivity

Previous communication technologies including 4G, LTE-M, and NB-IoT as well as the current 5G technology have paved the way allowing large-scale deployment of IoT in smart cities. Indeed, the earlier 4G technology provided the communication resources that can support a large variety of IoT applications, LTE-M and NB-IoT technologies were planned to specifically support machine-to-machine and IoT deployments. Later, the 5G improved the communication capabilities of the previous technologies by providing URLLC, eMBB, and mMTC. Currently, 5G paves the way for using AI for 6G, the next-generation communication technology [72].

6G is expected not only to enhance communication capabilities (i.e., by URLLC+, eMBB+, and mMTC+) but it will offer AI and wireless sensing as new network services. In practice, 6G will see the physical objects through electromagnetic waves and will improve communication performance by providing high-resolution sensing, localization, imaging, and environment reconstruction capabilities. 6G will provide joint communication and sensing that will integrate localization, sensing and communication, and will facilitate edge intelligence and enable the transformation from connected things and people to connected intelligence [73].

The edge intelligence will also offer intelligence at the edge and will enable the processing of large datasets for critical IoT applications and computations. The edge intelligence will thus provide swift replies with precise decisions for the requested services by the specific IoT applications. Moreover, 6G is expected to provide high-density IoT connections and support one million connections per square kilometer [4]. Benefiting from these advanced features, therefore, 6G will support a wide variety of IoT applications at very large scales and with very high dense deployments. Examples of such applications would include but are not limited to activity detection, gesture recognition, mobility identification, remote sensing, simultaneous localization and mapping, object tracking, and security screening [74].

Blockchain

While AI-based technologies provide the intelligence required for insights generation and decision-making automation in smart cities, ensuring the security and integrity of the data utilized by AI algorithms is equally crucial. Blockchain emerges as a key enabler, safeguarding data collected and transmitted by AI-enabled Smart City systems, providing a secure, reliable, and trustworthy environment [75].

Furthermore, blockchain not only enhances the security and efficiency of IoT-enabled smart city applications but also mitigates data vulnerability and addresses single-node failures inherent in cloud-based solutions [76]. Though cloud-based architectures are widely used, they are susceptible to cyber-attacks, including data tampering and false data injection, and can experience reliability issues due to single-node failures [77]. In this context, blockchain, with its decentralized Distributed Ledger Technology (DLT), offers a robust and transformative alternative. It ensures transparency, data immutability, and integrity while providing pseudonymity. This technology is vital for smart cities, offering secure, resilient, and dynamic services across various sectors, including smart grids and Intelligent Transportation Systems (ITS) [76]. Blockchain facilitates trust-free, peer-to-peer transactions without central authorities and protects users' identities through public pseudonymous addresses [78].

The use of smart contracts, which is related to the *Data Sharing* case discussed in Section 3.4, automates transactions between parties, streamlining smart city operations seamlessly. The convergence of blockchain's security features, 6G connectivity, and AI intelligence is fundamental for the development of secure, resilient, and adaptable smart cities, ready to meet the evolving requirements of future applications [18].

6 Conclusion

Information and communication technologies are advancing rapidly, causing an increase in the deployed network infrastructure and fostering an increase in the variety and scale of IoT applications that support smart cities. This chapter addressed the requirements and challenges associated with large-scale IoT deployments in smart cities considering the advances in emerging communication and computing technologies. The chapter also highlighted the roles of AI, and 5G beyond networks as well as the computing technologies that are needed to enable massive-scale IoT deployments in cities. To showcase the benefits of IoT deployments in cities, the chapter also presented the results obtained from a real-world case study of deploying two air quality IoT devices in the city of Helsinki, deployed at two separate locations. The results explain how these IoT devices can benefit decision-making by providing local air pollution monitoring, traffic management, and urban planning. Finally, the chapter explains the role of AI and emerging technologies by addressing the advances toward blockchain, digital twin systems, on-device machine learning, and 6G connectivity that would play a fundamental role in the creation of future smart cities.

7 Acknowledgement

This research is supported by the Nokia Center for Advanced Research (NCAR) and by the Business Finland project with grant number 8782/31/2022. The research is also supported by the Academy of Finland grant numbers 335934, 345008, 339614, and 355330. The research is also supported by the Finland Centennial Foundation to Urban Air Quality 2.0 project.

Bibliography

[1] W. Melibari, H. Baodhah, and N. Akkari, "IoT-based smart cities beyond 2030: Enabling technologies, challenges, and solutions," in *2023 1st International Conference on Advanced Innovations in Smart Cities (ICAISC)*. IEEE, 2023, pp. 1–6.

[2] S. Zeadally, J. Guerrero, and J. Contreras, "A tutorial survey on vehicle-to-vehicle communications," *Telecommunication Systems*, vol. 73, pp. 469–489, 2020.

[3] D. Jiang and G. Liu, "An overview of 5G requirements," *5G Mobile Communications*, pp. 3–26, 2016.

[4] R. Gupta, D. Reebadiya, and S. Tanwar, "6G-enabled edge intelligence for ultra-reliable low latency applications: Vision and mission," *Computer Standards & Interfaces*, vol. 77, p. 103521, 2021.

[5] R. K. Tripathi, A. S. Jalal, and S. C. Agrawal, "Suspicious human activity recognition: a review," *Artificial Intelligence Review*, vol. 50, pp. 283–339, 2018.

[6] N. H. Motlagh, E. Lagerspetz, P. Nurmi, X. Li, S. Varjonen, J. Mineraud, M. Siekkinen, A. Rebeiro-Hargrave, T. Hussein, T. Petäjä, M. Kulmala, and S. Tarkoma, "Toward massive scale air quality monitoring," *IEEE Communications Magazine*, vol. 58, no. 2, pp. 54–59, 2020.

[7] X. Su, X. Liu, N. H. Motlagh, J. Cao, P. Su, P. Pellikka, Y. Liu, T. Petäjä, M. Kulmala, P. Hui, and S. Tarkoma, "Intelligent and scalable air quality monitoring with 5g edge," *IEEE Internet Computing*, vol. 25, no. 2, pp. 35–44, 2021.

[8] N. H. Motlagh, L. Lovén, J. Cao, X. Liu, P. Nurmi, S. Dustdar, S. Tarkoma, and X. Su, "Edge computing: The computing infrastructure for the smart megacities of the future," *Computer*, vol. 55, no. 12, pp. 54–64, 2022.

[9] B. Rashid and M. H. Rehmani, "Applications of wireless sensor networks for urban areas: A survey," *Journal of Network and Computer Applications*, vol. 60, pp. 192–219, 2016.

[10] S. Kim, S. Pakzad, D. Culler, J. Demmel, G. Fenves, S. Glaser, and M. Turon, "Health monitoring of civil infrastructures using wireless sensor networks," in *Proceedings of the 6th International Conference on Information Processing in Sensor Networks*, 2007, pp. 254–263.

[11] C. Sun, V. O. Li, J. C. Lam, and I. Leslie, "Optimal citizen-centric sensor placement for air quality monitoring: a case study of city of cambridge, the united kingdom," *IEEE Access*, vol. 7, pp. 47 390–47 400, 2019.

[12] C. Badii, P. Bellini, A. Difino, and P. Nesi, "Smart city IoT platform respecting gdpr privacy and security aspects," *IEEE Access*, vol. 8, pp. 23 601–23 623, 2020.

[13] L. Cui, G. Xie, Y. Qu, L. Gao, and Y. Yang, "Security and privacy in smart cities: Challenges and opportunities," *IEEE Access*, vol. 6, pp. 46 134–46 145, 2018.

[14] Y. Mehmood, F. Ahmad, I. Yaqoob, A. Adnane, M. Imran, and S. Guizani, "Internet-of-Things-based smart cities: Recent advances and challenges," *IEEE Communications Magazine*, vol. 55, no. 9, pp. 16–24, 2017.

[15] E. Lee, Y.-D. Seo, S.-R. Oh, and Y.-G. Kim, "A survey on standards for interoperability and security in the internet of things," *IEEE Communications Surveys & Tutorials*, vol. 23, no. 2, pp. 1020–1047, 2021.

[16] A. Hazra, M. Adhikari, T. Amgoth, and S. N. Srirama, "A comprehensive survey on interoperability for IoT: Taxonomy, standards, and future directions," *ACM Computing Surveys (CSUR)*, vol. 55, no. 1, pp. 1–35, 2021.

[17] R. Morabito and J. Jiménez, "IETF protocol suite for the internet of things: Overview and recent advancements," *IEEE Communications Standards Magazine*, vol. 4, no. 2, pp. 41–49, 2020.

[18] A. R. Javed, F. Shahzad, S. ur Rehman, Y. B. Zikria, I. Razzak, Z. Jalil, and G. Xu, "Future smart cities: Requirements, emerging technologies, applications, challenges, and future aspects," *Cities*, vol. 129, p. 103794, 2022.

[19] N. H. Motlagh, M. A. Zaidan, L. Lovén, P. L. Fung, T. Hänninen, R. Morabito, P. Nurmi, and S. Tarkoma, "Digital twins for smart spaces-beyond iot analytics," *IEEE Internet of Things Journal*, 2023.

[20] P. R. Singh, V. Singh, R. Yadav, and S. Chaurasia, "6G networks for artificial intelligence-enabled smart cities applications: A scoping review," *Telematics and Informatics Reports*, p. 100044, 2023.

[21] A. Gharaibeh, M. A. Salahuddin, S. J. Hussini, A. Khreishah, I. Khalil, M. Guizani, and A. Al-Fuqaha, "Smart cities: A survey on data management, security, and enabling technologies," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 4, pp. 2456–2501, 2017.

[22] J. Pan and J. McElhannon, "Future edge cloud and edge computing for Internet of Things applications," *IEEE Internet of Things Journal*, vol. 5, no. 1, pp. 439–449, 2017.

[23] N. Hassan, S. Gillani, E. Ahmed, I. Yaqoob, and M. Imran, "The role of edge computing in Internet of Things," *IEEE Communications Magazine*, vol. 56, no. 11, pp. 110–115, 2018.

[24] B. Diène, J. J. Rodrigues, O. Diallo, E. H. M. Ndoye, and V. V. Korotaev, "Data management techniques for Internet of Things," *Mechanical Systems and Signal Processing*, vol. 138, p. 106564, 2020.

[25] S. B. Atitallah, M. Driss, W. Boulila, and H. B. Ghézala, "Leveraging deep learning and IoT big data analytics to support the smart cities development: Review and future directions," *Computer Science Review*, vol. 38, p. 100303, 2020.

[26] W. M. da Silva, A. Alvaro, G. H. Tomas, R. A. Afonso, K. L. Dias, and V. C. Garcia, "Smart cities software architectures: a survey," in *Proceedings of the 28th Annual ACM Symposium on Applied Computing*, 2013, pp. 1722–1727.

[27] E. F. Z. Santana, A. P. Chaves, M. A. Gerosa, F. Kon, and D. S. Milojicic, "Software platforms for smart cities: Concepts, requirements, challenges, and a unified reference architecture," *ACM Computing Surveys (Csur)*, vol. 50, no. 6, pp. 1–37, 2017.

[28] "What is data sharing? - data sharing explained - aws," Amazon Web Services, 2022, accessed: 11/October/2023. [Online]. Available: https://aws.amazon.com/what-is/data-sharing/

[29] F. C. Delicato, P. F. Pires, T. Batista, E. Cavalcante, B. Costa, and T. Barros, "Towards an IoT ecosystem," in *Proceedings of the first international workshop on software engineering for systems-of-systems*, 2013, pp. 25–28.

[30] A. M. Shahat Osman and A. Elragal, "Smart cities and big data analytics: a data-driven decision-making use case," *Smart Cities*, vol. 4, no. 1, pp. 286–313, 2021.

[31] C. Badii, P. Bellini, D. Cenni, A. Difino, P. Nesi, and M. Paolucci, "Analysis and assessment of a knowledge based smart city architecture providing service apis," *Future Generation Computer Systems*, vol. 75, pp. 14–29, 2017.

[32] J. C. Jiang, B. Kantarci, S. Oktug, and T. Soyata, "Federated learning in smart city sensing: Challenges and opportunities," *Sensors*, vol. 20, no. 21, p. 6230, 2020.

[33] H. Y. Teh, A. W. Kempa-Liehr, and K. I.-K. Wang, "Sensor data quality: A systematic review," *Journal of Big Data*, vol. 7, no. 1, pp. 1–49, 2020.

[34] T. Chakraborty, A. U. Nambi, R. Chandra, R. Sharma, M. Swaminathan, and Z. Kapetanovic, "Sensor identification and fault detection in IoT systems," in *Proceedings of the 16th ACM Conference on Embedded Networked Sensor Systems*, 2018, pp. 375–376.

[35] H. Zhang, J. Liu, and A.-C. Pang, "A Bayesian network model for data losses and faults in medical body sensor networks," *Computer Networks*, vol. 143, pp. 166–175, 2018.

[36] J. Ye, G. Stevenson, and S. Dobson, "Detecting abnormal events on binary sensors in smart home environments," *Pervasive and Mobile Computing*, vol. 33, pp. 32–49, 2016.

[37] M. A. Zaidan, Y. Xie, N. H. Motlagh, B. Wang, W. Nie, P. Nurmi, S. Tarkoma, T. Petäjä, A. Ding, and M. Kulmala, "Dense air quality sensor networks: validation, analysis, and benefits," *IEEE Sensors Journal*, vol. 22, no. 23, pp. 23 507–23 520, 2022.

[38] R. Y. Wang and D. M. Strong, "Beyond accuracy: What data quality means to data consumers," *Journal of Management Information Systems*, vol. 12, no. 4, pp. 5–33, 1996.

[39] Y. Li and L. E. Parker, "Nearest neighbor imputation using spatial–temporal correlations in wireless sensor networks," *Information Fusion*, vol. 15, pp. 64–79, 2014.

[40] M. A. Zaidan, N. H. Motlagh, P. L. Fung, D. Lu, H. Timonen, J. Kuula, J. V. Niemi, S. Tarkoma, T. Petäjä, M. Kulmala, and T. Hussein, "Intelligent calibration and virtual sensing for integrated low-cost air quality sensors," *IEEE Sensors Journal*, vol. 20, no. 22, pp. 13 638–13 652, 2020.

[41] Y. Liu, T. Dillon, W. Yu, W. Rahayu, and F. Mostafa, "Missing value imputation for industrial IoT sensor data with large gaps," *IEEE Internet of Things Journal*, vol. 7, no. 8, pp. 6855–6867, 2020.

[42] C. C. Aggarwal, *An introduction to outlier analysis*. Springer, 2017.

[43] N. F. Ahmad, D. B. Hoang, and M. H. Phung, "Robust preprocessing for health care monitoring framework," in *2009 11th international conference on e-Health networking, applications and services (Healthcom)*. IEEE, 2009, pp. 169–174.

[44] H. H. Bosman, G. Iacca, A. Tejada, H. J. Wörtche, and A. Liotta, "Spatial anomaly detection in sensor networks using neighborhood information," *Information Fusion*, vol. 33, pp. 41–56, 2017.

[45] M. A. Rassam, M. A. Maarof, and A. Zainal, "Adaptive and online data anomaly detection for wireless sensor systems," *Knowledge-Based Systems*, vol. 60, pp. 44–57, 2014.

[46] E. W. Dereszynski and T. G. Dietterich, "Spatiotemporal models for data-anomaly detection in dynamic environmental monitoring campaigns," *ACM Transactions on Sensor Networks (TOSN)*, vol. 8, no. 1, pp. 1–36, 2011.

[47] C. Zhao and Y. Fu, "Statistical analysis based online sensor failure detection for continuous glucose monitoring in type I diabetes," *Chemometrics and Intelligent Laboratory Systems*, vol. 144, pp. 128–137, 2015.

[48] M.-F. Harkat, G. Mourot, and J. Ragot, "Sensor failure detection of air quality monitoring network," *IFAC Proceedings Volumes*, vol. 33, no. 11, pp. 529–534, 2000.

[49] P. Ferrer-Cid, J. M. Barcelo-Ordinas, J. Garcia-Vidal, A. Ripoll, and M. Viana, "A comparative study of calibration methods for low-cost ozone sensors in IoT platforms," *IEEE Internet of Things Journal*, vol. 6, no. 6, pp. 9563–9571, 2019.

[50] J. Rabatel, S. Bringay, and P. Poncelet, "Anomaly detection in monitoring sensor data for preventive maintenance," *Expert Systems with Applications*, vol. 38, no. 6, pp. 7003–7015, 2011.

[51] M. A. Zaidan, N. H. Motlagh, P. L. Fung, A. S. Khalaf, Y. Matsumi, A. Ding, S. Tarkoma, T. Petäjä, M. Kulmala, and T. Hussein, "Intelligent air pollution sensors calibration for extreme events and drifts monitoring," *IEEE Transactions on Industrial Informatics*, vol. 19, no. 2, pp. 1366–1379, 2023.

[52] H. H. Bosman, G. Iacca, A. Tejada, H. J. Wörtche, and A. Liotta, "Ensembles of incremental learners to detect anomalies in ad hoc sensor networks," *Ad Hoc Networks*, vol. 35, pp. 14–36, 2015.

[53] World Health Organization, "World health statistics 2019: Monitoring health for the SDGs, sustainable development goals," 2021.

[54] P. Kortoçi, N. H. Motlagh, M. A. Zaidan, P. L. Fung, S. Varjonen, A. Rebeiro-Hargrave, J. V. Niemi, P. Nurmi, T. Hussein, T. Petäjä, M. Kulmala, and S. Tarkoma, "Air pollution exposure monitoring using portable low-cost air quality sensors," *Smart Health*, vol. 23, p. 100241, 2022.

[55] L. Järvi, H. Hannuniemi, T. Hussein, H. Junninen, P. P. Aalto, R. Hillamo, T. Mäkelä, P. Keronen, E. Siivola, T. Vesala, and M. Kulmala, "The urban measurement station SMEAR III: Continuous monitoring of air pollution and surface-atmosphere interactions in helsinki, finland," *Boreal Environment Research*, vol. 14, pp. 86–109, 2009.

[56] R. Hietikko, H. Kuuluvainen, R. M. Harrison, H. Portin, H. Timonen, J. V. Niemi, and T. Rönkkö, "Diurnal variation of nanocluster aerosol concentrations and emission factors in a street canyon," *Atmospheric Environment*, vol. 189, pp. 98–106, 2018.

[57] P. L. Fung, S. Sillanpää, J. V. Niemi, A. Kousa, H. Timonen, M. A. Zaidan, E. Saukko, M. Kulmala, T. Petäjä, and T. Hussein, "Improving the current air quality index with new particulate indicators using a robust statistical approach," *Science of The Total Environment*, vol. 844, p. 157099, 2022.

[58] P. Kumar, L. Morawska, C. Martani, G. Biskos, M. Neophytou, S. Di Sabatino, M. Bell, L. Norford, and R. Britter, "The rise of low-cost sensing for managing air pollution in cities," *Environment international*, vol. 75, pp. 199–205, 2015.

[59] A. Y. Bigazzi and M. Rouleau, "Can traffic management strategies improve urban air quality? a review of the evidence," *Journal of Transport & Health*, vol. 7, pp. 111–124, 2017.

[60] N. H. Motlagh, M. A. Zaidan, P. L. Fung, E. Lagerspetz, K. Aula, S. Varjonen, M. Siekkinen, A. Rebeiro-Hargrave, T. Petäjä, M. Matsumi, Yutaka an Kulmala, T. Hussein, P. Nurmi, and S. Tarkoma, "Transit pollution exposure monitoring using low-cost wearable sensors," *Transportation Research Part D: Transport and Environment*, vol. 98, p. 102981, 2021.

[61] H. Boogaard, A. Patton, R. Atkinson, J. Brook, H. Chang, D. Crouse, J. Fussell, G. Hoek, B. Hoffmann, R. Kappeler *et al.*, "Long-term exposure to traffic-related air pollution and selected health outcomes: A systematic review and meta-analysis," *Environment International*, vol. 164, p. 107262, 2022.

[62] R. M. Harrison, J. Allan, D. Carruthers, M. R. Heal, A. C. Lewis, B. Marner, T. Murrells, and A. Williams, "Non-exhaust vehicle emissions of particulate matter and voc from road traffic: A review," *Atmospheric Environment*, vol. 262, p. 118592, 2021.

[63] C. Falzone and A.-C. Romain, "Establishing an air quality index based on proxy data for urban planning part 1: Methodological developments and preliminary tests," *Atmosphere*, vol. 13, no. 9, p. 1470, 2022.

[64] C. He, K. Qiu, A. Alahmad, and R. Pott, "Particulate matter capturing capacity of roadside evergreen vegetation during the winter season," *Urban Forestry & Urban Greening*, vol. 48, p. 126510, 2020.

[65] Y. Feng, M. Ning, Y. Lei, Y. Sun, W. Liu, and J. Wang, "Defending blue sky in china: Effectiveness of the "air pollution prevention and control action plan" on air quality improvements from 2013 to 2017," *Journal of Environmental Management*, vol. 252, p. 109603, 2019.

[66] J. Zhang and D. Tao, "Empowering things with intelligence: a survey of the progress, challenges, and opportunities in artificial intelligence of things," *IEEE Internet of Things Journal*, vol. 8, no. 10, pp. 7789–7817, 2020.

[67] T. Deng, K. Zhang, and Z.-J. M. Shen, "A systematic review of a digital twin city: A new pattern of urban governance toward smart cities," *Journal of Management Science and Engineering*, vol. 6, no. 2, pp. 125–134, 2021.

[68] L. Dutta and S. Bharali, "TinyML meets IoT: A comprehensive survey," *Internet of Things*, vol. 16, p. 100461, 2021.

[69] R. Sanchez-Iborra and A. F. Skarmeta, "TinyML-enabled frugal smart objects: Challenges and opportunities," *IEEE Circuits and Systems Magazine*, vol. 20, no. 3, pp. 4–18, 2020.

[70] V. Rajapakse, I. Karunanayake, and N. Ahmed, "Intelligence at the extreme edge: A survey on reformable TinyML," *ACM Computing Surveys*, 2022.

[71] S. S. Hammad, D. Iskandaryan, and S. Trilles, "An unsupervised TinyML approach applied to the detection of urban noise anomalies under the smart cities environment," *Internet of Things*, vol. 23, p. 100848, 2023.

[72] E. C. Strinati, S. Barbarossa, J. L. Gonzalez-Jimenez, D. Ktenas, N. Cassiau, L. Maret, and C. Dehos, "6g: The next frontier: From holographic messaging to artificial intelligence using subterahertz and visible light communication," *IEEE Vehicular Technology Magazine*, vol. 14, no. 3, pp. 42–50, 2019.

[73] B. Rong, "6g: The next horizon: From connected people and things to connected intelligence," *IEEE Wireless Communications*, vol. 28, no. 5, pp. 8–8, 2021.

[74] W. Jiang, Q. Zhou, J. He, M. A. Habibi, S. Melnyk, M. E. Absi, B. Han, M. Di Renzo, H. D. Schotten, F.-L. Luo *et al.*, "Terahertz communications and sensing for 6g and beyond: A comprehensive view," *arXiv preprint arXiv:2307.10321*, 2023.

[75] S. Li, "Application of blockchain technology in smart city infrastructure," in *2018 IEEE international Conference on Smart Internet of Things (SmartIoT)*. IEEE, 2018, pp. 276–2766.

[76] A. Kumari, R. Gupta, and S. Tanwar, "Amalgamation of blockchain and IoT for smart cities underlying 6G communication: A comprehensive review," *Computer Communications*, vol. 172, pp. 102–118, 2021.

[77] X. Wang, X. Zha, W. Ni, R. P. Liu, Y. J. Guo, X. Niu, and K. Zheng, "Survey on blockchain for internet of things," *Computer Communications*, vol. 136, pp. 10–29, 2019.

[78] M. A. Rahman, M. M. Rashid, M. S. Hossain, E. Hassanain, M. F. Alhamid, and M. Guizani, "Blockchain and IoT-based cognitive edge framework for sharing economy services in a smart city," *IEEE Access*, vol. 7, pp. 18 611–18 621, 2019.