YA-DÁ: YÁng-Based DÁta Model for Fine-Grained IIoT Air Quality Monitoring

Yagmur Yigit*, Khayal Huseynov†, Hamed Ahmadi‡, and Berk Canberk§

* Department of Computer Engineering, Istanbul Technical University, Turkey
† Department of Electronic Engineering, University of York, United Kingdom
‡ Department of Artificial Intelligence and Data Engineering, Istanbul Technical University, Turkey
§ School of Computing, Engineering and The Build Environment, Edinburgh Napier University, United Kingdom

Index Terms—Industrial Internet of Things (IIoT), Air Quality Monitoring, Yet Another Next Generation (YANG) Data Model, Digital Twins (DT).

Abstract—With the development of industrialization, air pollution is also steadily on the rise since both industrial and daily activities generate a massive amount of air pollution. Since decreasing air pollution is critical for citizens’ health and well-being, air pollution monitoring is becoming an essential topic. Industrial Internet of Things (IIoT) research focuses on this crucial area. Several attempts already exist for air pollution monitoring. However, none of them are improving the performance of IoT data collection at the desired level. Inspired by the genuine Yet Another Next Generation (YANG) data model, we propose a YÁng-based DÁta model (YA-DÁ) to improve the performance of IIoT data collection. Moreover, by taking advantage of digital twin (DT) technology, we propose a DT-enabled fine-grained IIoT air quality monitoring system using YÁ-DÁ. As a result, DT synchronization becomes fine-grained. In turn, we improve the performance of IIoT data collection resulting in lower round-trip time (RTT), higher DT synchronization, and lower DT latency.

I. INTRODUCTION

The fast urbanization of the human population and the increasing industrial activities have a consequential impact on global air quality. The rapid development of cities significantly increases the pollutants produced by the ever-increasing number of vehicles, traffic volume, industrial sites, etc. According to data from the World Health Organization (WHO), ninety-nine percent of the global population is exposed to air pollution at different levels [1]. Air quality significantly impacts daily life, and prolonged exposure to air pollution can injure the respiratory, immune, and nervous systems and even result in the development of cancer [2]. Therefore, air quality monitoring is becoming essential to protect and improve the life quality of citizens in overcoming these problems. Thanks to the Internet of Things (IoT) technological advancement, the number of IoT-connected devices is rapidly increasing. According to the forecasts, the number of IoT-connected devices worldwide will reach around 29.4 billion by 2030 [3]. One of the most significant applications of the IoT is the Industrial Internet of Things (IIoT). Gartner Report also envisions that IIoT devices will populate more than half of IoT devices soon. IIoT focuses on the connectivity of intelligent machines in different domains such as transportation, manufacturing, healthcare systems, etc. Therefore, it is an essential concept for air quality monitoring.

Combining the robust data model-driven automation needed in IIoT with device and network efficiency is crucial. IIoT devices have interfaces to manipulate connected equipment. To correctly manage resources on each device, all relevant data model interfaces must be known during development. Yet Another Next Generation (YANG) is a data modeling language for device configuration. It is rich enough to express the characteristics of the related data model. The YANG model is a popular topic in the networking industry as it is easy to learn and can describe complex programmatic interfaces [4]. It provides many different mechanisms for IoT devices [5]. For instance, it supports distributed authorship to use modules from various sources together; language extensions, and model extensions without centralized control; it provides lifecycle, conformance model, platform-specific model deviations for actual deployment, etc. Furthermore, interoperability is essential between IoT devices and the rest of the IP world. In next-generation network systems, all devices will connect, called the Internet of Everything; network elements and IoT devices will configure together. At this point, YANG can be used as the data modeling language to manage network elements, and services [6]. Another promising technology is the digital twin (DT), which has rapidly become an essential part of industrial applications. It bridges the gap between physical and digital spaces by creating a real-time virtual replica of the objects in the system. DT provides a high-performance simulation model for analysis, testing, configuration, and optimization. Recently, DT-based visualization has become an important issue, especially with the smart city concept. With the help of DT-based visualization, air pollution levels can be visualized in real time across the city.

The amount of data being processed is increasing rapidly in parallel with the advancement of technology. For this reason, only relevant data should be processed in the related system so that the system can use it more efficiently for specific applications such as air quality monitoring, anomaly detection, etc. [7]. Thus, the overload in the system is reduced. Data collection efficiency is essential for IIoT systems since it can seriously affect the freshness of the data captured and the
decision-making process. Besides, efficient IIoT data collection reduces energy consumption, latency, network lifetime, and overall cost. However, existing air pollution monitoring works have high Round-trip time (RTT), relatively low DT synchronization, and relatively high DT delay. Therefore, IoT-based air quality monitoring was inefficient. In this paper, we are improving the performance of IoT data collection by proposing a new data model called the YAng-based DAData model (YA-DA), based on the YANG model. By taking advantage of DT technology, we propose a DT-enabled fine-grained IIoT air quality monitoring system using YA-DA. After implementing the YA-DA data structure, our approach has lower RTT, higher DT synchronization, and lower DT latency. We will provide more detail in the following sections.

The rest of the paper is organized as follows. Section II provides a survey of related studies in the literature. The proposed solution and the performance evaluation are explained in Section III and Section IV respectively. We conclude the paper in Section V.

II. RELATED WORK

Since data collection is an essential part of sensor devices and related technologies such as IIoTs and wireless sensor networks (WSN), there are many works published in this research area to propose different models, architectures, and contributions in the domains. I. Ali et al. gave a comprehensive review of the data collection in the IoT, WSN, and sensor cloud research domains [8]. The findings show that studies on data collection are relatively consistent with stable output in the last five years. They also provided critical research challenges and future research directions for data collection in related research domains. Z. Sheng et al. proposed a lightweight approach to enable device management of wireless sensor devices by taking advantage of the recent development of IPv6-based open standards for accessing wireless resource-constrained networks [9]. They developed a prototype to test the performance of the proposed approach in managing wireless sensor devices, thus contributing to IoT data collection development. W. Osamy et al. suggested an intelligent data collection technique to determine energy-aware disjoint dominating sets that work as data collection nodes in each round to improve overall WSN lifetime [10]. They mainly focused on energy conservation to enable the efficient functioning and lifetime of WSNs. They provided the efficiency of the proposed technique mathematically and in simulation. N. Koroniatis et al. presented a smart airport cyber twins security-oriented IIoT testbed SAir-IIoT [11]. It comprises multiple heterogeneous IIoT devices and communication protocols that can be accessed remotely as a service, automatically interconnected with each other. They proposed a data management tool that dynamically collects, analyzes, and labels network-acquired and telemetry IIoT data. In another work, S. Chen et al. proposed a privacy-preserving data collection and computation offloading scheme for efficient and secure big sensory data collection in fog-assisted IoT [12]. The simulation results show that the proposed method is an efficient data collection and computation offloading scheme with a strong privacy preservation property. C. T. Cheng et al. designed the concurrent data collection trees for IoT applications [13]. Their results show that the proposed tree structures perform better in data collection processes.

Some works focus on air quality monitoring taking advantage of DT technology in the literature. G. Mylonas et al. gave a detailed review of current DT research in the field of intelligent cities and also drew parallels with the applications of Industry 4.0 [14]. They emphasized the importance of pollution monitoring in the smart city domain and reviewed the current works of pollution monitoring. They also presented the open challenges of city-scale DTs. G. Schrotter et al. presented a city-scale DT of Zurich to facilitate use in several applications like air and noise pollution monitoring [15]. They presented a study that promotes the active digital participation of citizens in urban planning procedures by visualizing a variety of data covering various applications, such as thermal monitoring of the city. They note the significance of using open spatial data in encouraging dissemination and developing new applications. H. Lehner et al. discussed a new strategy for producing the three-dimensional city model and the geographical data of Vienna, which completely rethinks geographical data in the geographical data workflows [16]. After discussing the effect of the level of detail of three-
dimensional city representation, they provided the simulation results that calculate the pollution dispersion inside a city. In another work, T. Nochta et al. provided a city-scale DT prototype presented to tackle congestion, air pollution, growth management, and the limited capacity of the local energy infrastructure issues in the Cambridge city region [17]. They underlined the importance of city-scale DT to reflect the specifics of the urban and socio-political context. Y. Liu et al. proposed a federated learning-based aerial-ground air quality sensing framework for fine-grained three-dimensional air quality monitoring and forecasting [18]. They also proposed a graph convolutional neural network-based extended short-term memory model for ground sensing systems. Overall, most current air pollution monitoring DT works focus on the smart city domain, particularly the city-scale DT.

None of the previous studies specifically covered IIoT-based air quality monitoring in terms of data collection. So, their performance was relatively low. Thanks to the proposed approach, the performance in the particular DT-based air pollution monitoring model will be increased. The improved performance metrics with our solution are DT synchronization, DT latency, and RTT.

III. PROPOSED SOLUTION

Our proposed system consists of a physical network with physical objects and a digital twin network with digital twins and brain components. The proposed architecture is shown in Fig. 1. A replica of physical objects and synchronization between physical objects and digital twins are built. The IoT gateway collects sensor data and provides cyber-physical interaction between the physical objects and the digital twin. Then, sensor data is taken by the digital twin. After that, the air quality monitoring system takes only the desired data thanks to the software-defined YA-DA data model.

In the proposed system, we have spatial and temporal complexity. Spatial complexity is related to memory usage in the system. It needs to minimize by pulling only the relevant data into the air quality monitoring system. Temporal complexity pertains to end-to-end delay. It should be reduced to enhance the quality of service of the system. The software-defined YA-DA data model provides fine-grained air quality monitoring and reduces the computational time in the system by preventing the system from overloading. Thus, the system’s operation becomes more efficient, and we reduce the spatial and temporal complexity.

A. Software-defined YANG-based Data Model (YA-DA)

YANG is a data modeling language for configuration, and monitoring [21]. It is originally designed to model data for the Network Configuration (NETCONF) protocol. NETCONF is a network management protocol. It provides basic programming features for convenient and robust automation of network services. After a YANG module has defined data hierarchies, it can use for NETCONF-based operations. These can include state data, configuration, notifications, and remote procedure calls. All data sent between a NETCONF client and a server can be described. NETCONF is just one of many protocols that can be used with YANG. For instance, RESTCONF and the Constrained Application Protocol (CoAP) Management Interface (CoMI), etc. NETCONF is more CPU and memory efficient than RESTful and gRPC services [22].

The hierarchical organization of data is modeled as a tree in which each node has a name-value pair or a set of child nodes in YANG. It provides clear and brief descriptions of nodes along with the interaction between these nodes. A YANG module includes a combination of related definitions. It can import definitions from external modules and contain explanations from submodules. Moreover, the hierarchy can extend by allowing a module to add data nodes to the order defined in another module.

By taking the above advantages of the YANG model, we designed the YA-DA data model to improve the performance of the IoT data collection. We developed the YA-DA model in tree data structure because tree data structure performs better in IoT data collection. We defined four main data node types in the YA-DA data model as in the YANG for data modeling:

| TABLE I | SEMANTIC MODEL OF AIR MONITORING SENSORS BASED ON THE YA-DA DATA STRUCTURE |
|---------------------------------|---------------------------------------------------------------------------------------------------|
| container AirParticleURI {      | description “Air Monitoring Particle Sensor”                                                      |
| list value {                    | key “pm2.5-data”;                                                                             |
| leaf pm1-data {type air-sensor;}| leaf pm2.5-data {type air-sensor;}                                                              |
| leaf pm10-data {type air-sensor;}|                                               |
| container AirTemperatureURI {   | description “Air Monitoring Temperature Sensor”                                                 |
| leaf value {type air-sensor;}    |                                                   |
| container AirHumidityURI {      | description “Air Monitoring Humidity Sensor”                                                   |
| leaf value {type air-sensor;}    |                                                   |
| container AirGasesURI {         | description “Detectable Air Quality Gases”                                                     |
| container value {               | leaf carbon-monoxide-data {type air-sensor;}                                                   |
| leaf nitric-oxide {type air-sensor;}| leaf nitrogen-dioxide {type air-sensor;}                                                       |
| leaf sulphur-dioxide {type air-sensor;}| leaf carbon-monoxide-data {}                                                                 |
| leaf ethanol {type air-sensor;}  | leaf hydrogen {type air-sensor;}                                                               |
| leaf ammonia {type air-sensor;}  | leaf ethylamine {type air-sensor;}                                                             |
| leaf ozone {type air-sensor;}    |                                                   |
| Dataset        | Number of Features | Number of Samples | Number of Used Samples |
|---------------|--------------------|-------------------|------------------------|
| Fridge Activity | 6                  | 59944             | 1000                   |
| Garage Door Activity | 6              | 59587             | 800                    |
| GPS Tracker Activity | 6              | 58960             | 2200                   |
| Modbus Activity | 7                  | 51106             | 2000                   |
| Motion Light Activity | 6              | 59488             | 1000                   |
| Thermostat Activity | 6              | 52774             | 1000                   |
| AQ&U [20]    | 11                 | 522000            | 2000                   |

leaf nodes, leaf-list nodes, container nodes, and list nodes.

- **Leaf Node** contains at most one instance in the data tree. A leaf has a value but no child nodes.
- **Leaf-List Node** defines a set of uniquely identifiable nodes rather than a single node. Each node has a value but no child nodes.
- **Container Node** is used to group related nodes in a subtree. A container has no value but rather a set of child nodes of any type, such as leaves, lists, containers, leaf lists, etc.
- **List Node** defines a sequence of list entries. Each entry is defined like a container and is uniquely identified by the values of the key leaves. A list can define multiple key leaves and may contain any number of child nodes of any type, including leaves, lists, containers, etc.

The semantic model of air quality monitoring sensors based on YA-DA data structure is shown in table II.

Since the amount of sensor data in the system is more than desired and needed, the air quality monitoring system can become very slow. It is necessary to import only the required data into the system. Therefore, we also use a software-defined YA-DA model to accomplish this. In this work, we defined an air quality KPI. Thanks to the YA-DA Paths, the system receives only the relevant data from the digital twin. Thus, we get an increasingly efficient, fine-grained system. The hierarchy between the mentioned terms can be better represented in Fig. 2.

The above explanation is not a full DTDL definition of a sensor entity but contains main parts. As a result, we have connected the twins into a graph that reflects their interactions. For data analysis and storage, the ADT platform is able to stream the data through an external output plugin [24]. After implementing the twin models with predefined interfaces and relations among entities, we transmitted data to the brain part of the proposed system as the output plugin of ADT using the DT API.

We tested the performance of the software-defined YA-DA data model. To this end, we used TON-IoT and AQ&U as two real-world datasets [19] [20]. The ToN-IoT dataset is comprised of heterogeneous data sources to evaluate the fidelity and efficiency of different applications for IoT and IIoT. The AQ&U dataset is created to monitor air quality in Salt Lake City. The properties of these datasets are as in table II. We constituted a new dataset of ten thousand samples by taking particular samples from each dataset.

We reconfigured the sensors of the YA-DA data model according to the created dataset. Since we used an available dataset, not real-time data, we tested the proposed approach in the digital twin and brain areas of the digital twin network section of the proposed system. We set the DT synchronization to be between 0 and 1. The value of 1 means full synchronization, and 0 implies out-of-synchronization. We compared the proposed solution’s performance implementation with and without the YA-DA data model. The results showed that all the required sensors to monitor air quality are classified successfully. As can be seen in table III, the system has better
DT synchronization when using the YA-DA data model. Thus, DT synchronization becomes fine-grained. Then, we tested the average end-to-end delay between the digital twin and brain areas of the digital twin network section of the proposed system. The results are shown in Fig. 3. Since the YA-DA data model has a tree structure, its average end-to-end delay is better. Hence, we can provide that the DT network has lower latency using YA-DA data model.

After that, we examined the performance of the digital twin network. Fig. 4 depicts the RTT result of the network. The results showed that the data payload became around forty percent better when we used the software-defined YA-DA data model based on YANG. The reason for this result is that we pulled data into the system using the YA-DA data model instead of all raw data. Moreover, the network load is reduced by getting specific data using YA-DA.

Finally, we scrutinized the air quality data in the monitoring system of the brain part. We used Grafana to visualize and analyze our air quality data. Grafana is a web-based, open-source analytics and interactive visualization program. Fig. 5 demonstrates the graphical representation of the air quality data. This data includes particulate matter, air gases, humidity, and temperature values. According to the values, it can be stated that the air quality in this region is at an acceptable level.

| Number of Nodes | DT Synchronization (∗) with YA-DA | without YA-DA |
|-----------------|-----------------------------------|----------------|
| 4               | 0.65                              | 0.47           |
| 6               | 0.61                              | 0.42           |
| 16              | 0.74                              | 0.57           |

(∗) The interval for the value is [0,1], in which closeness to 0 indicates poor and closeness to 1 represents a desirable outcome.

Fig. 3. The comparison of average end-to-end delay.

Fig. 4. The comparison of RTT results.

Fig. 5. The monitoring of air quality data in the system.
V. CONCLUSION

In this paper, we design a YANG-based DATA model (YA-DA) to improve the performance of IIoT data collection inspired by the authentic YANG data model. DT synchronization becomes fine-grained thanks to the YA-DA data structure. Here, we propose a DT-enabled fine-grained IIoT air quality monitoring system by taking advantage of DT and IIoT technologies. Thus, we increase the efficiency of the air quality monitoring system. Our simulation results showed that our solution successfully reduces RTT and DT latency and improves DT synchronization thanks to the software-defined YA-DA data model.

ACKNOWLEDGMENT

Yagmur Yigit would like to thank the DeepMind Scholarship Programme and the ITU-Turkcell Graduate Research Scholarship Programme for their support. This paper is supported by the Royal Society under Grant IES/R3/213169.

REFERENCES

[1] W. H. Organization. The Global Health Observatory: Air Pollution. [Online]. Available: https://www.who.int/data/gho/data/themes/air-pollution?tab=countryprofile [Accessed July 8, 2022].
[2] M. A. Fekih, W. Bechkit, H. Rivano, M. Dahan, F. Renard, L. Alonso, and F. Pineau, “Participatory Air Quality and Urban Heat Islands Monitoring System,” IEEE Transactions on Instrumentation and Measurement, vol. 70, pp. 1–14, 2021.
[3] M. Hung. Leading the IoT: Gartner Insights on How to Lead in a Connected World. [Online]. Available: https://www.gartner.com/imagesrv/books/iot [Accessed July 2, 2022].
[4] M. Bjorklund. RFC 6020: YANG - A Data Modeling Language for the Network Configuration Protocol (NETCONF). [Online]. Available: https://ietf.org/rfc/rfc6020.txt, Accessed July 10, 2022.
[5] P. V. D. Stok and A. Bierman. “YANG-Based Constrained Management Interface (CoMI),” in IoT Semantic Interoperability Workshop, Ericson Research, 2016.
[6] D. Bogdanović, B. Claise, and C. Moberg. RFC 8199: YANG Module Classification. [Online]. Available: https://www.rfc-editor.org/rfc/rfc8199 [Accessed July 10, 2022].
[7] E. Horsanali, Y. Yigit, G. Secinti, A. Karameseoglu, and B. Canberk, “Network-Aware AutoML Framework for Software-Defined Sensor Networks,” in 2021 17th International Conference on Distributed Computing in Sensor Systems (DCOSS), 2021, pp. 451–457.
[8] I. Ali, I. Ahmedy, A. Gani, M. U. Munir, and M. H. Anisi, “Data Collection in Studies on Internet of Things (IoT), Wireless Sensor Networks (WSNs), and Sensor Cloud (SC): Similarities and Differences,” IEEE Access, vol. 10, pp. 33,909–33,931, 2022.
[9] Z. Sheng, H. Wang, C. Yin, X. Hu, S. Yang, and V. C. M. Leung, “Lightweight Management of Resource-Constrained Sensor Devices in Internet of Things,” IEEE Internet of Things Journal, vol. 2, no. 5, pp. 402–411, 2015.
[10] W. Osamy, A. Salim, A. M. Khedr, and A. A. El-Sawy, “IDCT: Intelligent Data Collection Technique for IoT-Enabled Heterogeneous Wireless Sensor Networks in Smart Environments,” IEEE Sensors Journal, vol. 21, no. 18, pp. 21,099–21,112, 2021.
[11] N. Korontziotis, N. Moustafa, F. Schiliro, P. Gauravaram, and H. Janicke, “The SAIrIoT Cyber Testbed as a Service: A Novel Cybertwins Architecture in IoT-Based Smart Airports,” IEEE Transactions on Intelligent Transportation Systems, pp. 1–14, 2021.
[12] S. Chen, X. Zha, H. Zhang, C. Zhao, G. Yang, and K. Wang, “Efficient Privacy Preserving Data Collection and Computation Offloading for Fog-Assisted IoT,” IEEE Transactions on Sustainable Computing, vol. 5, no. 4, pp. 526–540, 2020.
[13] C.-T. Cheng, N. Ganganath, and K.-Y. Fok, “Concurrent Data Collection Trees for IoT Applications,” IEEE Transactions on Industrial Informatics, vol. 13, no. 2, pp. 793–799, 2017.
[14] G. Mylonas, A. Kalogerias, G. Kalogerias, C. Anagnostopoulos, C. Alexakos, and L. Muñoz, “Digital Twins From Smart Manufacturing to Smart Cities: A Survey,” IEEE Access, vol. 9, pp. 143,222–143,249, 2021.
[15] G. Schrotter and C. Hürzeler, “The Digital Twin of the City of Zurich for Urban Planning,” PFG – Journal of Photogrammetry, Remote Sensing and Geoinformation Science, vol. 88, pp. 99–112, 2020.
[16] H. Lehner and L. Dorflner, “Digital geoTwin Vienna: Towards a Digital Twin City as Geodata Hub,” PFG – Journal of Photogrammetry, Remote Sensing and Geoinformation Science, vol. 88, pp. 63–75, 2020.
[17] T. Nochta, L. Wan, J. M. Schooling, and A. K. Parlikad, “A Socio-Technical Perspective on Urban Analytics: The Case of City-Scale Digital Twins,” Journal of Urban Technology, vol. 28, no. 1-2, pp. 263–287, 2021. [Online]. Available: https://doi.org/10.1080/10630732.2020.1798177.
[18] Y. Liu, J. Nie, X. Li, S. H. Ahmed, W. Y. B. Lim, and C. Miao, “Federated Learning in the Sky: Aerial-Ground Air Quality Sensing Framework With UAV Swarms,” IEEE Internet of Things Journal, vol. 8, no. 12, pp. 9827–9837, 2021.
[19] N. Moustafa. ToN IoT datasets. [Online]. Available: https://ieee-dataport.org/documents/toniot-datasets [Accessed July 10, 2022].
[20] T. Becnel, K. Kelly, and P.-E. Gaillardon. University of Utah AirU Pollution Monitoring Network, Salt Lake City UT, 2019-07-26 / 2021-05-14. [Online]. Available: https://ieee-dataport.org/documents/university-utah-airu-pollution-monitoring-network-salt-lake-city-ut-2019-07-26-2021 [Accessed July 10, 2022].
[21] M. Bjorklund. RFC 7950: The YANG 1.1 Data Modeling Language. [Online]. Available: https://www.rfc-editor.org/rfc/rfc7950.html [Accessed July 10, 2022].
[22] M. Erel-Ozc ¸evik and B. Canberk, “OFaaS: OpenFlow Switch as a Service for Multi Tenant Slicing in SD-CDN,” IEEE Transactions on Network and Service Management, vol. 18, no. 1, pp. 362–373, 2021.
[23] Microsoft. Azure Digital Twins Documentation. [Online]. Available: https://docs.microsoft.com/en-us/azure/digital-twins [Accessed Sept. 27, 2022].
[24] Y. Yigit, B. Bal, A. Karameseoglu, T. Q. Duong, and B. Canberk, “Digital Twin-Enabled Intelligent DDoS Detection Mechanism for Autonomous Core Networks,” IEEE Communications Standards Magazine, vol. 6, no. 3, 2022.
[25] G. Labs. Grafana: The open observability program. [Online]. Available: https://grafana.com/products/cloud/ [Accessed Sept. 28, 2022].