Khatri, Vikramajeet; Monshizadeh, Mehrnoosh; Hojjatinia, Sina; Kriaa, Siwar; Mahonen, Petri

Automated Data Correlation for IoT Anomaly Detection with B5G Networks

Published in:
2024 32nd International Conference on Software, Telecommunications and Computer Networks, SoftCOM 2024

DOI:
10.23919/SoftCOM62040.2024.10721776

Published: 01/01/2024

Document Version
Peer-reviewed accepted author manuscript, also known as Final accepted manuscript or Post-print

Please cite the original version:
Khatri, V., Monshizadeh, M., Hojjatinia, S., Kriaa, S., & Mahonen, P. (2024). Automated Data Correlation for IoT Anomaly Detection with B5G Networks. In D. Begusic, J. Radic, & M. Saric (Eds.), *2024 32nd International Conference on Software, Telecommunications and Computer Networks, SoftCOM 2024* (SoftCOM). IEEE. https://doi.org/10.23919/SoftCOM62040.2024.10721776

Automated Data Correlation for IoT Anomaly Detection with B5G Networks

Vikramajeet Khatri
Nokia Bell Labs, Finland
vikramajeet.khatri@nokia-bell-labs.com

Mehrnoosh Monshizadeh
Nokia Bell Labs, France
Department of Information and Communication Engineering, Aalto University, Finland
mehrnoosh.monshizadeh@nokia-bell-labs.com
mehrnoosh.monshizadeh@aalto.fi

Sina Hojjatinia
Nokia, France
Department of Information and Communication Engineering, Aalto University, Finland
sina.hojjatinia@nokia.com
sina.hojjatinia@aalto.fi

Siwar Kriaa
Nokia Bell Labs, France
siwar.kriaa@nokia-bell-labs.com

Petri Mähönen
Department of Information and Communication Engineering, Aalto University, Finland
petri.mahonen@aalto.fi

***Abstract*— Smart city monitoring technologies, including IoT devices like sensors and smart cameras, enable real-time anomaly detection by analyzing data from various locations. While video and audio can identify unsafe activities, camera coverage is limited, necessitating audio detectors for out-of-sight incidents. Static methods do not perform well under conditions like low-quality voice due to illness or mood, highlighting the need for a dynamic mechanism to orchestrate data collection, clean background noise, correlate data, and identify public safety incidents. This paper addresses challenges in correlating data from IoT devices at different locations, orchestrating information among various IoT service providers, and ensuring communication between IoT and network domains. The proposed architecture leverages AI to analyze IoT data in real-time for automatic anomaly detection, making it well-suited for AI-enabled Beyond 5G (B5G) networks. Analysis results are sent to operators via orchestrators to pinpoint the location of anomalous IoT devices. This information is also relayed to public safety agencies for appropriate action. Unlike existing systems focused on audio and video data, the proposed architecture can be applied to any IoT data, enhancing monitoring and detection capabilities.**

Keywords—machine learning, IoT, anomaly detection, B5G

I. Introduction and Motivation

The concept of smart cities envisions smart parking meters, smart electric meters, smart streetlights, and smart transportation systems, as well as public safety. For these applications, data mining techniques have been introduced as efficient tools to detect both unknown and known patterns and to discover the relationships in the large quantity of data collected from sensors and devices for further analysis. In this context, Machine Learning (ML) represents a potential means of identifying abnormal behavior in Internet of Things (IoT) devices within smart cities.

In Artificial Intelligence (AI) enabled 6th Generation (6G) networks, it is relevant to note that the entire network solution space is moving towards the integrations of AI/ML. These techniques are becoming significantly important within 3rd Generation Partnership Project (3GPP). While vendor-specific AI/ML solutions can be implemented today, 3GPP is enhancing support for data collection and exposure of analytics. In addition, 5th Generation Advanced (5G-Advanced) and forthcoming 6G networks are expected to offer AI/ML capabilities across the Radio Access Network (RAN), core, and management network domains.

Current 5G core network domain includes the LoCation Services (LCS) architecture which can be used to collect location data [1] [2] [3] [4].

In 6G networks, a client function (similar to LCS) can collect the location data from core network using similar functions to Gateway Mobile Location Centre (GMLC) or Network Exposure Function (NEF). The client will invoke Mobile Terminated Location Requests (MT-LR) request for the location of a target User Equipment (UE) which is a moving object. There are different request types:

- Immediate Location Request is used to get location information for the UE (moving object) or group of target UEs within a short time period. This is usually employed by emergency services and can act in real-time fashion.
- Deferred Location Request is used to get location information for the target UE (moving object) or group of target UEs at some future time (or times), which may be associated with specific events associated with the target UE (moving object) or group of target UEs. The specific events related to this document are following:
    - Area. UE (moving object) enters, leaves, or remains within a pre-defined geographical area.
    - Periodic Location. Location report is generated periodically.
    - Motion. An event where the UE (moving object) moves by more than some predefined straight-line distance from a previous location.

- Core domain can also provide analytics data (via similar function as NetWork Data Analytics Function (NWDAF) in 5G), collects data from Network Functions (NFs) and Operations, Administration and Maintenance (OAM). Following analytics data is related to current paper:
    - NF load analytics (NF status, load, and resource usage)
    - Network performance analytics (radio status information, number of UEs (moving object) in the area, Protocol Data Unit (PDU) session setup success rate in the area and handover success ratio in the area)
    - UE (moving object) mobility analytics for UE (moving object) location
    - UE (moving object) communication analytics, UE (moving object) communication patterns and type of communication
    - Abnormal UE (moving object) behavior parameters, UE (moving object) with unexpected location or communication pattern
    - User data congestion analytics (areas with congestion)
    - Quality of Service (QoS) sustainability analytics (areas where QoS Key Performance Index (KPI) have exceeded expected value thresholds)

Advanced monitoring technologies used in smart cities including IoT devices (e.g., sensors and smart cameras) can be installed at various locations, where the real-time data can be analyzed to detect anomalies. An example, video and audio data can be utilized to identify and track individual unsafe activities. However, cameras are useful only for limited areas that within their visible range; for example, in cases where activity occurs out of the camera's coverage, only the audio detectors that are deployed throughout the area would be able to provide assistance. In addition, for the conditions such as low-quality voice due to illness, mood and other factors, the current static methods tend to underperform. Hence, an intelligent and dynamic mechanism is needed to orchestrate data collection between devices installed in various location and the network, clean background noise from the collected data, correlate the collected data and identify the type of public safety incident and furthermore to inform the corresponding location of a target or situation. On the other hand, cameras can provide the location of the event, which can be then fed into an analytic function to make predictions. This data can be correlated with the prediction mobility or depending on the situation then a specific mobility can be asked (e.g., a criminal may not use the most probable path). This process is very challenging or impossible without the help of ML and AI.

This paper proposes a novel, scalable and distributed software architecture that enables deployment of AI/ML enabled orchestrator for situation awareness from IoT data. The proposes architecture allows event triggered alerts for emergencies or anomalies and can be integrated as secure application module for existing 5G and coming 6G networks.

Consequently, this paper addresses the following challenges:

- Correlating various types of data collected from IoT devices that are located in different location.
- Orchestrating collected information among various IoT service providers in order to track object with or without SIM card.
- Communication between IoT domain and network domain.

In the paper we illustrate the framework with a couple of different scenarios, but we emphasize that the proposed architecture is general even and data aggregation-based framework that allows location and time correlation of any aggregated IoT data in multi-operator environment. This has a capability of increasing the system resilience and efficiency.

The proposed software architecture in this paper is *not* limited to an audio/video context for data correlation, but the scope is any data and information that can be collected from IoT devices in general, and that can be used toward a certain monitoring/detection goal.

This architecture employs AI to analyze the received IoT data in real-time and automatically detect an anomaly. The results of the analysis are transmitted to the respective operator via orchestrators in order to obtain the location of the IoT device carrying the anomalous behavior. Furthermore, the extracted information will be sent to public safety agencies or the like, for taking proper action. In the proposed architecture, the details of anomaly detection and orchestrating the result to operator are presented in order to provide a security alert.

In the case of IoT cameras, the mentioned technology can be integrated with response mechanisms such as Automated Targeting System (ATS) for situational awareness and to support enhanced mission-critical operations. The proposed architecture can also support search and rescue missions in difficult environments, such as tunnels or collapsed and damaged structures.

The main contributions of out solution are as follows:

- Introduction of the Intelligent & Autonomous Security Service (IASS) framework to enable the communication between IASS platform and 5G/6G networks in a non-roaming scenario (involving NEF, User Data Registry (UDR), GMLC and NWDAF).
- Enabling the communication between IASS platform and 5G/6G networks in roaming scenario.
- Implementing load balancing or load prediction mechanisms, which also involve the Management Data Analytics Service (MDAS) to determines the appropriate IASS platform to use based on the movement of the object.

The rest of the paper is organized as follows. Section II examines various studies related to smart city safety, IoT anomaly detection, and audio and image analysis. Section III introduces the proposed architecture for IoT data correlation and anomaly detection. In Section IV, the scenarios that utilizes mentioned architecture are described. Finally, Section V draws a conclusion and outlines the scope for future research.

II. Related Work

In [5], Shankar and Maple propose a hybrid framework for securing the IoT-enabled smart city infrastructure. They introduce a secure IoT network architecture for smart cities combining blockchain and deep learning to safeguard privacy and credibility. Data from sensors, wearables, and Closed-Circuit Television (CCTV) cameras are gathered and analyzed within a blockchain layer. Subsequently, a deep learning algorithm processes this information to ensure efficient resource utilization. While this paper proposes an efficient architecture to preserve the privacy and trustworthiness of collected data in smart cities, it does not discuss how the data from various sources (e.g., sensors and cameras) are correlated.

In [6], Al-Amri et al. discuss the importance of the massive amounts of data collected from numerous IoT devices. The authors highlight the challenges posed by the dynamic nature of data. Accordingly, the paper benchmarks different studies that apply machine learning techniques for detecting anomalies in the data collected from IoT devices. The study covers various aspects such as the nature of the data, types of anomalies, detection learning modes, window models, datasets, and evaluation criteria. The main focus of this paper is to discover anomalies in the data collected from IoT devices rather than employing the collected data to detect and mitigate security threats in smart cities. In [7] [8], the authors focus on the security aspect of IoT networks, exploring the effectiveness of machine learning algorithms in detecting anomalies within network data. Like the aforementioned research, these studies employ various machine learning algorithms for intrusion detection, rather than using data collected from IoT devices such as cameras and microphones to identify security threats in a smart city context.

Eaton et al. [9] propose a machine-learning engine which learns movements and/or activities in a video over time and distinguishes between normal and abnormal behavior within a scene. The system is trained with normal behaviors and detects anomaly in case of abnormal behaviors. However, the audio has not been taken into consideration and the proposed engine is suitable for streaming video data only with main focus on vehicle use cases.

In an audio and video analysis management system, allowing users who own the audio and video collection device to perform the audio and video analysis calculations based on the audio and video analysis management platform. The patent claims that data processing capabilities can filter out audio and use or interfere information in the video, automatically analyze, extract specific useful information from audio and video sources, and establish a mapping between audio, video and objects, events or their behavior. The audio and video analysis management system comprises of variety of algorithms designed according to user requirements. The patent does not mention more details are mentioned about the process of how this information is linked.

The [10] patent proposes a method and apparatus for automatically indexing the locations of specified events on a video tape. The events, for example, include touchdowns, fumbles and other football-related events. An index to the locations where these events occur are created by using both speech detection and video analysis algorithms. A speech detection algorithm locates specific words in the audio portion of the video tape.

Previous studies primarily focus on enhancing security through ML-algorithms for IoT intrusion detection, or on object recognition from video streams, often neglecting the analysis of audio and the correlation between audio and video data. In contrast, our architecture leverages AI to perform real-time analysis of IoT data, enabling automatic anomaly detection. The outcomes of this analysis are promptly communicated to the appropriate operators via orchestrators, allowing them to pinpoint the location of the IoT device exhibiting anomalous behavior.

III. Architecture

Fig. 1 introduces proposed automated data correlation framework for IoT anomaly detection, consisting of an IASS module, an orchestrator and interface towards the 6G core. For example, in the context of smart cities, the IASS framework analyses data from the IoT devices, detects suspicious activities that may pose threats to public safety and security, identifies the location of IoT device exhibiting anomalous behavior and asks help from relevant public welfare and emergency authorities.

Fig. 1. IASS interaction with 6G mobile network.

The IASS module is composed of four logical functions: Content Extraction, Anomaly Detection, Geolocator and Anomaly Interpreter, as shown in Fig. 2. In the first step, the Content Extraction function examines the input to identify its type. Subsequently, Anomaly Detection function employs AI/ML algorithms to classify the extracted information. This step labels the suspicious behavior and forwards the results to Geolocator and Anomaly Interpreter logical functions for further actions. The Geolocator function interfaces with core network through orchestrator and determines the location of the IoT devices. Finally, the Anomaly Interpreter function determines the urgency of the situation e.g., whether it is a medical emergency situation, threat to people in the area of interest etc. The Anomaly Interpreter logical function may also communicate with Geolocator function to obtain the location information of the threat and to share it with the relevant public safety agencies, such as health services or fire departments.

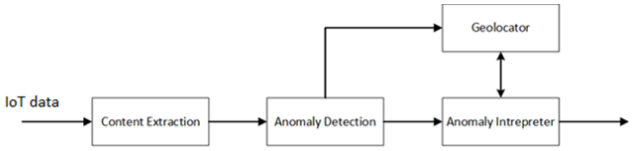

Fig. 2. IASS module.

The Orchestrator, based on the received information from IASS module(s), sends the IoT data and/or the location of the respective IoT device to the public safety agency. Through the orchestrator, the geolocator should communicate with 6G Core to obtain real-time location and update the same with the public safety agency. If the Orchestrator requests the 6G Core to provide the location information of the IoT device, the 6G Core will obtain this information from a location database (similar

to GMLC) of the 6G Core and forwards it to the Orchestrator. This will help the public safety agency to reach the exact location in time and tackle the situation faster. Additionally, based on the location, the Orchestrator may determine the ideal IASS module to be used, allowing for proactive service function chaining of Virtual NFs (VNFs) depending on the evolution of a public safety use case.

In brief, this architecture performs the followings subtasks:

- Prepares the data
- Analyzes the data and extract incidents
- Geotags the data
- Triggers (informs) action

In order to perform the mentioned tasks, the communication between IASS and core network elements is illustrated in Fig. 3.

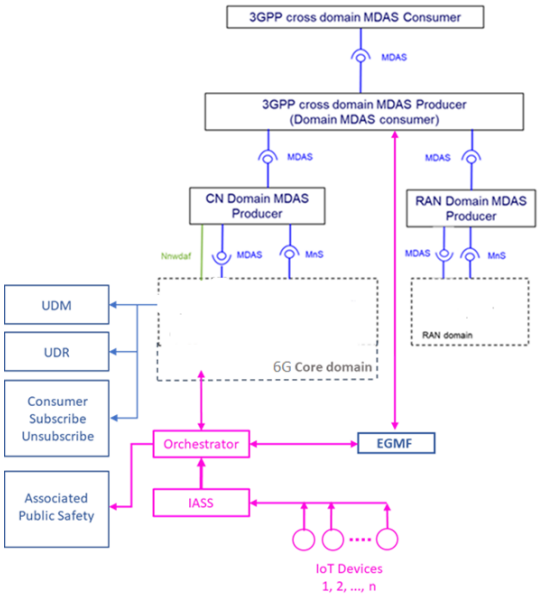

Fig. 3. Communication between IASS and core network elements.

A. *Intelligent & Autonomous Security Service*

As discussed above the IASS comprises of four modules:

- Content Extraction receives the incoming data from different sources around the city and does some processing such as data cleaning, normalization, dimensionality reduction and so on that are further analyzed by Anomaly Detection module.
- Anomaly Detection consists of pre-trained classification and/or other ML algorithms to classify, and label received data and detect suspicious information.
- Geolocator determines the location of the IoT devices. The location of the IoT device of interest can be detected by techniques that will be introduced in future paper. After determining the position of the IoT device, the geo-locator module will send the location information to Anomaly Interpreter module.
- Anomaly Interpreter receives the results from various modules explained earlier. If a threat or abnormal behavior is detected, the model proceeds to find the geo-location of that IoT device. Based on this information, this module determines the urgency of the situation and shares the information to the associated public safety agency to take further steps. In addition, this module informs the Orchestrator of the location of device of interest in real-time.

B. *Orchestrator and interaction with core and radio network*

The Orchestrator module has four major tasks:

- Sharing the information about abnormal behavior to relevant public safety agency.
- Prioritizing IASS services based on the application.
- Tracking the location of one or multiple IoT devices involved in the situation by interacting with core network; share its location information and live feed from IoT devices in surroundings of determined location with public safety agency.
- Updating network configuration.

The Orchestrator is architecturally location agnostic and flexible, i.e. it can be located in the enterprise network or as an application in operator network. This increases its usability, and the location can be tailored to take into account e.g., security, business, and regulatory constraints.

When multiple instances of IASS modules exist (due to load balancing, various service providers / operators, coverage reasons or geographic location) the information is forwarded to the Orchestrator, which connects to core network in order to update network configuration and track the location of object of interest.

If multiple instances of IASS belong to different operators, a single operator manages overseeing Orchestrator, which aggregates data from other operators through their respective orchestrators. The Orchestrator may also be operated by public authority or third-party service provider. As objects move, different functions or applications can be chained together; and service chaining concept can also be utilized.

The link between “EGMF” and “3GPP cross domain MDAS” for inter operator communication between orchestrators is a new interface to be introduced in 6G standard.

IV. Example Scenarios

Depending on the situation, whether the object carries an active radio device with (e)SIM card or not, there are different scenarios and therefore information exchange will be through various network modules. A reasonable assumption here is that IoT devices or IASS will have some a sort of radio equipment, which will capture radio signals for the object to determine which operator does it belong. In scenarios where an object of interest belongs to a different network operator, efficient cross-operator communication is essential to ensure timely and accurate information exchange. The proposed architecture addresses this need by leveraging a collaborative approach involving multiple network operators and their respective network functions.

A. *Scenario 1: Object carries an active radio device with (e)SIM card*

In this scenario as shown in Fig. 4, IoT devices forward the collected information (image/audio/video) to IASS. In case of anomality, the information will be forwarded to the Orchestrator. The Orchestrator sends information about object to Core MDAS via Exposure Governance Management Function (EGMF), which subsequently forwards this information to LCS/NWDAF via NEF for location tracking of the object.

LCS/NWDAF maintains user level configuration and can provide an estimated motion path for the object from the last location based on aggregated motion data of users in the vicinity. After receiving probable motion path of object, the Orchestrator determines IoT devices or drones that fall within the predicted motion path. The Orchestrator then gathers and shares live data feeds of IoT devices or drones with public safety agency. This process enables real-time tracking of the object.

If object belongs to the other operator, the Orchestrator broadcasts the object Subscriber Identity Module (SIM) information to other operators and asks for both user information and location. The other operator, in response, extracts the user information and location via NEF, NWDF and so on. The acquired information is then forwarded to the Orchestrator and accordingly shared with the associated public safety agency.

B. *Scenario 2: Object does not carry an active (e)SIM*

If target device is SIMless as shown in Fig. 5, object can be traced using probable location (trajectory prediction) received from radio network and MDAS to track the object in real time.

Another solution involves collecting user information from a database that contains samples of users' audio and images maintained by network analytic module.. In this scenario, IoT device forwards the collected image/voice/audio data to IASS. In case of anomality, the received information will be forwarded to the public safety agency to be compared with image and voice registry. Simultaneously, this information is broadcasted to all operators that provide IoT cameras in order to track object via AI techniques such as image recognition to spot target in the new location.

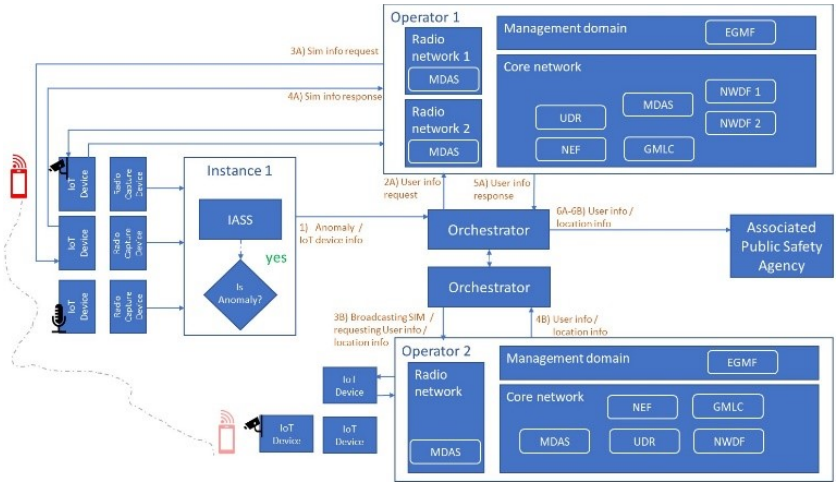

Fig. 4. Object carries SIM

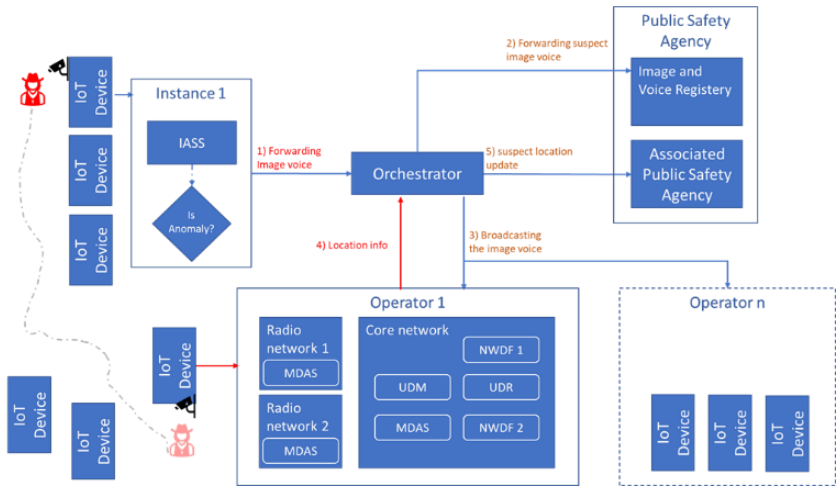

Fig. 5. Object does not have subscription or mobile is off

V. Conclusion and Future Discussion

The automated data correlation framework for IoT anomaly detection in smart cities represents integration of machine learning technologies aimed at strengthening public safety and security. At its core, the framework comprises the Intelligent & Autonomous Security Service (IASS) module, an Orchestrator, and an interface with the 6G Core network. This architecture is designed to process vast amounts of data generated by IoT devices dispersed throughout environments, swiftly detect anomalies, determine device locations, and efficiently alert relevant public safety agencies in real-time.

The IASS module, with its four logical functions - Content Extraction, Anomaly Detection, Geolocator, and Anomaly Interpreter, serves as the frontline defense mechanism. Content Extraction analyzes incoming data, cleansing and normalizing it for subsequent analysis. Anomaly Detection leverages machine learning algorithms to categorize data and pinpoint suspicious activities. Geolocator then steps in to geotag IoT devices, allowing for precise location determination. Finally, the Anomaly Interpreter evaluates the urgency of detected anomalies and shares pertinent information with the appropriate public safety agencies.

Complementing the IASS module is the Orchestrator, which orchestrates communication with the 6G Core network, facilitates the prioritization of services, tracks IoT device locations, and dynamically updates network configurations as needed. Its ability to interact with core network elements enables efficient information exchange and enhances the responsiveness of public safety agencies.

Operational scenarios of the paper illustrated the framework's adaptability to diverse situations. In Scenario 1, where objects carry active radio devices with SIM cards, the framework efficiently tracks and traces objects in real-time through coordinated efforts between the IASS module, Orchestrator, and core network functions. Conversely, Scenario 2 addresses instances where objects lack active radio devices, showcasing alternative methods such as trajectory prediction and database analysis to achieve similar outcomes. In 5G/6G networks this can be done in minimal basis and securely. And only such SIM-information is exchanged for which there is need and IoT device owner has agreed upon.

In the future, several opportunities for improvement and implementation can be identified. Firstly, refining inter-operator communication channels, such as the introduction of the EGMF and 3GPP cross-domain MDAS interface, will strengthen collaboration and information sharing across operators, thereby improving overall response capabilities. Additionally, advancements in data analytics and anomaly detection algorithms will be crucial in enhancing the framework's ability to rapidly identify and respond to emerging threats. Moreover, efforts to optimize the implementation of this architecture within existing urban infrastructures, including considerations for scalability, interoperability, and resource allocation, will be paramount to its widespread adoption and effectiveness in ensuring public safety and security in smart cities.

References

[1] 3GPP, “Technical Specification Group Services and System Aspects; 5G System (5GS) Location Services (LCS); Stage 2, TS 23.273 (Release 17),” 2023.

[2] 3GPP, “Technical Specification Group Core Network and Terminals; 5G System; Network Exposure Function Northbound APIs; Stage 3, TS 29.520,” 2023.

[3] 3GPP, “Technical Specification Group Services and System Aspects; Procedures for the 5G System (5GS); Stage 2, TS 23.502,” 2023.

[4] 3GPP, “Study on enhancement for data collection for NR and ENDC, TR 37.817,” 2022.

[5] A. Shankar and C. Maple, “Securing the Internet of Things-enabled smart city infrastructure using a hybrid framework.,” in *Computer Communications,* vol. 205, pp. 127-135, 2023.

[6] R. Al-amri, R. K. Murugesan, M. Man, A. F. Abdulateef, M. A. Al-Sharafi and A. A. Alkahtani, “A review of machine learning and deep learning techniques for anomaly detection in IoT data,” in *Applied Sciences,* vol. 11, p. 5320, 2021.

[7] M. Hasan, M. M. Islam, M. I. Islam Zarif and M. Hashem, “Attack and anomaly detection in IoT sensors in IoT sites using machine learning approaches,” in *Internet of Things,* vol. 7, p. 100059, 2019.

[8] A. Singh, A. Mishra, A. Antil, B. Bhushan and A. Chauhan, “Anomaly Based IDS in Industrial IoT,” in *International Conference on Smart Systems for applications in Electrical Sciences (ICSSES)*, 2023.

[9] J. E. Eaton, W. K. Cobb, D. G. Urech, D. S. Friedlander, G. Xu, M.J. Seow, L. W. Risinger, D. M. Solum, T. Yang, R. K. Gottumukkal and K. A. Saitwal, “Semantic representation module of a machine-learning engine in a video analysis system,” United States of America Patent US20180204068A1, 19 July 2018.

[10] Y.L. Chang and W. Zeng, “Method and apparatus for extracting indexing information from digital video data,” United States of America Patent USOO5828.809A, 27 October 1998.