Edge Data Analytics

Subjects: Computer Science, Cybernetics

Contributor: Poornima Mahadevappa , Redhwan Al-amri , Gamal Alkawsi , Ammar Ahmed Alkahtani , Mohammed Fahad Alghenaim , Mohammed Alsamman

Edge data analytics refers to processing near data sources at the edge of the network to reduce delays in data transmission and, consequently, enable real-time interactions. However, data analytics at the edge introduces numerous security risks that can impact the data being processed. Thus, safeguarding sensitive data from being exposed to illegitimate users is crucial to avoiding uncertainties and maintaining the overall quality of the service offered. Most existing edge security models have considered attacks during data analysis as an afterthought.

edge computing edge data analytics security threats security models edge applications

1. Introduction

Edge computing is a distributed systems paradigm that aims to offload selected services of applications from the cloud and bring them closer to the end-user. It is a generic term that captures associated paradigms, such as fog computing, mist computing, mobile edge computing, and cloudlet computing. Services are hosted at the edge of the network on nodes, such as routers, gateways, and micro-data centers. Data generated from end-users or sensors on Internet of Things (IoT) devices or sensors are analyzed and processed on the edge, which is nearer to the data source. Although edge nodes may be resource-limited when compared to the cloud, data analytics on the edge offers three benefits: (i) better responsiveness by reducing round-trip communication latency, (ii) a higher degree of data privacy, and (iii) minimizing the ingress bandwidth demand to the cloud [1].

Security is of paramount importance when using edge nodes for data processing since a large attack surface is exposed. User-generated or sensor data that are transferred to edge nodes must be protected for confidentiality and integrity, even if an edge node is attacked [2]. Data must be protected even when attacks, such as Man-In-The-Middle (MITM), Denial of Service (DoS), eavesdropping, and others (to be discussed later), are underway. Many attacks while performing data analytics have been previously understood in the context of the cloud and are inherited by the edge (for example, Man-In-The-Middle (MITM), Denial of Service (DoS), or eavesdropping). Recent security breaches or incidents in edge computing serve to highlight the severity of security risks in this domain. For example, in 2020, a vulnerability known as “Ripple20” was discovered, affecting millions of IoT devices across various industries, including edge computing devices. This vulnerability allowed attackers to remotely execute malicious code, potentially leading to data breaches or system compromise [3]. Another notable incident involved the exploitation of a vulnerability in the “Treck” TCP/IP stack, affecting numerous IoT and edge devices. This vulnerability, dubbed “AMNESIA:33,” enabled attackers to execute remote code execution, DoS attacks, and

other malicious activities [4]. These incidents underscore the importance of addressing security vulnerabilities in edge computing environments to mitigate the risk of data breaches, system compromise, and other cyber threats.

2. Background of Edge Data Analytics

Edge data analytics allows preprocessing data for obtaining real-time decisions. The data flow is similar to that on the cloud, with the difference that edge resources process data. The data analytics process will need to consider the following five aspects: (a) data source, (b) content format, (c) data storage, (d) data staging, and (e) data processing [5]. Data processing on edge nodes enables real-time interactions. The flow of data in an edge computing layer sandwiched between the cloud and end-user devices layer (referred to as the Internet of Things (IoT)) is shown in **Figure 1**. In edge-based IoT applications, sensing, collecting, and analyzing the data depend on the types of services they provide.

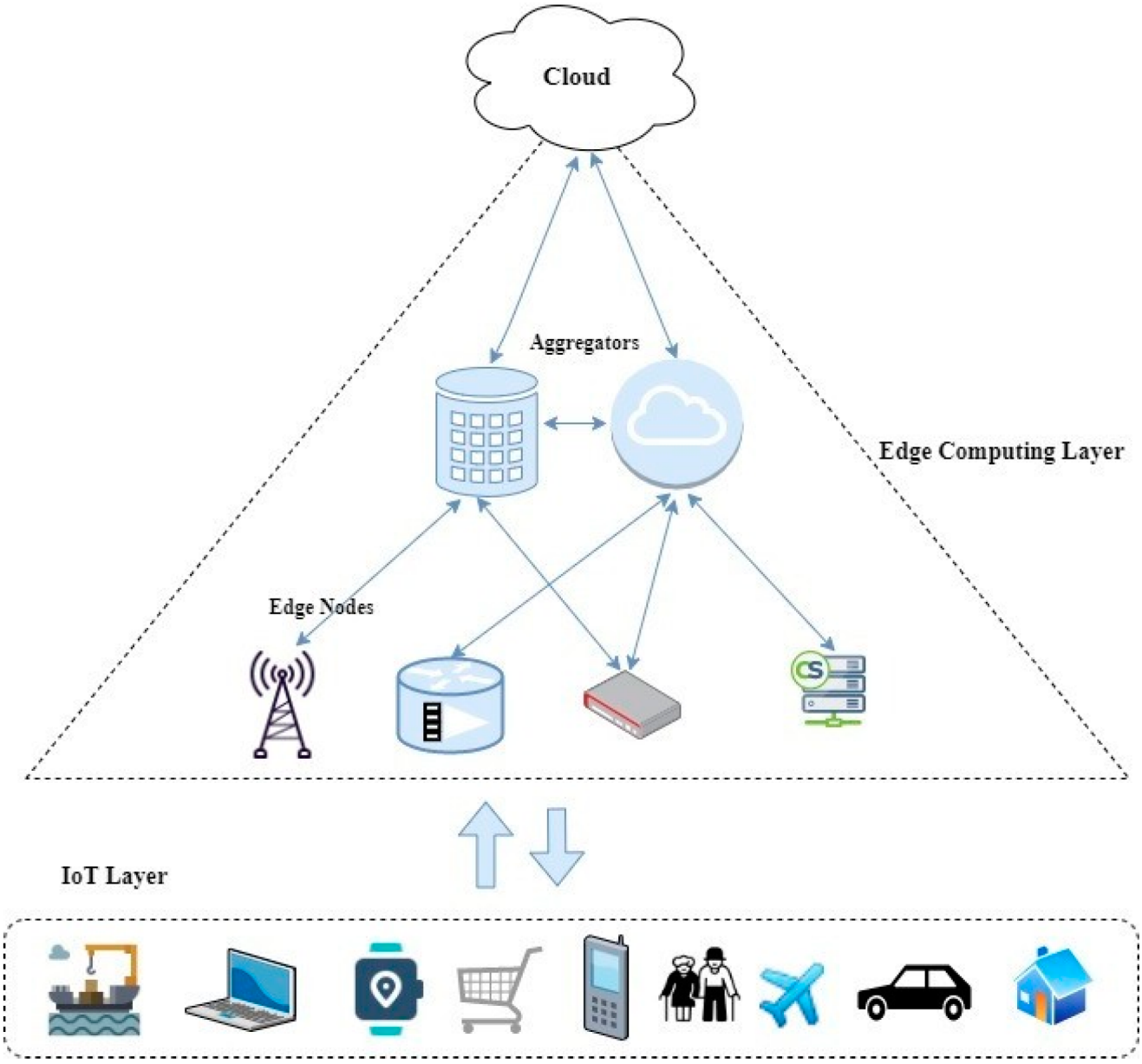

Figure 1. Flow of data in the edge computing layer.

A data processing model has been proposed for edge computing systems [6]. Heterogeneous data are collected from ubiquitous devices and pushed forward through communication channels to preprocess. Real-time analysis and decision-making occur to support quick responses to the applications on IoT devices. The services offering real-time analysis may be transferred to the cloud. Data processing depends on the information gathered from the hierarchical edge layer, how quickly the data are collected, and how they trigger the specific services for decision-making. The components that support this process are shown in **Figure 2**.

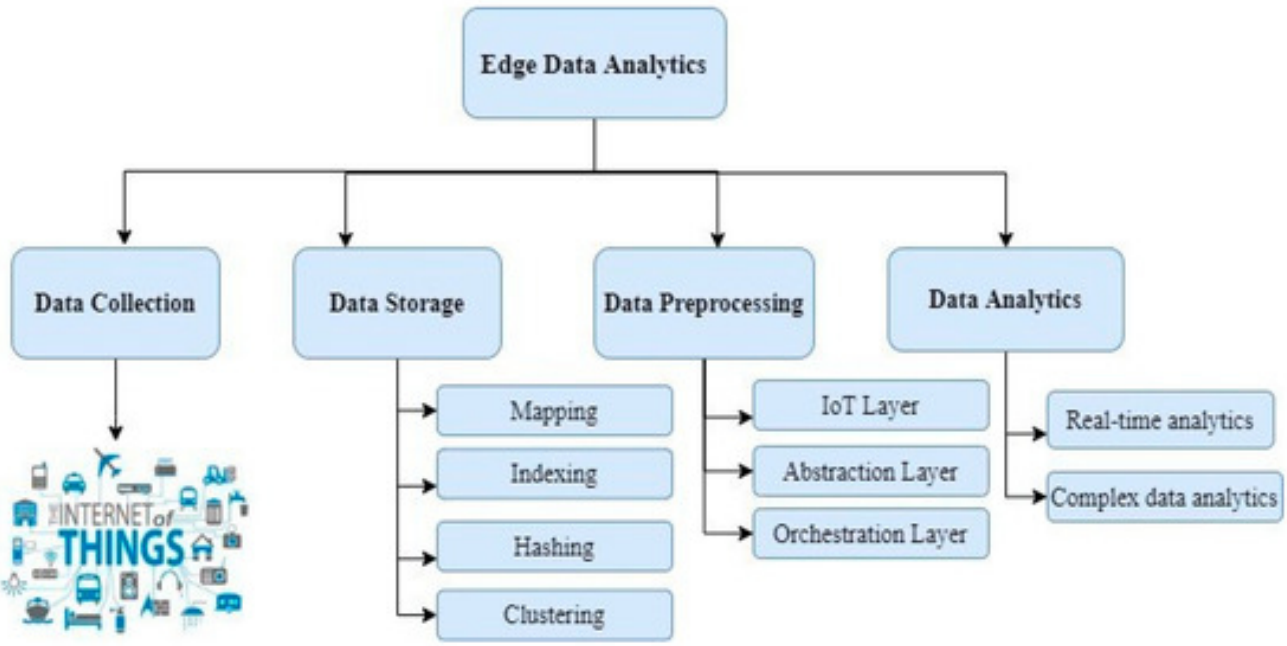

Figure 2. Components of edge data analytics.

Data Collection—All devices are the primary source to generate data. The devices may be electrical appliances, homes, or embedded systems connected with the unique Internet Protocol (IP) to establish connection and communication among them. Edge nodes closer to devices collect data and support computation for the IoT devices' applications by offloading tasks across the cloud and edge nodes. Various deployment models deploy the task as middleware between the cloud and IoT devices with efficient resource utilization [7].

Data Storage—Data collected from devices can be stored in either the device or on edge nodes in virtual machines or containers [8]. Typical efficient storage relies on techniques such as mapping, hashing, clustering, replication, indexing, and so on. Data are collected in clusters and sent to the storage devices [9]. In indexing, indexes are created based on the extraction, recognition, and labeling of real-time data, such as video streams or social media data [10]. In replication, the data are duplicated to support the data-intense applications by encapsulating the coherent data logically [11].

Data Processing—IoT verticals, abstraction layers, and orchestration layers are the three components responsible for data processing in edge computing architectures. IoT verticals include the application that is in use. They

provide multitenancy to host the application on edge data servers and provide flexibility and interoperability to the edge nodes. The abstraction layer provides a uniform virtualized platform through a generic API to monitor, provision, and control physical resources. The orchestration layer includes data API and orchestration layer API, which are responsible for node placement or node selection, run-time monitoring, control during execution, and optimizing data-driven decisions [12][13].

Data Analytics—Data collected from IoT devices are preprocessed on the edge nodes through intensive real-time task analysis. This establishes real-time interactions between the edge nodes and the users. For example, generating a diagnosis report for a doctor to treat the patient remotely [14] or traffic signal detection for unmanned autonomous vehicles [11]. The volume of data that may be challenging for the edge nodes to analyze is pushed to the cloud for more complex data analysis [15][16]. Machine learning (ML) algorithms are usually employed to provide long-term predictive decisions [17].

Decision-Making in Edge Data Analytics

Data analytics and decision management are two critical components of decision-making. The report generated from data analytics is used by the decision management component to identify what decisions should be made. For example, in traffic management applications, information about traffic density, vehicle-specific data, and movement of other vehicles and pedestrians are collected to perform quick data analytics and generate decisions on traffic flow. Hence, agility in decision-making triggers the business process, resource utilization, and customer satisfaction. Based on agility, decision-making is divided into predictive and reactive models: (i) Predictive models rely on the cloud to collect large amounts of data and perform long-term data analysis to identify the best decisions. They evaluate decisions based on various policies in the applications and improve the predictive analysis over time. (ii) Reactive models respond to an event with reactive decisions within a short time interval. These models achieve real-time support without focusing on what the system might look like in the future. The key characteristics of real-time support are the most suitable for edge computing applications. To obtain a decision at an adequate response time, edge nodes have to be placed closer to IoT devices [18]. Whether services need to be placed on the cloud or edge is an optimization problem [19].

However, when edge nodes are scattered and placed closer to IoT devices, monitoring these nodes will be challenging. Geographical factors, such as network infrastructure and regulatory environments, significantly influence the design and deployment of edge security solutions [20]. In regions with limited network infrastructure, edge security solutions must adapt to unreliable or slow connectivity, potentially requiring decentralized architectures to ensure data processing and threat detection can occur locally [21]. Regulatory environments, such as GDPR (General Data Protection Regulation) or HIPAA (Health Insurance Portability and Accountability Act of 1996), dictate strict requirements for data privacy and security, impacting how data are stored, processed, and transmitted in edge computing environments [22]. Compliance with these regulations may necessitate additional encryption measures, data residency requirements, or auditing protocols in the design of edge security solutions. Moreover, variations in network latency due to geographical distances can affect the responsiveness of security measures, prompting the optimization of algorithms or deployment strategies to accommodate latency-sensitive

applications. Additionally, geographical factors influence physical security considerations, as edge devices deployed in remote or inaccessible locations may require robust physical protection against tampering or unauthorized access [23]. Overall, accounting for geographical factors is essential in designing and deploying effective edge security solutions that address the unique challenges posed by different environments. Intruders can easily compromise and gain access to the edge layer, and thus they can mine or steal data that are exchanged among edge nodes [24]. In cloud computing, there are regulations and obligations for data protection, as per the European Commission [8]. However, no such standards exist in edge computing, which makes them vulnerable to security attacks. In the next section, the security models that affect decision-making and the normal functioning of an application are reviewed.

References

1. Yahuza, M.; Bin Idris, M.Y.I.; Wahab, A.W.B.A.; Ho, A.T.S.; Khan, S.; Musa, S.N.B.; Taha, A.Z.B. Systematic review on security and privacy requirements in edge computing: State of the art and future research opportunities. IEEE Access 2020, 8, 76541–76567.
2. Hu, P.; Dhelim, S.; Ning, H.; Qiu, T. Survey on fog computing: Architecture, key technologies, applications and open issues. J. Netw. Comput. Appl. 2017, 98, 27–42.
3. Röckl, J.; Wagenhäuser, A.; Müller, T. Veto: Prohibit Outdated Edge System Software from Booting. In Proceedings of the International Conference on Information Systems Security and Privacy, Lisbon, Portugal, 22–24 February 2023; pp. 46–57.
4. Rajkumar, V.S.; Stefanov, A.; Musunuri, S.; de Wit, J. Exploiting Ripple20 to Compromise Power Grid Cyber Security and Impact System Operations. IET Conf. Proc. 2021, 2021, 3092–3096.
5. Hashem, I.A.T.; Yaqoob, I.; Anuar, N.B.; Mokhtar, S.; Gani, A.; Khan, S.U. The rise of “big data” on cloud computing: Review and open research issues. Inf. Syst. 2015, 47, 98–115.
6. Dautov, R.; Distefano, S.; Bruneo, D.; Longo, F.; Merlino, G.; Puliafito, A. Data processing in cyber-physical-social systems through edge computing. IEEE Access 2018, 6, 29822–29835.
7. Zhang, J.; Ma, M.; He, W.; Wang, P. On-demand deployment for IoT applications. J. Syst. Archit. 2020, 111, 101794.
8. Tychalas, D.; Karatza, H. A Scheduling Algorithm for a Fog Computing System with Bag-of-Tasks Jobs: Simulation and Performance Evaluation. Simul. Model. Pract. Theory 2020, 98, 101982.
9. Bonomi, F.; Milito, R.; Zhu, J.; Addepalli, S. Fog computing and its role in the internet of things. In The First Edition of the MCC Workshop on MOBILE Cloud Computing—MCC’12; ACM Press: New York, NY, USA, 2012; p. 13.

10. Nikouei, S.Y.; Xu, R.; Nagothu, D.; Chen, Y.; Aved, A.; Blasch, E. Real-Time Index Authentication for Event-Oriented Surveillance Video Query using Blockchain. In Proceedings of the 2018 IEEE International Smart Cities Conference (ISC2), Kansas City, MO, USA, 16–19 September 2018.
11. Ouyang, Z.; Niu, J.; Ren, T.; Li, Y.; Cui, J.; Wu, J. MBBNet: An edge IoT computing-based traffic light detection solution for autonomous bus. J. Syst. Archit. 2020, 109, 101835.
12. Wen, Z.; Yang, R.; Garraghan, P.; Lin, T.; Xu, J.; Rovatsos, M. Fog orchestration for IoT Services: Issues, Challenges and Directions. IEEE Internet Comput. 2017, 21, 16–24.
13. Dsouza, C.; Ahn, G.-J.; Taguinod, M. Policy-driven security management for fog computing: Preliminary framework and a case study. In Proceedings of the 2014 IEEE 15th International Conference on Information Reuse and Integration (IEEE IRI 2014), Redwood City, CA, USA, 13–15 August 2014; pp. 16–23.
14. Dou, C.; Zhang, S.; Wang, H.; Sun, L.; Huang, Y.; Yue, W. ADHD fMRI short-time analysis method for edge computing based on multi-instance learning. J. Syst. Archit. 2020, 111, 101834.
15. Alkawsi, G.; Al-Amri, R.; Baashar, Y.; Ghorashi, S.; Alabdulkreem, E.; Tiong, S.K. Towards lowering computational power in IoT systems: Clustering algorithm for high-dimensional data stream using entropy window reduction. Alex. Eng. J. 2023, 70, 503–513.
16. Al-Amri, R.; Murugesan, R.K.; Almutairi, M.; Munir, K.; Alkawsi, G.; Baashar, Y. A Clustering Algorithm for Evolving Data Streams Using Temporal Spatial Hyper Cube. Appl. Sci. 2022, 12, 6523.
17. Varghese, B.; Gohil, B.N.; Ray, S.; Vega, S. Research challenges in query processing and data analytics on the edge. In Proceedings of the CASCON 2019 Proceedings—Conference of the Centre for Advanced Studies on Collaborative Research—Proceedings of the 29th Annual International Conference on Computer Science and Software Engineering, Toronto, ON, Canada, 4–6 November 2019; pp. 317–322.
18. Bellavista, P.; Berrocal, J.; Corradi, A.; Das, S.K.; Foschini, L.; Zanni, A. A survey on fog computing for the Internet of Things. Pervasive Mob. Comput. 2018, 52, 71–99.
19. Brogi, A.; Forti, S.; Guerrero, C.; Lera, I. How to place your apps in the fog: State of the art and open challenges. Softw.—Pract. Exp. 2020, 50, 719–740.
20. Shruti; Rani, S.; Srivastava, G. Secure hierarchical fog computing-based architecture for industry 5.0 using an attribute-based encryption scheme. Expert Syst. Appl. 2024, 235, 121180.
21. Mamatas, L.; Demiroglou, V.; Kalafatidis, S.; Skaperas, S.; Tsaoussidis, V. Protocol-Adaptive Strategies for Wireless Mesh Smart City Networks. IEEE Netw. 2023, 37, 136–143.
22. Jumani, A.K.; Shi, J.; Laghari, A.A.; Hu, Z.; Nabi, A.U.; Qian, H. Fog computing security: A review. Secur. Priv. 2023, 6, e313.

23. Ali, M.; Naeem, F.; Kaddoum, G.; Hossain, E. Metaverse Communications, Networking, Security, and Applications: Research Issues, State-of-the-Art, and Future Directions. In IEEE Communications Surveys & Tutorials; IEEE: Piscataway, NJ, USA, 2023; p. 1.
24. Mukherjee, M.; Shu, L.; Wang, D. Survey of fog computing: Fundamental, network applications, and research challenges. In IEEE Communications Surveys and Tutorials; IEEE: Piscataway, NJ, USA, 2018; Volume 20, pp. 1826–1857.