An Examination of Virtualization Technologies for Enabling Intelligent Edge Computing

Kevin Afachao[1]
[1]*Department of Electrical, Electronic and Computer Engineering*
University of Pretoria
Pretoria, 0028, South Africa
u22851217@tuks.co.za

Adnan M. Abu-Mahfouz[1,2]
[2]*Council for Scientific and Industrial Research (CSIR)*
Pretoria, 0083, South Africa
a.abumahfouz@ieee.org

***Abstract*— At the frontier of computing today is the Internet of Things, rapidly connecting things and incorporating them into its larger ecosystem. Driving the rapid rise of the Internet of Things is the Edge Computing framework. While the Edge Computing framework has its shortcomings, it can be enhanced with the use of Artificial Intelligence and Virtualization technologies. This paper discusses several papers and earlier surveys focused on enabling virtual services in Intelligent Edge Computing. Few works are dedicated to advancing virtualisation's incorporation in Intelligent Edge Computing. However, the current works studied provide an insight into the research's direction. Furthermore, future directions are stated and highlighted to encourage research in this domain.**

Keywords—*Internet of things, Intelligent edge, Microservices, Network function virtualization, Software-defined network, Distributed systems, Resource management*

I. INTRODUCTION

The IoT has led to a huge paradigm shift giving machines and users access to the internet by various means. The IoT empowers various computing technologies like Cloud computing and Edge computing. However, the IoT environment is volatile and heterogeneous. Designing a framework that takes into account all the different factors affecting performance in a single framework is a challenge since such a framework may have many points of failure. Cloud computing, on one hand, fails to provide users with a relevant privacy guarantee, location-aware services and low latency services [1]. The aforementioned areas give Edge computing leverage over cloud computing for IoT applications. On the other hand, Edge computing faces some challenges. The heterogeneous nature of the environment means that user mobility, device constraints, user requests, and available bandwidth all affect performance, and all of these factors are volatile [2].

Central to Artificial Intelligence (AI) today is the deep learning model, a subclass of machine learning [3]. Significant amounts of research are going into this area because of its ability to scale on structured and unstructured data, as well as the optimal performance-to-error rate ratio. Applications arising from deep learning and edge computing are beginning to emerge. While it is not a surprise, AI has brought several benefits to Edge Computing to augment its weaknesses such as network optimization to reduce network congestion, AI hardware to accelerate computing and varied application context scenarios for user satisfaction [4]. The inter-discipline, Intelligent Edge (IE) is now a curious subject for both industry and research as it brings the concept of Edge Computing and AI together. Furthermore, Intelligent Edge Computing introduces additional computational demand on limited network resources and optimisation problems. Thus, virtualization and its associated technologies are proposed to handle this constraint.

The contribution of this literature survey is to show the scope of available technologies in virtualization for Intelligent Edge computing, its advances as well as challenges and opportunities. Also, a contribution is made to the ongoing discussion of how to best manage and optimize large-scale Edge networks for improved reliability and performance. The paper is outlined as follows. The next section presents the background, followed by a discussion of the various applications of virtualized services in the Intelligent Edge and opportunities for research. Finally, the paper ends with a conclusion.

II. RELATED WORKS

Application deployment can be achieved using virtualization for the Intelligent Edge. Modern hardware on IoT end nodes allows lightweight virtualization to be adopted such as containers and Unikernel [5]. Many surveys address the expanse of literature related to virtualization in Edge computing [6]–[8]. On the other hand, surveys related to Intelligent Edge Computing for virtualization are few. The survey conducted by Zhu et al. [6] explores Deep Learning (DL) techniques in Mobile Edge Computing (MEC) for virtualized networks and presents a discussion on research investigations ongoing. Uncertainties for next-generation mobile network systems are also explored. However, the paper emphasizes Mobile Edge Caching and its development. The literature review done by Duc et al. [7] presents orchestration parameters or areas to be considered in designing systems for the virtualized Edge network, from a machine learning perspective with no mention of the Intelligent Edge. The paper by Murkherjee et al [8] extensively reviews AI microservices deployed in the edge computing framework and security implementations to minimize risks and vulnerabilities. The article addresses a specific niche in Intelligent Edge Computing which is the security and privacy architecture but lacks a broad view of advances in the field. The scope of IoT and Intelligent Edge Computing surveys discuss technologies in the trend of MEC such as NFV, SDN and ICN, however, few consider other virtual technologies like microservices [1], [9]. The IoT research domain is expanding rapidly. Ongoing research explores new paradigms that can be incorporated into Intelligent Edge Computing such as microservice. This literature review considers microservices in addition to other mainstream virtual technology like Software Defined Networking (SDN) and Network Function Virtualization (NFV). Table I shows a summary of the works presented.

III. ENABLING TECHNOLOGIES

A. *Software Defined Networking*

Software Defined Networking (SDN) addresses the challenge of network flexibility faced by heterogeneous, interoperable networks such as Edge-IoT networks. In heterogeneous networks, the flexibility of a network refers to its ability to adapt the available network resources, such as flows or topology, to changes in design requirements, e.g., shorter latency budgets or different traffic distributions. Network flexibility is a challenge because of the differences in the operating systems and protocols available in the system architecture. It facilitates a decentralized network traffic and routing scheme for both virtual and physical devices. The SDN architecture is multi-layered consisting of a data plane, application plane and control plane [10]. The control plane and data plane work together to ensure that the network is functioning properly and that the packets are being forwarded to the correct destination. Whereas the application plane acts as an intermediary between applications and the rest of the SDN. The SDN framework gives the network added functionalities to make organization easier. The SDN employs a central controller to monitor and coordinate the entire network. The data plane contains forwarding devices such as virtual switches and physical switches and the control plane oversees the functions of the data plane. A policy on how to execute data packet forwarding and manipulation at the data plane comes from the control plane. The control plane plays a central part in SDN functionality. By organizing the data plane and the application plane, the central controller in the control plane optimizes the network to deliver on performance as well as translate software requirements into policy for the data plane. The application plane contains applications which adjust performance settings to suit network conditions to prevent overloading of the network. By decoupling the network control plane from the data plane, the SDN controller has access to a global view of the network as well as tracking all network data, such data is profitable for network analytics and machine learning (ML) insights.

The heterogeneity of networks poses a challenge to system maintenance and application orchestration in SDN [10]. Cognition was proposed to handle the system dynamics, however, the network nodes proved incapable of performing the additional instructions. This is being attempted once more with the improvement in device performance [11]. In [12], the authors encounter the dynamic routing problem in an IoT network, due to the time cost incurred. The routing problem initially was solved with the Shortest Path First (SPF) algorithm [13], however, on large scale, the algorithm does not meet performance requirements. The paper does a comparison between a few Machine Learning (ML) algorithms to determine which has the best performance: Max-Min Ant Swarm, Neural Network, and SPF. This leads to the discovery that Neural Networks outperform other ML algorithms in the execution time of services for SDN. Recognizing the potential of Neural Networks, Cui et al [14] propose to push the potential of SDN by aiding it with a Neural Network to carry out effective load balancing for network traffic management. From simulations, the Neural Network achieves a 19% decrease in latency over the Round Robin algorithm. Unlike Li et al [5], Stampa et al [15], attempts the routing optimization problem for heterogeneous networks with Deep Reinforcement Learning (DRL). Deep Reinforcement Learning employs a software agent without any memory of the environment to explore and exploit a system until it can generate the most optimal policy for system performance. Despite Deep Reinforcement Learning requiring larger datasets and computing time for training, they stand a chance of producing overfitting data nevertheless due to its adaptability over long periods and its ability to reach a globally optimal solution it is the most preferred and in the case of [15], it reduces the network delay. The results showed a slight reduction in network delays compared to previous algorithms for routing optimization. The authors in [16] tried to uncover the ongoing process in the DRL neural network that makes it exceptional at reducing network delay, however, despite their efforts the method failed to make any discovery. Resource management in Vehicular Ad hoc Networks (VANETS) which consider caching, computing and networking altogether is complex. Thus, following the work done by Stampa et al [8], He et al [17] attempt to model this situation as a joint optimization problem for DRL to solve. DRL is used to formulate a policy for resource allocation. The results did not meet expectations, since the energy utilization is not efficient when DRL is employed compared to implementations without DRL. A similar attempt is made on smart city environments in [18], and likewise, the results did not exceed the performance of the baseline. As a preliminary work, Nakao et al [19] consider the research question of what the other application scenarios of Deep Learning (DL) in Mobile Virtual Networks (MVN) could be. Through monitoring network traffic data with SDN and categorizing the data with neural network pattern recognition, potential application cases are recognized with the bottleneck being minimizing errors in data collection.

B. *Microservices*

Commercially, applications are designed as a unit, with one programming language and function dependent on one another [20]. However, the application cannot be augmented with new functionalities quickly since that would mean redesigning the whole application. Also, application deployment features a single programming language-driven application, dependent on other functions working together. However, when one function fails during runtime the whole system terminates. This can be detrimental to security as a service [21]. On the other hand, Microservices are service-oriented modular units for software development. The microservices are deployed in containers such as Docker and orchestrated with Kubernetes. With a divide-and-conquer pattern and lightweight communication protocols, it achieves application task execution. As a new paradigm, it enables the deployment of multiple virtual units to carry out tasks without the task having a single point of failure from any of the units [20]. Due to the flexibility of microservices, they are deployed rapidly and new instances are created dynamically according to network traffic, service requests, service assurance and security. Microservices deploy AI applications for a varied number of services, security as a service, infrastructure as a service, and software as a service [22].

TABLE I. ENABLING TECHNOLOGIES FOR VIRTUAL SERVICE AT THE INTELLIGENT EDGE

Ref.	Objective	Contribution	Limitation
[12]	Traffic engineering with ML	Using ML to interface between the control plane and data plane of SDN to provide faster time gains for Edge network	Implementation does not perform better than the baseline
[13]	Load balancing with SDN-enabled network	Using Neural Network to enhance network load balancing performance in SDN	Implementation was carried out in small-scale simulation
[14]	SDN routing optimization	Using DRL for routing optimization to enhance the performance of SDN in Edge network	Performance decreases with a decrease in training time
[15]	SDN resource management	Using DRL for Network resource allocation policy in Edge networks enabled with SDN	Implementation is not energy efficient compared to baselines
[16]	Network latency prediction	Using Neural Network to predict network delay in SDN enabled Edge networks	Implementation does not perform better than the baseline
[17]	Predictive dynamic network routing	Using Neural Networks to solve the routing optimization problems in SDN enabled Edge networks	Implementation carried out on a small scale
[18]	SDN optimization for Smart Cities	Using DRL for Network resource allocation policy in Edge networks enabled with SDN	Implementation does not perform better than the baseline
[19]	Identifying mobile application scenarios	Using Neural Networks to identify application scenarios from data distributions in SDN enabled Edge networks	Added data introduces computational error
[23]	Microservice IoT application deployment	Considers AI applications in microservice deployment for IoT	Implementation is carried out on a small scale
[24]	Predictive Analysis of Data aided by Microservices	Using microservices to integrate various applications in IoT	Implementation is carried out on a small scale
[27]	Microservice IoT application deployment for Environmental Monitoring	Using microservices to integrate individual applications into one system	There is no implementation carried out
[28]	Microservice IoT application deployment for Industrial Monitoring	Using microservices for anomaly detection in Industrial machines.	There is no implementation carried out
[25]	Microservice IoT application deployment	Using microservices to design a robust service deployment scheme for Edge-Cloud, with DRL	Service often breaks down during implementation
[26]	Microservice IoT application deployment	Using DRL to solve microservice deployment optimization problems for Edge-Cloud environment	Implementation records slight improvement in performance
[31]	Identification of Cognitive 5G use cases	Using NFV to provide edge network flexibility with AI for optimization	Performance declines due to false negatives and false recall
[32]	Malfunction Detection for NFV networks	Using AI for anomaly detection in NFV-enabled edge networks	Implementation is not robust and would not withstand large-scale evaluation
[33]	Dynamic identification and selection of VNF	Using AI for routing optimization in NFV-enabled edge network	Performance declines when service requests increase

Traditional IoT applications on a large scale are difficult to maintain, restructure and extend to other use cases. IoT needs to incorporate architecture which is flexible, interoperable, heterogeneous and robust against the harsh environment. In [23], the authors propose an open IoT microservice framework which uses device plugins and third-party service plugins to encourage other applications to be introduced. However, the paper does not sufficiently consider AI applications. Advancing the earlier work done, Ali et al [24] incorporate microservices, ML algorithms and Virtual Objects into the IoT framework. The microservices grant the flexibility of the application, whiles the ML algorithms render predictive performance to the applications and Virtual Objects extending the reach of the IoT network. The performance of this scheme showed low query time from Virtual Objects and a low error rate in activity prediction accuracy. Another direction of research in this area has been to employ microservices with Deep Learning techniques like DRL to the problem of optimization for IoT networks at the Edge. In [25], Chen et al show a decrease in system performance in terms of user service average waiting time after DRL is employed in Edge-Cloud networks. About the baseline algorithms, genetic algorithm and random algorithm, the DRL algorithm achieves 32% and 44% better service waiting time respectively. Also, Debauche et al [26] show that the optimization problem needs to be solved in microservice deployment optimization, with pattern recognition from neural networks and the results attained show a 5% reduction in latency and the corresponding increase in service time deployment by 28 seconds. Researchers have applied microservices in various Intelligent Edge application deployment scenarios. For example, microservices and Deep Learning (DL) are applied in environmental monitoring at the edge of the network [27]. Also, microservices are applied in Intelligent Edge Computing for industrial processes [28]. Wide-scale adoption of microservices is hindered by potential drawbacks in microservices such as the time-consuming nature of developing microservices because of their complexity. Also, microservices rely predominantly on communication traffic, which tends to create a significant overhead than expected during the peak demand for services.

C. *Network Function Virtualization.*

Network Function Virtualization (NFV) allows distributed and otherwise proprietary hardware in the form of servers, switches and routers to perform virtual network functions (VNF) [29]. NFV takes advantage of virtualization technologies like Virtual Machines and Containers to deploy VNFs. NFV supports functions of services that change frequently, such as switching functions, tunnelling functions, service assurance functions, converged functions, application functions and security functions [30]. These service functions are defined as VNF and implemented on servers. NFV offers the advantages such as hardware agnosticism and service orchestration among others.

To provide a general overview of NFV and AI operation for Intelligent Edge Computing, it is important to note that the edge computing environment is not well designed to deal with dynamic function requests under virtualization for mobile users since there are privacy concerns when it comes to the vulnerability of VM systems and the sizes of VM images which exceed the minimum memory capacity of most end devices [30]. NFV can meet these challenges. To deal with privacy concerns, using NFV standards such as security management and monitoring data breaches can be detected quickly and addressed. Also by using compression techniques available in the NFV network such as gzip or bzip2 the sizes of VM images can be reduced. Next-generation networks envision the connection of whole continents to the internet. The problem is how to manage such a massive network. AI is the option considered by Yahia et al [31] to achieve such an ambition with NFV providing flexibility of the intelligent network. The work shows that Deep Learning methods such as LSTM and RNN achieve bandwidth prediction values better than the baseline ML algorithms such as Decision trees. Ahrens et al. deployed a self-supervised neural network associated with SDN technology to detect anomalies such as bugs and software update errors in VNF systems [32]. The system predicts anomalies based on CPU and memory utilization metrics. However, the algorithm is not robust and would not withstand large-scale implementation. In [33], the authors proposed a DL technique called Deep Belief Networks to achieve the best routing performance. Deep Belief Network learns the optimal routing scheme by feature extraction and classification thus achieving decent network traffic prediction for NFV orchestration. However, there is a performance decline when the service requests increase due to prediction errors. Deep Learning can be applied in the context of NFV to model NFV placement problems considering the reliability requirement of the services. Some of the problems of NFV in this domain include the complexity and difficulty of deploying NFV at scale. The breadth of the architecture and the number of distinct components make it challenging to design, build and support [34].

IV. FUTURE DIRECTIONS

A. *Security*

The Intelligent Edge computing paradigm incorporates many frameworks avoiding a centralized administration. The network infrastructure and virtual services infrastructure are all independent, however, they collaborate to ensure user satisfaction. These distributed frameworks are areas malicious agents can exploit to launch attacks against the network. A security feature that offers global protection over the whole Intelligent Edge infrastructure would be preferred over individual security schemes which would present many points of failure. The entire perimeter of the network needs to be protected to ensure adversaries are kept away from crucial services. With the collaboration of different paradigms, a single vulnerability can lead to the exploitation of the entire network. Many attempts are being made to employ SDN, and DRL due to the nature of these two technologies to provide a global optimum solution to network monitoring and routing, enhancing security in networks by first identifying what patterns in network traffic are normal and then escalating abnormal cases for inspection. By doing so popular malware and exploitation hacks can be avoided [8], [35].

B. *Privacy*

Organizations such as the European Union are taking steps to ensure that user data is only used with user authorization, in response to recorded cases of data manipulation and infringement [36]. Current efforts are trying to make procedures as transparent as possible avoiding black boxes and misleading policies. Attempts are being made to uncover what exactly happens in hidden layers of deep neural networks [13]. Researchers are also developing new paradigms by incorporating current enabling techniques [1]. Two such paradigms are Privacy by Design (PbD) and Software Defined Privacy (SDP). PbD is a scheme that uses microservices to augment the whole system's security as a Service (SaaS). Privacy by Design (PbD) as a data privacy concept calls for the incorporation of data privacy protections into the design of information systems, products, and services. PbD aims to prevent data privacy breaches and protect the privacy of individuals by proactively incorporating data privacy safeguards into systems and processes. PbD is based on some principles: proactive not reactive; privacy as the default setting; privacy embedded into the design; full functionality – positive-sum, not zero-sum; end-to-end security – full lifecycle protection. Software Defined Privacy (SDP) seeks to modify SDN to offer users enhanced privacy based on transparent policies. Software Defined Privacy is built to protect the privacy of its users by controlling or limiting the amount of information made available to third parties. The software can apply encryption or filtering of various kinds.

C. Standardization

Industry players such as service providers, researchers, network operators, and other stakeholders are providing input to form standards for Intelligent Edge Computing, to establish a trustworthy paradigm. The terms Intelligent Edge and Edge Intelligence are used loosely and interchangeably. According to Wang et al [21], Intelligent Edge is used to describe techniques used to incorporate DL into Edge Computing whiles Edge Intelligence is attaining AI-enabled Edge Computing infrastructure which is independent of Cloud computing for reliability, scalability and flexibility. However, there is no universally accepted definition for Intelligent Edge Computing. On the other hand, in 2017 European Telecommunications Standards Institute (ETSI) led to the change of MEC is Mobile Edge Computing to Multi-Access Edge Computing to recognize the wireless networks and enhanced user mobility introduced into mobile computing. The term is now widely used in technology under cellular networking. Other paradigms like Cloudlet and Fog Computing are employed and governed mainly by private corporations. Universal standards are needed to enhance product and service integrity.

V. CONCLUSIONS

Due to the growth of AI and Edge computing, it is more possible to bring complex AI applications into the edge computing volatile environment. This makes Intelligent Edge feasible with current trends. Virtualization and Network Function Virtualization are thus introduced to facilitate the realization of an efficient Intelligent Edge. Also in this literature review, consideration is given to the several approaches that have been used to achieve virtual service deployment in the Intelligent Edge and in what direction current research is driving the Intelligent Edge.

REFERENCES

[1] P. Porambage, J. Okwuibe, M. Liyanage, M. Ylianttila, and T. Taleb, "Survey on Multi-Access Edge Computing for Internet of Things Realization," IEEE Communications Surveys and Tutorials, vol. 20, no. 4. Institute of Electrical and Electronics Engineers Inc., pp. 2961–2991, Oct. 01, 2018. doi: 10.1109/COMST.2018.2849509.

[2] P. Wei et al., "Reinforcement Learning-Empowered Mobile Edge Computing for 6G Edge Intelligence," IEEE Access, vol. 10. Institute of Electrical and Electronics Engineers Inc., pp. 65156–65192, 2022. doi: 10.1109/ACCESS.2022.3183647.

[3] "Ian Goodfellow, Yoshua Bengio, Aaron Courville - Deep Learning [pre-pub version]-MIT Press (2016)".

[4] Z. Zhou, X. Chen, E. Li, L. Zeng, K. Luo, and J. Zhang, "Edge Intelligence: Paving the Last Mile of Artificial Intelligence With Edge Computing," Proceedings of the IEEE, 2019, doi: 10.1109/JPROC.2019.2918951.

[5] E. Harjula et al., "Decentralized Iot edge nanoservice architecture for future gadget-free computing," IEEE Access, vol. 7, pp. 119856–119872, 2019, doi: 10.1109/ACCESS.2019.2936714.

[6] H. Zhu, Y. Cao, W. Wang, T. Jiang, and S. Jin, "Deep Reinforcement Learning for Mobile Edge Caching: Review, New Features, and Open Issues," IEEE Netw, vol. 32, no. 6, pp. 50–57, Nov. 2018, doi: 10.1109/MNET.2018.1800109.

[7] T. le Duc, R. G. Leiva, P. Casari, and P. O. Östberg, "Machine learning methods for reliable resource provisioning in edge-cloud computing: A survey," ACM Comput Surv, vol. 52, no. 5, Sep. 2019, doi: 10.1145/3341145.

[8] M. Mukherjee, R. Matam, C. X. Mavromoustakis, H. Jiang, G. Mastorakis, and M. Guo, "Intelligent Edge Computing: Security and Privacy Challenges," IEEE Communications Magazine, vol. 58, no. 9, pp. 26–31, Oct. 2020, doi: 10.1109/MCOM.001.2000297.

[9] H. Tabatabaee Malazi et al., "Dynamic Service Placement in Multi-Access Edge Computing: A Systematic Literature Review," IEEE Access, vol. 10, pp. 32639–32688, 2022, doi: 10.1109/ACCESS.2022.3160738.

[10] J. Xie et al., "A survey of machine learning techniques applied to software-defined networking (SDN): Research issues and challenges," IEEE Communications Surveys and Tutorials, vol. 21, no. 1. Institute of Electrical and Electronics Engineers Inc., pp. 393–430, Jan. 01, 2019. doi: 10.1109/COMST.2018.2866942.

[11] S. Bansal and D. Kumar, "IoT Ecosystem: A Survey on Devices, Gateways, Operating Systems, Middleware and Communication," Int J Wirel Inf Netw, vol. 27, no. 3, pp. 340–364, Oct. 2020, doi: 10.1007/s10776-020-00483-7.

[12] Y. Li, X. Li, and O. Yoshie, "Traffic engineering framework with machine learning based meta-layer in software-defined networks," in Proceedings of 2014 4th IEEE International Conference on Network Infrastructure and Digital Content, IEEE IC-NIDC 2014, Dec. 2014, pp. 121–125. doi: 10.1109/ICNIDC.2014.7000278.

[13] U. di Bologna, I. C. Society, I. S. D. N. Initiative, I. of Electrical, and E. Engineers, 2017 IEEE Conference on Network Softwarization : July 3-7, 2017, Bologna, Italy.

[14] C. X. Cui and Y. bin Xu, "Research on load balance method in SDN," International Journal of Grid and Distributed Computing, vol. 9, no. 1, pp. 25–36, 2016, doi: 10.14257/ijgdc.2016.9.1.03.

[15] G. Stampa, M. Arias, D. Sanchez-Charles, V. Muntes-Mulero, and A. Cabellos, "A Deep-Reinforcement Learning Approach for Software-Defined Networking Routing Optimization," Sep. 2017, [Online]. Available: http://arxiv.org/abs/1709.07080

[16] A. Mestres et al., "NeuRoute: Predictive Dynamic Routing for Software-Defined Networks," Computer Communication Review, vol. 47, no. 3, pp. 1–10, Jul. 2017, doi: 10.1145/3138808.3138810.

[17] Y. He, F. R. Yu, N. Zhao, H. Yin, and A. Boukerche, "Deep reinforcement learning (DRL)-based resource management in software-defined and virtualized vehicular ad hoc networks," in DIVANet 2017 - Proceedings of the 6th ACM Symposium on Development and Analysis of Intelligent Vehicular Networks and Applications, Co-located with MSWiM 2017, Nov. 2017, pp. 47–54. doi: 10.1145/3132340.3132355.

[18] Y. He, F. Richard Yu, N. Zhao, V. C. M. Leung, and H. Yin, "Software-Defined Networks with Mobile Edge Computing and Caching for Smart Cities: A Big Data Deep Reinforcement Learning Approach," IEEE Communications Magazine, vol. 55, no. 12, pp. 31–37, Dec. 2017, doi: 10.1109/MCOM.2017.1700246.

[19] A. Nakao and P. Du, "Toward in-network deep machine learning for identifying mobile applications and enabling application specific network slicing," IEICE Transactions on Communications, vol. E101B, no. 7, pp. 1536–1543, Jul. 2018, doi: 10.1587/transcom.2017CQI0002.

[20] F. Al-Doghman, N. Moustafa, I. Khalil, Z. Tari, and A. Zomaya, "AI-enabled Secure Microservices in Edge Computing: Opportunities and Challenges," IEEE Trans Serv Comput, 2022, doi: 10.1109/TSC.2022.3155447.

[21] X. Wang, Y. Han, V. C. M. Leung, D. Niyato, X. Yan, and X. Chen, "Convergence of Edge Computing and Deep Learning: A Comprehensive Survey," IEEE Communications Surveys and Tutorials, vol. 22, no. 2. Institute of Electrical and Electronics Engineers Inc., pp. 869–904, Oct. 2020. doi: 10.1109/COMST.2020.2970550.

[22] J. Chen and X. Ran, "Deep Learning With Edge Computing: A Review," Proceedings of the IEEE, 2019, doi: 10.1109/JPROC.2019.2921977.

[23] L. Sun, Y. Li, and R. A. Memon, "An open IoT framework based on microservices architecture," China Communications, vol. 14, no. 2, pp. 154–162, Feb. 2017, doi: 10.1109/CC.2017.7868163.

[24] S. Ali, M. A. Jarwar, and I. Chong, "Design methodology of microservices to support predictive analytics for IoT applications," Sensors (Switzerland), vol. 18, no. 12, Dec. 2018, doi: 10.3390/s18124226.

[25] L. Chen et al., “IoT Microservice Deployment in Edge-Cloud Hybrid Environment Using Reinforcement Learning,” IEEE Internet Things J, vol. 8, no. 16, pp. 12610–12622, Aug. 2021, doi: 10.1109/JIOT.2020.3014970.

[26] O. Debauche, S. Mahmoudi, S. A. Mahmoudi, P. Manneback, and F. Lebeau, “A new edge architecture for AI-IoT services deployment,” in Procedia Computer Science, 2020, vol. 175, pp. 10–19. doi: 10.1016/j.procs.2020.07.006.

[27] J. Huang, Y. Liu, and N. Lu, “Research on Environmental Monitoring System Based on Microservices and Data Mining”, doi: 10.1051/e3sconf/20.

[28] L. Rychener, F. Montet, and J. Hennebert, “Architecture Proposal for Machine Learning Based Industrial Process Monitoring,” in Procedia Computer Science, 2020, vol. 170, pp. 648–655. doi: 10.1016/j.procs.2020.03.137.

[29] Y. Y. Shih, H. P. Lin, A. C. Pang, C. C. Chuang, and C. T. Chou, “An NFV-Based Service Framework for IoT Applications in Edge Computing Environments,” in IEEE Transactions on Network and Service Management, Dec. 2019, vol. 16, no. 4, pp. 1419–1434. doi: 10.1109/TNSM.2019.2948764.

[30] S. R. Chaudhry, A. Palade, A. Kazmi, and S. Clarke, “Improved QoS at the Edge Using Serverless Computing to Deploy Virtual Network Functions,” IEEE Internet Things J, vol. 7, no. 10, pp. 10673–10683, Oct. 2020, doi: 10.1109/JIOT.2020.3011057.

[31] I. G. ben Yahia, J. Bendriss, A. Samba, and P. Dooze, “CogNitive 5G networks: Comprehensive operator use cases with machine learning for management operations,” in Proceedings of the 2017 20th Conference on Innovations in Clouds, Internet and Networks, ICIN 2017, Apr. 2017, pp. 252–259. doi: 10.1109/ICIN.2017.7899421.

[32] J. Ahrens, M. Strufe, L. Ahrens, and H. D. Schotten, “An AI-driven Malfunction Detection Concept for NFV Instances in 5G,” Apr. 2018, [Online]. Available: http://arxiv.org/abs/1804.05796

[33] J. Pei, P. Hong, and D. Li, “Virtual network function selection and chaining based on deep learning in SDN and NFV-Enabled Networks,” in 2018 IEEE International Conference on Communications Workshops, ICC Workshops 2018 - Proceedings, Jul. 2018, pp. 1–6. doi: 10.1109/ICCW.2018.8403657.

[34] R. S. Alonso, J. Prieto, F. De La Prieta, S. Rodriguez-Gonzalez, and J. M. Corchado, “A Review on Deep Reinforcement Learning for the management of SDN and NFV in Edge-IoT,” in 2021 IEEE Globecom Workshops, GC Wkshps 2021 - Proceedings, 2021. doi: 10.1109/GCWkshps52748.2021.9682179.

[35] Y. Xiao, Y. Jia, C. Liu, X. Cheng, J. Yu, and W. Lv, “Edge Computing Security: State of the Art and Challenges,” Proceedings of the IEEE, 2019, doi: 10.1109/JPROC.2019.2918437.

[36] W. Shi, J. Cao, Q. Zhang, Y. Li, and L. Xu, “Edge Computing: Vision and Challenges,” IEEE Internet Things J, vol. 3, no. 5, pp. 637–646, Oct. 2016, doi: 10.1109/JIOT.2016.2579198.