Abstract—Edge computing that leverages cloud resources to the proximity of user devices is seen as the future infrastructure for distributed applications. However, developing and deploying edge applications, that rely on cellular networks, is burdensome. Such network infrastructures are often based on proprietary components, each with unique programming abstractions and interfaces. To facilitate straightforward deployment of edge applications, we introduce open-source software (OSS) based radio access network (RAN) on over-the-air (OTA) commercial spectrum with Development Operations (DevOps) capabilities. OSS allows software modifications and integrations of the system components, e.g., Evolved Packet Core (EPC) and edge hosts running applications, required for new data pipelines and optimizations not addressed in standardization. Such an OSS infrastructure enables further research and prototyping of novel end-user applications in an environment familiar to software engineers without telecommunications background. We evaluated the presented infrastructure with end-to-end (E2E) OTA testing, resulting in 7.5MB/s throughput and latency of 21ms, which shows that the presented infrastructure provides low latency for edge applications.

I. INTRODUCTION

Edge computing is a prominent part of current and future cellular networks. It provides cloud resources, e.g., application-specific virtualization of computation and data, in the proximity of end-users. Virtualization at the edge decouples software from hardware, e.g., based on containers, that enables a variety of software-based deployment and coordination scenarios atop common hardware. As a coherent system, this promises application performance with more bandwidth and reduced latency for user equipments (UEs) accessing the edge resources, while improving controls on privacy.

The ETSI multi-access edge computing (MEC) standards aim to deliver these promises for the local RAN atop Fifth Generation (5G) networks, allowing effective use of the up-and-coming 5G radio features. At the same time, software-defined network (SDN) and network function virtualization (NFV) coupled RANs start to resemble edge platforms, as virtualization facilitates mobile network operators (MNOs) building RAN features on general-purpose hardware. This can address several scaling problems of RANs [1], including dynamic network resource allocation, and network hardware ossification.

However, edge application development and deployment are currently challenging in RANs. First, RANs, with or without MEC, have challenging requirements on E2E latency, energy consumption and reliability with both up and downlinks [2, 3]. Second, edge applications are envisioned as both data- and computation-intensive. Thus, application deployment is not straightforward concerning the capabilities of edge hosts, application requirements, and environmental parameters of the RAN. Only the consideration of all these aspects makes envisioned zero-latency applications possible. These issues are aggregated with multiple parallel applications on the same infrastructure.

Further, the current MEC standardization offers less application programming interface (API) endpoints for software development than existing cloud computing environments. Therefore, linking development and actual deployment may be hard for end-user application developers, more familiar with current cloud environments. Coincidentally, MNOs may be unable to introduce the required APIs for developers, for the RAN components being proprietary. However, if the RAN would be based on OSS then MNO could create the required bridging interfaces. Using existing mature OSS development platforms, which support the current mainstream software engineering processes, the challenges mentioned above could be thus addressed. In this context, we present an OSS infrastructure for Software-based 5G RAN development, where DevOps simplifies the deployment of E2E OTA applications to the edge, that also satisfies the attributes of a micro-operator, as discussed in [4].

The paper is organized as follows. Section II discusses the related paradigms in software and telecommunications research. Section III details the infrastructure design. Section IV evaluates the proposed infrastructure to deploy the network OTA with RAN hardware and spectrum licenses provided by the University of Oulu. Discussion follows in Section V and lastly the contribution is concluded in Section VI.

II. BACKGROUND

IoT and edge computing necessitate application development for a distributed environment, where virtualization provides a deployment mechanism. In 5G virtualization is realized through SDN and NFV technologies. Combined, these technologies pave the way for the microservice architectural model [5], which can with DevOps practices speed up software development for the edge.
A. Virtualization

SDN and NFV use virtualization to decouple application control and data into separate planes and to provide programmability and flexibility to network management through software. This approach supports adjusting the edge infrastructure and the network to user demands while considering the locally available resources.

To realize virtualization on general-purpose hardware, containers have been reported to produce less operational overhead and better performance \[6, 7\] than virtual machines that are based on virtualized operating system layer instead of hardware layer \[8\].

From a developer’s perspective, managing containers means managing applications rather than machines. This enables the same deployment environment reducing inconsistencies between development and production, which improves deployment reliability and speed.

For MNOs, containers provide an isolation mechanism which avoids edge applications from interfering with the host computer’s, or other edge applications’ execution. As the host computers solely manage containers, limiting operating system versions on edge hosts is easier, making the infrastructure more maintainable \[9\]. Container deployments spanning many physical host computers can then be managed with an orchestrator, which automates container creation, removal, restoration, and updating without interfering with the availability of the RAN or the end-user edge applications.

B. Microservices and -operators

Microservices offer a distributed, cloud-native architectural model, which refines and simplifies service-oriented architecture \[3, 6\]. For example, microservices follow the share-nothing philosophy, to support agile business methods and to promote software isolation and autonomy \[10\]. Microservices may be based on various programming languages, middleware and data stores. Each service runs in isolation while communicating through lightweight APIs. The microservice model is seen to provide isolation and scalability elastically and platform-agnostically. However, stable functional decomposition of the application into services is required with supporting system services, e.g., for service discovery, system configuration, load balancing, and fault detection.

In general, microservice development requires knowledge of distributed systems deployment, e.g., the tradeoffs between communication and computation overheads to realize the expected benefits \[5, 11\]. This is further challenging with MEC which generally aims to optimize the connection between UEs and MEC services.

From the MNOs perspective, predicting the required computational and communication capabilities is challenging for a single site of operation in a traditional RAN. A distributed MEC platform adds extra variables into the challenge, such as application-specific resource migration within and between sites. Therefore, facilities for testing, deploying, and monitoring microservices becomes crucial in a cellular network due to multiple sites with varying edge computing capabilities.

In dense cellular deployments, MNOs also face challenges in allocating radio resources with limited spectrum, managing interference, and in load-balancing cells. Local geographical MNOs may thus be needed to manage wireless connectivity for UEs effectively while maximizing throughput, latency, connectivity, or any other goal \[12, 13, 14\]. Further, MNOs infrastructure is challenged by ever-increasing throughput and latency requirements of the 5G applications. If resources on wider area networks need to be accessed, latencies will grow, thus on-site data centers are a viable choice for edge applications on 5G. As such, micro-operators could be the first MNOs relying solely on SDN and NFV principles, due to lack of legacy hardware and legacy application requirements from previous generations of cellular connectivity. Combined with agile business perspectives of microservices, such as the ability to cut long-term commitment and bottlenecks by realizing software as a set of complete, cohesive, and coupled services, micro-operators could thus provide an edge platform with customizability, low latency, and high reliability, that is unrealistic for traditional MNOs.

C. DevOps

DevOps paradigm provides good software quality and frequent delivery mechanisms, for development teams of any size, facilitating enhanced value creation for end-users \[13\]. DevOps allows software platforms to be offered as a Platform-as-a-Service (PaaS) for E2E application development. MEC applications could be independently developed and tested without the intervention of the MNO.

DevOps utilizes continuous deployment (CD) and continuous integration (CI) and continuous monitoring (CM) \[6\]. When the number of deployments increases, as with microservices, CD and CI become essential for frequent software releases. CM then provides microservice developers and the MNO with performance-related feedback of the releases, e.g., to detect operational anomalies, and profiling.

D. Related work

Virtualized cellular networks without OTA testing have been studied widely \[7, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25\]. OTA testbeds without MEC have been installed on LTE \[26, 27, 28, 29\] and, to avoid regulated spectrum, on WiFi \[30, 31\]. Proprietary OTA testbeds with MEC exist \[32, 33, 34, 35\], but without consideration of DevOps practices. In general, DevOps practices have rarely been installed on Long Term Evolution (LTE) but one study did provide Infrastructure-as-a-Service (IaaS) services \[36\] making such deployments possible, and another one has offered DevOps \[37\] on WiFi. As such, our study essentially combines the OTA environment of \[36\], but replaces IaaS as in \[37\] with PaaS thus relieving the developer to only maintain an application instead of a server, as described in \[II-A\] To the best of our knowledge the research in this niche is yet uncharted.
Development for our needs, we can replace it more cost-effectively than a provider. Should a DevOps component prove to be unsuitable of the testbed on OSS, avoiding dependency on a single tool DevOps tools on Kubernetes we base the initial configuration (e.g., Prometheus) are easy to integrate. By installing these control (e.g., Gitea), CI and CD (e.g., Drone), and CM ware engineering, open-source DevOps services like revision bernetes. Thanks to industry adoption of Kubernetes in soft-where DevOps services are installed on MEC hosts on Ku-
possible with minimized hardware requirements.

Application components; thus widespread deployments become
Docker provides lightweight virtualization for distributed edge
for microservice development in the proposed infrastructure.
for container-based application deployment, which we used
separating developer applications from the actual deploy-
in isolation, prioritization, and scheduling mechanisms for
As an example, current off-the-shelf software-deﬁned radio
(SDR) hardware allows building cellular deployments on virtualization and OSS [27], but the performance and features are underdeveloped.

In the corporate market, Kubernetes has emerged as the leading solution for development deployments for the edge [38], [9]. Moreover, microservice infrastructures have been promoted by major cloud computing infrastructure providers, such as Google, Microsoft, and Alibaba, with varying capabilities for management.

As we set the requirement to use the same infrastructure in both development and deployment, we built our platform with Kubernetes. In this regard, Kubernetes provides built-in isolation, prioritization, and scheduling mechanisms for separating developer applications from the actual deployments on the same infrastructure. Kubernetes supports Docker for container-based application deployment, which we used for microservice development in the proposed infrastructure. Docker provides lightweight virtualization for distributed edge application components; thus widespread deployments become possible with minimized hardware requirements.

Fig. 1 presents an overview of the proposed workflow, where DevOps services are installed on MEC hosts on Kubernetes. Thanks to industry adoption of Kubernetes in software engineering, open-source DevOps services like revision control (e.g., Gitea), CI and CD (e.g., Drone), and CM (e.g., Prometheus) are easy to integrate. By installing these DevOps tools on Kubernetes we base the initial configuration of the testbed on OSS, avoiding dependency on a single tool provider. Should a DevOps component prove to be unsuitable for our needs, we can replace it more cost-effectively than a proprietary cloud-based service. As a result, the unity of these components contains most of the functionality specified in the proprietary cloud-based service. As a result, the unity of these components contains most of the functionality specified in the MEC reference architecture [39] (Fig. 6-1). The reference architecture is further described in [40] (§V.B), with which we can translate the ETSI standard to Kubernetes terminology: APP is a deployment in Kubernetes, VIM is a container registry, CFS is kubectl interfacing through Web UI, and LCM is kubectl interfacing through JSON API.

New E2E edge applications are deployed as follows (see Fig. 1): a developer places application source-code in a git-repository. Once the developer pushes changes to the master branch of the repository (1), an automatic post-commit git hook pulls the source-code to the CI server (2), which runs the developer-specified tests (3), and if successful, attempts to build the application using its Dockerfile. If successful, the CI creates a Docker image, passes it to CD (4), which saves it (5), and commands Kubernetes to deploy the application using the built Docker image (6, 7). If Kubernetes is able to provision the application (8), the CD server returns the application’s DNS address from Kubernetes (9), which is then passed onto the developer to be used to access the application from an UE.

Fig. 2 shows the software stacks of the primary host running Kubernetes master and the EPC and the edge hosts meant solely for hosting the edge applications. The primary host is built on Arch Linux, as it provides straightforward installation of a real-time Linux kernel. The edge hosts operate on CoreOS, that runs containers natively, and be used to form a cluster of bare-metal servers. The clusterization allows the CoreOS hosts to update the container runtime and kernel components automatically and in succession, resulting in zero-downtime operating system upgrades [41]. Each physical host then runs the standard Kubernetes architecture, e.g., flannel as a CNI (Container Network Interface) for establishing a virtual network that attaches IPs for containers, Kube Proxy for routing between pods, and Kubelet as orchestration agent responsible for the control interface.

### A. Network architecture

Fig. 3 shows the network architecture of our infrastructure. UEs are assigned an IP from the `tun/tap` device of the EPC.
Fig. 2: Software stacks of the edge hosts. Left: the primary host running Kubernetes master and the EPC. Right: the edge hosts hosting the edge applications.

As a part of the Kubernetes cluster, the EPC can route UEs to access the rest of the Kubernetes deployment, on which the load-balanced edge applications can run in a single network hop from the EPC facilitating guarantees on low latency.

The EPC implementation must support low latency edge applications on the RANs. Thus a real-time Linux kernel is a requirement to optimize its network stack [7]. As discussed in [42], EPC should be left unvirtualized for NFV management. Further, we installed kernel passthrough for the network interface controller (NIC) [7], [42]. As such, the EPC is a scalable software component instead of a hard to upgrade proprietary digital signal processor (DSP) hardware.

In this respect, we reviewed the current OSS EPC projects, as shown in Table I. We chose NextEPC as our environment for its ease of installation and support for 3rd Generation Partnership Project (3GPP) Release 13. Besides having the most up-to-date 3GPP conformance, NextEPC is also the only EPC installable from a package manager and with an easy-to-use web-based subscriber management interface.

TABLE I: Currently available OSS EPC packages.

| Project     | Language | Anecdotal summary               |
|-------------|----------|---------------------------------|
| NextEPC     | C        | 3GPP Release 13 support         |
| corenet     | Python   | Minimal 3G and LTE EPC          |
| vEPC        | C++      | For performance testing [7]     |
| openair-cn  | C        | 3GPP Release 10 support [43]    |
| srsLTE      | C++      | 3GPP Release 8 support          |
| openLTE     | C++      | Elegant, but incomplete [45]    |

IV. EVALUATION AND RESULTS

We evaluated the proposed infrastructure with regard to communication latency between one UE and the EPC as infrastructure end-point, that is the focus of the edge paradigm. The evaluations are conducted using Nokia picocell as a base station, set at 10Mhz on band 7, using spectrum licenses of the University of Oulu. The EPC is installed on commodity hardware running i3-6100. The results establish a baseline on which to set the performance expectations with application-specific virtualization.

We measured latencies in four data sets, as shown in Table II and Fig. 4. The first set served an empty web page. The rest of the tests served pre-generated binary blogs generated by `/dev/rand`. The second test served a single 1MB binary file. The third test served a single 10MB binary file, and the fourth served 10MB binary file over WiFi as a baseline. All the tests sent 250 requests over 5 seconds in chunks of 50 requests per second. We used HTTP-based Vegeta for benchmarking.

The detailed datasets, software versions, and hardware component specifications are available on Github. Given the results, further research is inevitable to debug anomalies such as the first request latency of an empty webpage of 65.6ms as shown in Fig. 4. While WiFi does yield better performance, it is worth noting that LTE has wider operation range thus is more suitable for applications requiring high mobility. Further, we note the EPC uses userland implementations of the GPRS Tunneling Protocol (GTP), and Stream Control Transmission Protocol (SCTP), which are the transport protocols LTE communication is based on. As both protocols have Linux kernel implementations, the source code of the EPC could be modified to use the more privileged versions thus facilitating higher performance and lower latency. Moreover, integrating Data Plane Development Kit bindings would provide further network performance gains [7].

TABLE II: Latency and throughput

| Test                  | Mean  | 99th   | Throughput  |
|-----------------------|-------|--------|-------------|
| empty HTML document   | 20.6ms| 45.9ms | 9.38MB/s    |
| 1MB blob              | 33.6s | 40.5s  | 6.72MB/s    |
| 10MB blob             | 332.5s| 336.0s | 7.52MB/s    |
| 10MB blob on WiFi     | 62.5s | 68.7s  | 38.65MB/s   |

1https://github.com/tsenart/vegeta
2https://github.com/toldjuuso/haavisto2019infra
architecture and Open Compute Project network hardware. These implementations could reside on more profound levels of software abstraction, thus resulting in better performance and topics of future research.

Finally, the University of Oulu owns the RAN hardware, and has the regulated spectrum licenses to deploy the network OTA. This enables a future scenario with eSIMs, in which the campus area has an open cellular network for students and faculty members. In such scenario, subscribers could programmatically have SIM credentials created using QR codes on the campus hallways while aspiring developers could deploy software both as RAN infrastructure components, but also as end-user E2E applications accessible over-the-air within the campus periphery.

VI. CONCLUSION

An infrastructure for RAN was presented, where the target is to enable low latency connections for UEs in the context of 5G networks for MEC. Our evaluation shows that the network, at the premises of the University of Oulu, can handle latency of 21ms and bandwidth of 7.5MB/s in E2E OTA testing. These results provide a real-world baseline for further improving such infrastructures towards the goals of 5G and edge computing. The OSS infrastructure allows off-the-spec RAN modifications, this is difficult to realize in propri- etary environments. Our future work addresses improvements concerning the EPC and edge host hardware. Moreover, we envision further real-world OTA application deployments and studies with industry developers, researchers, and students on the presented open cellular network.

VII. ACKNOWLEDGEMENTS

This research is financially supported by Academy of Finland 6Genesis Flagship (grant 318927) and by the AI Enhanced Mobile Edge Computing project, funded by the Future Makers program of Jane and Aatos Erkko Foundation and Technology Industries of Finland Centennial Foundation.

REFERENCES

[1] Y. Li and M. Chen, “Software-defined network function virtualization: A survey,” IEEE Access, vol. 3, pp. 2542–2553, 2015.
[2] M. Condoluci, G. Araniti, T. Mahmoodi, and M. Dohler, “Enabling the iot machine age with 5G: Machine-type multicast services for innovative real-time applications,” IEEE Access, vol. 4, pp. 5555–5569, 2016.
[3] M. Mezzavilla, M. Zhang, M. Polese, R. Ford, S. Dutta, S. Rangan, and M. Zorzi, “End-to-end simulation of 5G mmwave networks,” IEEE Communications Surveys & Tutorials, 2018.
[4] P. Ahokangas, S. Mogaddam, M. Matinmikko, A. Abouzeid, I. Aiko, J. F. Gomes, and M. Iivari, “Future micro operators business models in 5G,” The Business & Management Review, vol. 7, no. 5, p. 143, 2016.
[5] A. Balalaie, A. Heydarnoori, and P. Jamshidi, “Migrating to cloud-native architectures using microservices: an experience report,” in European Conference on Service-Oriented and Cloud Computing, pp. 201–215, Springer, 2015.
[6] M. Amaral, J. Polo, D. Carrera, I. Mohomed, M. Unuvar, and M. Steinder, “Performance evaluation of microservices architectures using containers,” in Network Computing and Applications (NCA), 2015 IEEE 14th International Symposium on, pp. 27–34, IEEE, 2015.
[7] C.-N. Mao, M.-H. Huang, S. Padhy, S.-T. Wang, W.-C. Chung, Y.-C. Chung, and C.-H. Hu, “Minimizing latency of real-time container cloud for software radio access networks,” in Cloud Computing Technology and Science (CloudCom), 2015 IEEE 7th International Conference on, pp. 611–616, IEEE, 2015.

[8] J. Larisch, J. Mickens, and E. Kohler, “Altos: lightweight vms using virtualization-aware managed runtimes,” in Proceedings of the 15th International Conference on Managed Languages & Runtimes, p. 8, ACM, 2018.

[9] B. Burns, B. Grant, D. Oppenheimer, E. Brewer, and J. Wilkes, “Borg, omega, and kubernetes,” Queue, vol. 14, no. 1, p. 10, 2016.

[10] M. Richards, “Microservices vs. service-oriented architecture.” O'Reilly Media, 2015.

[11] N. Dmitry and S.-S. Manfred, “On micro-services architecture,” International Journal of Open Information Technologies, vol. 2, no. 9, 2014.

[12] A. Gudipati, D. Perry, L. E. Li, and S. Katti, “Softron: Software defined radio access network,” in Proceedings of the second ACM SIGCOMM workshop on Hot topics in software defined networking, pp. 25–30, ACM, 2013.

[13] Y. Elkhatib, B. Porter, H. B. Ribeiro, M. F. Zhani, J. Qadir, and M. Peng, S. Yan, K. Zhang, and C. Wang, “Fog-computing-based radio access networks: issues and challenges,” in Proceedings of the ACM SIGMETRICS joint international conference on Measurement and modeling of computer systems, pp. 317–328, ACM, 2011.

[14] Z. Wang, Z. Qian, Q. Xu, Z. Mao, and M. Zhang, “An Untold Story of Middleboxes in Cellular Networks,” SIGCOMM Computer Communication Review, vol. 41, pp. 374–385, Aug. 2011.

[15] T. Killalea, “The hidden dividends of microservices,” Communications of the ACM, vol. 59, no. 8, pp. 42–45, 2016.

[16] C. Vallati, A. Virdis, E. Mingozzi, and G. Stuo, “Exploiting lte d2d communications in m2m fog platforms: Deployment and practical issues,” in 2015 IEEE 2nd World Forum on Internet of Things (WF-IoT), pp. 585–590, IEEE, 2015.

[17] X. Liu, Y. Zheng, Y. Q. Zhang, and C. Wang, “Fog-computing-based radio access networks: issues and placement,” in IEEE Network, vol. 30, no. 4, pp. 46–53, 2016.

[18] Y. Elkhitib, B. Porter, H. B. Ribeiro, M. F. Zhan, J. Qadir, and E. Rivi`ere, “On using micro-clouds to deliver the fog,” arXiv preprint arXiv:1703.08375, 2017.

[19] H.-C. Hsieh, C.-S. Lee, and J.-L. Chen, “Mobile edge computing platform with container-based virtualization technology for rti applications,” Wireless Personal Communications, vol. 102, no. 1, pp. 527–542, 2018.

[20] R. Muñoz, R. Vilalta, N. Yoshikane, R. Casellas, R. Martínez, T. Tsuritani, and I. Morita, “Integration of iot, transport sdn, and edge/cloud computing for dynamic distribution of iot analytics and efficient use of network resources,” Journal of Lightwave Technology, vol. 36, no. 7, pp. 1420–1428, 2018.

[21] K. Amemiya, Y. Akiyama, K. Kobayashi, Y. Inoue, S. Yamamoto, and K. Amemiya, Y. Akiyama, K. Kobayashi, Y. Inoue, S. Yamamoto, and K. Amemiya, Y. Akiyama, K. Kobayashi, Y. Inoue, S. Yamamoto, and K. Amemiya, Y. Akiyama, K. Kobayashi, Y. Inoue, S. Yamamoto, and K. Amemiya, Y. Akiyama, K. Kobayashi, Y. Inoue, S. Yamamoto, “5G slicing with downlink slicing technology,” in 2018 5th IEEE International Conference on Mobile Edge Computing (SEC), IEEE, 2018.

[22] L. Gauthier, “4G Core Network Training.” https://www.coreos.com/docs/2037/coreoscontainer-update-strategies.html. Jan. 2019.

[23] T. K¨am¨ar¨ainen, M. Siekkinen, A. Yl¨a-J¨a¨aski, W. Zhang, and P. Hui, “A measurement study on achieving imperceptible latency in mobile cloud gaming,” in Proceedings of the 8th ACM on Multimedia Systems Conference, pp. 88–99, ACM, 2017.

[24] F. Eckermann, P. Gorczak, and C. Wietfeldt, “tinyLTE: Lightweight, Ad-Hoc Deployable Open Network for Vehicular Communication,” arXiv preprint arXiv:1802.09262, 2018.