Intelligent Ranking for Dynamic Restoration in Next Generation Wireless Networks

Navrati Saxena, Prasham Jain, Abhishek Roy, Harman Jit Singh, Sukhdeep Singh, Madhan Raj Kanagarathinam

Abstract—Emerging 5G and next generation 6G wireless are likely to involve myriads of connectivity, consisting of a huge number of relatively smaller cells providing ultra-dense coverage. Guaranteeing seamless connectivity and service level agreements in such a dense wireless system demands efficient network management and fast service recovery. However, restoration of a wireless network, in terms of maximizing service recovery, typically requires evaluating the service impact of every network element. Unfortunately, unavailability of real-time KPI information, during an outage, enforces most of the existing approaches to rely significantly on context-based manual evaluation. As a consequence, configuring a real-time recovery of the network nodes is almost impossible, thereby resulting in a prolonged outage duration. In this article, we explore deep learning to introduce an intelligent, proactive network recovery management scheme in anticipation of an eminent network outage. Our proposed method introduces a novel utilization-based ranking scheme of different wireless nodes to minimize the service downtime and enable a fast recovery. Efficient prediction of network KPI (Key Performance Index), based on actual wireless data demonstrates up to ~54% improvement in service outage.

I. INTRODUCTION

Next generation wireless networks envision a phenomenal growth in wireless connectivity, fueled by rapid penetration of smart devices, IoT and myriads of mobile applications. Gradual roll out of 5G wireless networks is expected to result in a massive improvement in data rates, latency and user capacity, while providing diverse service, like enhanced mobile broadband (eMBB), ultra-reliable and low-latency communications (uRLLC), and massive machine type communications (mMTC) [1]. However, the introduction of ultra dense 5G wireless networks, with a wide number of heterogeneous cells, makes the network configuration, management and restoration extremely challenging. Unfortunately, present architecture of 5G wireless is not endowed with enough intelligence and flexibility required for efficient network management and recovery. Naturally, wireless evolution towards sixth generation (6G) [2] is likely to adopt an architectural transformation, required to support intelligent coordination of different network elements, for on-demand network management and fast service recovery after any network outage.

Emergence of Cloud Radio Access Networks (C-RAN) was the first step towards this architectural transformation [3]. The concept of C-RAN is based on distributed base stations (BS), where Remote Radio Heads (RRH) are separated from digital Base Band Unit (BBU) by high speed (typically of several Gbps) fiber optical fronthaul cables. Two major doctrines of C-RAN were: (i) centralization and (ii) virtualization of BBU processing. The vRAN basically extends C-RAN concepts one more step by executing the virtualized baseband functions in commodity servers. This makes the implementation of vRAN based on the core principles of Network Function Virtualization (NFV) [4]. Over the last few years, some independent alliances and forums have initiated research on intelligent NFV for agile network management and restoration. Recently, Open Radio Access Network (O-RAN) [5] is formed from C-RAN alliance to introduce openness on top of NFV for network evolution towards 6G wireless.

Once a network outage occurs, the objective is to restore or recover the services as fast as possible. Typically, a major outage is considered when the network services are compromised for a large set of subscribers (e.g. \( \geq 0.000 \)) of subscribers. Unfortunately, restoration in a large-scale network is already proven as NP-hard [5]. Naturally, most of the solutions depend on specific heuristics, using gradual, progressive recovery under uncertain network conditions [6] or some cooperation between underlying networks [7]. Considering a virtualized wireless network and data center availability not always guaranteed, an outage arising from data center error or failure has a non-zero probability. After the outage, restoration of network services might require re-instantiating affected virtual network functions (VNFs). VNFs may additionally require re-configuration, which is generally performed by an Auto-Configuration Server (ACS). A wireless subscriber will experience network service outage until all the associated VNFs are restored. Unfortunately, unavailability of real-time Key Performance Index (KPI) information, compels the network management to mostly rely on some manual evaluation. Existing manual evaluations and recovery process thwart fast and real-time restoration of the network nodes, thus resulting in increased outage duration. This paves the way for increasing popularity of intelligent Self Organizing Networks (SON) [9] among most of the major cellular operators. SON provides significant reduction in both capital expenditure (capex) and operating expenditure (opex) in all the phases of the network life cycle, i.e. planning, deployment, and operation-management. The use of SON is crucial, if not inevitable, to most of the operators, running multi-vendor and multi-layer VNFs, involving an overwhelming number of configuration and optimization parameters.

Understanding intelligence as the harbinger of self-healing in next generation wireless systems (e.g. 6G wireless), in this article, we explore avenues to improve the expected network outage experienced by the subscribers.
optimism from network slicing, introduced in 5G wireless, we believe that embedded intelligence in the Network Slice Management Function (NSSMF) is essential to perform network restoration in a complex wireless system. More specifically, our contributions are:

- Realizing the increasing penetration of cloudification and virtualization of wireless networks, we deviate from existing hardware-centric self-healing mechanisms and explore software-based network restoration for a large-scale cellular network.
- We exploit long short-term memory (LSTM) recurrent neural network (RNN) to learn the traffic patterns of a real-world 4G cellular network, deployed in a densely populated area in India. This learning is used to make intelligent predictions about the impact of each network element in terms of expected traffic and major network KPIs.
- Subsequently, we introduce a rank-based strategy to recover the network element(s) (NE(s)) in the decreasing order of their impact on the subscribers. Note that, any failure in the network, involving wide number of VNFs result in huge losses. Our solution resolves the problems and takes care of network restoration in an efficient manner.
- Implementation results, using actual 4G wireless network traffic, demonstrates that our framework can achieve fast network restoration with minimum user impact.

II. NETWORK VIRTUALIZATION AND SELF ORGANIZATION

Network functions virtualization (NFV) is a network architectural evolution that explores information technology (IT) to virtualize functions of NEs into inter-connected software building blocks for communication services. In 5G wireless and beyond, NFV is expected to enable network slicing - a virtual network architecture, allowing multiple virtual networks to be created atop a shared physical infrastructure. SON has gradually evolved to make automated planning, configuration, management, optimization and healing of mobile access networks in a fast and efficient manner. It has been already ratified within 3GPP Release 8 and subsequent standard specifications. SON enables auto-configuration of new base stations in line with a "plug-and-play" paradigm, as well as periodic self-optimization of existing, operational base stations. Furthermore, self-healing mechanisms are expected to temporarily compensate for a detected equipment outage, while awaiting a more permanent solution.

Fig. 1 shows a schematic diagram of NFV using Centralized SON (cSON), where Network Slice Subnet Management Function (NSSMF) is responsible for VNF operations, like provisioning and instantiation; and NF Management Function (NFMF) provides Configuration Management (CM), Fault Management (FM) and Performance Management (PM) services to the VNFs. PM collects KPI data from VNF(s) and shares it with SON server, FM monitors the VNF(s) and identifies if any VNF(s) are down, whereas CM (ACS) is responsible for remedial action for VNF recovery. Remedial actions might include re-provisioning and re-instantiation of the VNF, thus triggering NFMF to send new VNF provisioning request to NSSMF. One of the major functionalities of SON lies in supporting self-healing [10] of the network elements. Challenges associated with self-healing [11] of the network elements. Challenges associated with self-healing across low-powered small cells [11] has identified the role of mobile user equipment (UE) to achieve a graceful solution. Cell outage classification and estimation of lost traffic is performed to develop efficient outage compensation [12]. Root cause
Fig. 2. Message Flow in Proposed Dynamic Recovery Management

analysis and dimensionality reduction is also recently used for self-healing in complex wireless networks [13]. With the gradual evolution of towards 6G wireless, intelligence is likely be the new tenet in next generation cSON. Naturally, future wireless networks envision efficient learning and prediction for estimating network outage and trigger fast recovery.

Before going into the details of our Deep Learning based network restoration framework, incorporated in cSON, we briefly highlight the major steps involved in any ML/AI assisted solution:

1) The first step is model capability query, performed when the model is going to be executed for the very first time (or updated). These capabilities include hardware processing power, ML engine, and available data sources.

2) The next step is model selection and training, where the ML training host initiates the model training and sends the trained model back to the system for deployment.

3) The ML inference host is then configured with the model description file, and the online data is used for inference. Depending on the outcome of the model inference, the corresponding actions are taken using the related actors. Based on the location of the ML inference and the actors and type of actions different network interfaces are used.

4) Finally, upon monitoring the performance of the model, the inference host feeds back the model performance to the training host for the purpose of model redeployment or model update.

In the remainder of this article, we discuss our new predictive rank-based network restoration mechanism, which explores efficient machine learning techniques to predict user KPIs and perform dynamic ranking of network elements.

III. PREDICTIVE RANK-BASED NETWORK RECOVERY

A wireless subscriber typically experiences network outage until the associated NEs or VNFs are restored. Naturally, the outage duration is sensitive to the order of the restoration of required VNFs. Intuitively, early restoration of network elements, involving larger number of subscribers should result in improved recovery. Unfortunately, during the outage, the NEs might be incapable of measuring the necessary KPIs. Thus, a viable alternative for the network is to rely on efficient learning and prediction of the associated KPIs. This is exactly what efficient machine learning based KPI prediction offers.

In the remaining of this Section we will discuss our proposed KPI prediction and rank-based network restoration method.

A. Machine Learning based KPI Prediction

Fig. 1 also depicts an overview of our LSTM Neural Network, introduced in cSON for learning the temporal traffic patterns inside each network element and predict the expected current utilization during an outage. As shown in Fig. 1 we have used two hidden layers, LSTM1 and LSTM2, with hyperbolic tangent activation function, mean squared error (MSE) loss function, 4 : 1 proportion of training to validation and Adam algorithms of Stochastic Gradient Descent [14]. When a network outage occurs, the system uses the predicted KPI values, in lieu of real time KPI measurements, to restore the services.

In our framework, the major cellular network parameters, such as user’s downlink throughput (IP throughput), number of
Radio Resource Control (RRC) connected users and Downlink Physical Resource Block (DL-PRB) utilization are continuously recorded for each cell in the eNBs of an actual 4G wireless network. The RNN model, mentioned in Fig. 1, is then used to learn the temporal patterns of these parameters and subsequently predict the future KPIs. Any erroneous prediction results in updating the weights of the RNN model, thus resulting in a continuous improvement of the model accuracy. Fig. 2 also points out that our model is comprised of three layers of 168 LSTM units each, for learning a week-long traffic pattern. For training of the RNN model, we have used hourly KPI measurements (DL-PRB Utilization, User’s DL Throughput and number of RNC Connected users), captured across 4989 cells of a popular 4G wireless network of India over 26 days.

B. Rank-based Dynamic Recovery Management

Intuitively it is clear that recovering network elements or VNFs in the decreasing order of their ranks (i.e. recovering better ranked VNFs first) should improve the network outage and proactively recover more subscribers. Fig. 2 demonstrates the overall message flow in our proposed dynamic recovery management. VNF sends the raw Operation Management (OM) data to NFMF. NFMF shares the KPI data with SON server. The recurrent neural network module, in SON server, uses this KPI data for training and performs prediction of future KPI(s). During any network outage, the VNF informs the outage information to the NFMF. Alternatively, NFMF itself can also detect the outage. Interestingly, during the network outage the network elements, e.g. VNF(s) might be incapable of monitoring the necessary KPI(s). At this crucial point, NFMF inquires the SON server about KPI prediction. The recurrent neural network module in SON server performs the KPI prediction and shares the prediction results with NFMF. If the outage affects more than one VNFs, which is a typical scenario in large-scale cellular networks, the NFMF uses the KPI prediction results to rank the individual VNFs.

Depending on the choice of KPI, used for ranking, a higher KPI value may indicate a higher or lower subscriber impact, thus resulting in a better or worse rank respectively. As we have used user’s DL data rate (IP throughput), number of RNC connected users and DL PRB utilization as our KPIs, a higher value indicates higher subscriber impact and better rank. The VNF recovery operations are now queued and processed in the descending order of ranks, determined by the predicted KPI values. Interestingly, multiple KPIs can be simultaneously used for queuing and sorting, by considering them as primary, secondary and tertiary sources. For example, RRC user count can be used as the primary source and DL-PRB utilization can be used as the secondary source for the sorting operation. The NFMF now triggers VNF re-provisioning with NSSMF and NFVO. Subsequently, NFMF initiates recovery by re-configuring the VNFs depending on their individual rankings, i.e. the VNFs having better ranking are recovered prior to the VNFs with lower rankings. As better ranking infers higher subscriber impact, such a rank-based mechanism always results in better performance, compared to the existing recovery management schemes.

IV. Performance Evaluation

We have developed our RNN model using a high-level tensor-flow programming library, Keras for Python 3.7.7. In addition, we developed a Java-based platform to emulate a large-scale network scenario, the virtual network functions and associated management systems. Our platform emulates an outage scenario for 377 eNBs, consisting of 4,989 cells, by sending fault notifications to the management system and measures the recovery performance based on outage duration of each eNB. In our gNB, every eNB is considered as a NE, managed by a single VNF. The training of LSTM-based Deep Learning is performed on a Dell-R740 server, having 80 CPUs, 512GB of RAM, operating with CentOS 7. The RNN model is trained for a period of 26 days, by using actual 4G wireless network statistics, captured hourly, over 4,989 cells. The data is split into a 21 day training set and a 5 day validation set for each of these 4,989 cells. An outage is then emulated at a randomly selected time within the validation period, i.e. the KPIs of the 4,989 emulated cells are matched with the real traffic at the selected time instance. The predicted KPIs are fed into the recovery module, where each KPI is independently used to rank the cells, resulting in three distinct rankings. In order to compare with the baseline, a legacy approach, without any consideration for real time resource utilization and traffic prediction, is considered. Table 1 provides the major parameters used in our simulation.

Fig. 3 shows the comparison of actual KPI (RRC Connected Users) measurements with our LSTM-based KPI prediction module, running in the SON-server. The figure shows that the predicted KPI, in terms of RRC Connected Users, closely follow the actual KPI (RRC Connected Users). The average RRC Connected User based KPI prediction accuracy over the period is around 96%. Similarly, our LSTM-based module also produces more than 90% accuracy for KPI prediction, in terms of both User throughput and DL PRB utilization.

Fig. 4 demonstrates the network recovery trends in terms of estimated number of subscribers for whom network services have been restored during the recovery process. Our method recovers the cells at a uniform rate in the order prescribed by the KPI ranking. The figure points out that the RRC Connected User based ranking provides the best results, followed by DL-

| Network Parameters | Values |
|-------------------|--------|
| 1. Num. of eNBs (NEs) | 377 |
| 2. Num. of cells | 4,989 |

| LSTM Parameters | Values |
|-----------------|--------|
| 3. LSTM Hidden Layers | 3 |
| 4. LSTM units per layer | 168 |
| 5. Batch Size | 48 |
| 6. Num. of Epoch | 300 |
| 7. Activation | Tanh |
| 8. Loss | MSE |
| 9. Validation Split | 0.2 |
| 10. Optimizer | Adam |

1 https://keras.io/api/layers/recurrent_layers/lstm/
PRB Utilization based ranking and IP Throughput based ranking. The proposed strategies continuously outperform traditional approach, until the entire network recovery is completed. It is clear from the figure that both RRC User-based and DL PRB Utilization based rankings recover almost 70% of the subscribers during the first couple of time units, when around 2,000 cells are recovered. On the other hand, during this time the IP Throughput based ranking and legacy approach recovers around 50% and only 35% of subscribers respectively. This is a direct consequence of rank-based recovery, which recovers the network elements in order of particular KPI ranking. It is also apparent that both RRC User based ranking and DL PRB Utilization based rankings are better choice than IP throughput.

Another interesting metric to look into is the average outage duration (or downtime) experienced by the subscribers. Fig. 5 delineates that compared to the legacy approach, RRC User based ranking and DL PRB Utilization based ranking results in almost 54% and 46% improvement in average subscriber downtime. On the other hand, IP throughput based ranking results in a meagre 15% improvement over the legacy approach.

V. CONCLUSION AND FUTURE WORKS

With the advent of 5G wireless, consisting of a myriad of small-size cells, network management, maintenance and restoration or recovery has become significantly more complex. In this article we have proposed a new ML-based dynamic network restoration process. The process applies LSTM Neural Networks to learn and predict the wireless network KPIs, like RRC Connected Users, PRB Utilization and IP Throughput. Successful KPI prediction aids in ranking the network elements, which in turn, improves the network recovery by reducing the average recovery time for wireless subscribers. It should be noted that the gains might vary depending upon the variation of traffic conditions in the wireless network. Higher heterogeneity in service utilization patterns might yield higher gains. One future investigation is to identify an efficient collective ranking scheme, which incorporates simultaneous usage of multiple KPIs. For example, RRC user count can be used as the primary source and IP Throughput can be used as the secondary source for the sorting operation. As a result, if two VNFs (NEs) have the same predicted RRC user count, then the network elements with the higher predicted value for IP Throughput will be prioritized. This approach would in effect, target the more active subscribers and reduce interruption to ongoing usage. This is expected to improve overall Quality of Experience (QoE), perceived by the user. We would like to investigate and quantify the benefits of such a collective scheme in terms of the QoE improvement.

Moreover, one RNN model, made for one particular region, might not give good training and prediction results for all regions. The reason lies in the difference in user patterns or behaviour and surrounding infrastructure. For example, while typical downtown regions have ultra-dense deployment of high rise buildings, residential areas might have relatively lower height buildings. Hence, instead of having one common model for all, we need to optimize ML models based on the patterns and feedback. Considering nation-wide deployment, it is very difficult to maintain different models and their related predictions at one place. It might be better to perform model execution and prediction result storage in distributed servers (dedicated for intelligence). This leads to distributed learning concept, with an inherent trade-off or compromise between optimized learning and CAPEX of distributed learning infras-
structure.

REFERENCES

[1] M. Agiwal, A. Roy and N. Saxena, "Next Generation 5G Wireless Networks: A Comprehensive Survey," IEEE Communications Surveys and Tutorials, vol. 18, no. 3, pp. 1617-1655, 2016.

[2] K. B. Letaief, W. Chen, Y. Shi, J. Zhang and Y. A. Zhang, "The Roadmap to 6G: AI Empowered Wireless Networks," IEEE Communications Magazine, vol. 57, no. 8, pp. 84-90, 2019.

[3] A. Checko et al., "Cloud RAN for Mobile Networks-A Technology Overview," IEEE Communications Surveys and Tutorials, vol. 17, no. 1, pp. 405-426, 2015

[4] A. Laghrissi and T. Taleb, "A Survey on the Placement of Virtual Resources and Virtual Network Functions," IEEE Communications Surveys and Tutorials, vol. 21, no. 2, pp. 1409-1434, 2019.

[5] N. Bartolini, S. Ciavarella, T. F. La Porta, and S. Silvestri, "On Critical Service Recovery After Massive Network Failures", IEEE/ACM Transactions on Networking, Vol. 25, No. 4, 2017.

[6] Tootaghaj, N. Bartolini, H. Khamfroush and T. F. La Porta, "On Progressive Network Recovery From Massive Failures Under Uncertainty, IEEE Transactions on Network and Service Management, Vol. 16, No. 1, 2019.

[7] Y. Nakayama, K. Maruta, T. Tsuchumi and K. Sezaki, "Wired and Wireless Network Cooperation for Wide-Area Quick Disaster Recovery", IEEE Access, Vol. 6, pp. 2410 - 2424.

[8] ORAN Alliance, "O-RAN: Towards an open and smart RAN", 2018. [Online]. Available: https://www.o-ran.org/resources

[9] P. V. Klaine, M. A. Imran, O. Onireti, and R. D. Souza, "A Survey of Machine Learning Techniques Applied to Self-Organizing Cellular Networks", IEEE Communication Surveys and Tutorials, Vol. 19, No. 4, pp. , 2017.

[10] A. Asghar, H. Farooq, and A. Imran, "Self-Healing in Emerging Cellular Networks: Review, Challenges, and Research Directions", IEEE Communication Surveys and Tutorials, Vol. 20, No. 3, pp., 2018.

[11] S. Fortes, A. agual-farcAa, J. A. FernAnd dez-luque, A. G. Martin, and R. Barco, "Context-Aware Self-Healing: User Equipment as the Main Source of Information for Small-Cell Indoor Network", IEEE Vehicular Technology Magazine, Vol. 11, No. 1, 2016

[12] Isabel de la bandera, Pablo Munoz, InMaculada serrano, and raquel barco, "Improving Cell Outage Management through Data Analysis", IEEE Wireless Communications Magazine, Vol. 24, No. 4, pp. 113-119, 2017.

[13] D. Palacios, S. Fortes, I. de-la-Bandera and R. Barco, "Self-Healing Framework for Next-Generation Networks through Dimensionality Reduction", IEEE Communications Magazine, Vol. 56, No. 7, pp. 170-176, 2018.

[14] A. Savitzky and M. J. E. Golay, "Smoothing and Differentiation of Data by Simplified Least Squares Procedures", Analytical Chemistry, Vol. 36, No. 8, pp. 1627-39.

[15] O. K. Tonguz and W. Viriyasitavat, "A Self-Organizing Network Approach to Priority Management at Intersections," IEEE Communications Magazine, Vol. 54, No. 6, pp. 119-127, 2016.

Navrati Saxena (M’2010) is an Assistant Professor in the Computer Science Department of San Jose State University, USA. During 2007-2019 she was an Assistant and Associate Professor in SKKU, Korea and director of Mobile Ubiquitous System Information Center. She received her Ph.D. degree in 2005 from Univ. of Trento Italy. Her research interests include 5G/6G wireless networks, IoT and cloud RAN. She has more than 70 International journal publications and serves in the program committee of many International conferences. Email: navrati@sj-su.edu.

Prasham Jain is currently working as Lead Engineer at Samsung R&D India Bangalore. He received his B.Tech from IIT Kanpur in 2016. His research interest includes 4G/5G, Network Management, SON, Network Virtualization. Email: prasham.jain@samsung.com

Harman Jit Singh is currently working as Lead Engineer at Samsung R&D India Bangalore. He received his B.Tech from NITK in 2016. His research interest includes Anomaly detection for RAN, Network Optimization, SON, Network Virtualization. Email: harman.jit@samsung.com

Sukhdeep Singh is currently working as Senior Chief Engineer at Samsung R&D India Bangalore. He received his Ph.D. from Sungkyunkwan University, South Korea in 2016. His research interest includes 4G/5G RAN system design, O-RAN, vRAN, Cloud Native and TCP/QUIC for 5G/6G cellular networks. Email: Sukh.sandhu@samsung.com

Madhan Raj Kanagarathinam (M’2010) received the BE degree in computer science and engineering from Anna University, Chennai, India, in 2012. He has seven years of working experience in design and development of TCP/IP protocols, Multi-path TCP and UNIX flavored operating systems. Currently, he is working as a Chief Engineer for Samsung R&D Institute India Bangalore. Previously, he worked as an Engineer with Aricent Technology (India) Private Limited. He is the author of 10 articles, and more than 20 inventions. His current research interests include communication and network which include Pre-4G/Beyond 5G, Next Generation Mobile Network, software defined network architecture, 40 transport layer protocols, and cross layer optimization technique. He is a senior member of the ACM. Email: madhan.raj@samsung.com

Abhishek Roy (S’02-M’10-SM’17) is a senior technical manager at MediTek USA. He received his Ph.D. degree in 2010 from Sungkyunkwan University, his M.S. degree in 2002 from the University of Texas at Arlington, and his B.E. degree in 2000 from Jadavpur University, India. He has strong professional skills in 4G/5G/6G wireless system design, New Radio unlicensed, IoT, cloud RAN, network modeling, and simulation. Email: Abhishek.Roy@mediatek.com.