Large-Scale and High-Dimensional Cell Outage Detection in 5G Self-Organizing Networks

Po-Chiang Lin
Department of Electrical Engineering, Yuan Ze University, Taoyuan, Taiwan
Email: pclin@saturn.yzu.edu.tw
Tel: +886-3-4638800 Ext. 7333

***Abstract*—In this paper, we investigate the cell outage detection in Self-Organizing Networks. The purpose of cell outage detection is to automatically detect whether there exist some failures or degradation in the base stations, such that users could not obtain mobile services, or the obtained mobile services do not fulfill their requirements. The cell outage detection in 5G is with great challenge. The deployment of future 5G mobile communication networks would be heterogeneous and ultra-dense. The mobile communication environments are very complicated. They include the multipath transmission, fading, shadowing, interference, and so on. Users' mobility and usage pattern also vary. In such environments, the mobile data would be large-scale and high-dimensional. Traditional small-scale and low-dimensional anomaly detection methods would be unsuitable. Moreover, operational mobile communication networks should be normal almost all the time. Cell outage would be seldom. Therefore, the normal data and anomaly data would be imbalanced. In this paper, we formulate the cell outage detection problem as an anomaly detection problem. We propose an cell outage detection method using the autoencoder, which is a neural network that is trained by unsupervised learning. The network could be trained in advance even when the cell outage data is still not available. Moreover, the autoencoder is also useful for denoising. This proposed method could thus automatically detect the cell outage in complicated and time-varying mobile wireless communication environments. Comprehensive system-level simulations validate the performance of the proposed method.**

***Index Terms*—5G, Self-Organizing Network, Cell Outage Detection, Autoencoder.**

I. Introduction

In the fifth generation of mobile communication systems (5G), various key software technologies, especially those for mobile communication network management, would be adopted [1]. With the evolution of software technologies and the improvement of computation capabilities, the vision of automatic network management combined with machine learning techniques becomes realistic.

The vision of automatic network management is already defined as the self-Organizing Network (SON) [2], which includes Self Configuration, Self Optimization, and Self Healing. The developing trend in 5G mobile communication networks includes the heterogeneous network integration, ultra-dense networks, small cell base stations. Therefore, SON gets more and more important for 5G [3].

In this paper we investigate the cell outage detection (COD) problem, which is one of the critical issues in Self Healing. The objective of cell outage detection is to detect whether there exists any malfunction of degradation in base station(s) which leads to service unavailability or unsatisfactory. The cell outage detection in 5G is a very challenging problem. The deployment of 5G mobile communication networks would be heterogeneous and ultra-dense. The communication environments include multi-path fading, noise, and interference, and thus would be very complicated. User movement and service demand would change rapidly. In such environments, the mobile data would be large-scale and high-dimensional. Traditional small-scale and low-dimensional anomaly detection methods would be unsuitable. Moreover, operational mobile communication networks should be normal almost all the time. Cell outage would be seldom. Therefore, the normal data and anomaly data would be imbalanced. This data imbalance causes challenges to traditional classification techniques using supervised learning.

In this paper, we formulate the cell outage detection problem as an anomaly detection problem. We propose a cell outage detection method using the autoencoder, which is a neural network that is trained by unsupervised learning. The data for the network training comes from the Reference Signals Received Power (RSRP) and Reference Signals Received Quality (RSRQ) values from the measurement reports. The network could be trained in advance even when the cell outage data is still not available. Moreover, the autoencoder is also useful for denoising. This proposed method could thus automatically detect the cell outage in complicated and time-varying mobile wireless communication environments. Comprehensive system-level simulations validate the performance of the proposed method.

The major contributions of this paper are threefold:

1) We design a novel cell outage detection method called the Cell Outage Detection with Autoencoder (CODA) method. This method is unsupervised and thus avoid the data labeling work. The network could be trained in advance even when the cell outage data is still not available.
2) The proposed CODA method only uses the RSRP and RSRQ values of neighboring cells in the measurement report from mobile stations. It does not require the location information of mobile stations, and thus avoid the privacy and location accuracy issues.
3) The proposed CODA method is distributed and thus scalable to large-scale mobile service regions. The com-

putation of the proposed CODA method could be performed distributively in a cluster of several base stations or some mobile edge computing equipment.

The rest of this paper is organized as follows. In Section II we describe the related work in the literature. The system model and problem formulation are presented in Section III. The proposed CODA method is described in Section IV, followed by the simulation results and discussions in Section V. Finally, conclusions are presented in Section VI.

II. Related Work

Klaine et.al. provided a comprehensive literature survey for applying machine learning techniques in Self-Organizing Networks [4]. Moysen and Giupponi also provided the literature survey and analysis of similar issues [5]. They all pointed that the deep learning would be the trend of future Self-Organizing Networks, and there exist many challenges to solve.

The 3rd Generation Partnership Project (3GPP) defined the specification for the Minimization of Drive Tests (MDT), which is one of the key technologies in SON [6]. MDT defines the collection and transmission of measurement reports from mobile stations to base station, in order to predict the signal coverage and to save the time and cost of mobile communication service providers to perform drive tests. However, the location information of user equipment is not always accurate. Akbari et.al. provided an analytical model to quantify the effect of positioning errors on the coverage prediction [7]. In this paper, we use the measurement reports to obtain data to train our model.

Kumar, Farooq, and Imran investigated proactive network failure prediction methods [8]. They used realistic mobile communication network data, and compared the performance of the support vector machine and several neural networks.

Zoha et.al. used the MDT measurement reports and the Fuzzy Q-Learning (FQL), which combines Fuzzy-Logic and Q-Learning, to develop a cell outage detection platform [9]. They also use the Multi-Dimensional Scaling (MDS) to reduce the dimensionality.

Ma et.al. also used the MDT measurement reports to investigate the cell outage detection problem [10]. They proposed a Dynamic Affinity Propagation (DAP) algorithm. Based on computer simulation results for heterogeneous network environments, their proposed method could successfully detect cell outage, and indicate specific outage areas.

Alias, Saxena, and Roy classified the states of 5G base stations into four categories: healthy, degraded, crippled, and catatonic [11]. They developed a Hidden Markov Model (HMM) to automatically obtain the current states of base stations, and to predict the probability of cell outage. Simulation results show that the prediction accuracy of their method is 80% in dense 5G heterogeneous networks.

III. System Model and Problem Formulation

In the system model, there exist multiple base stations, as Fig. 1 shows. Each base station could be in the state of either normal or outage. For those mobile stations associated with normal base stations, the mobile communication service meet the users' requirements. For those mobile stations associated with outage base stations, the mobile communication service would be unavailable or unacceptable.

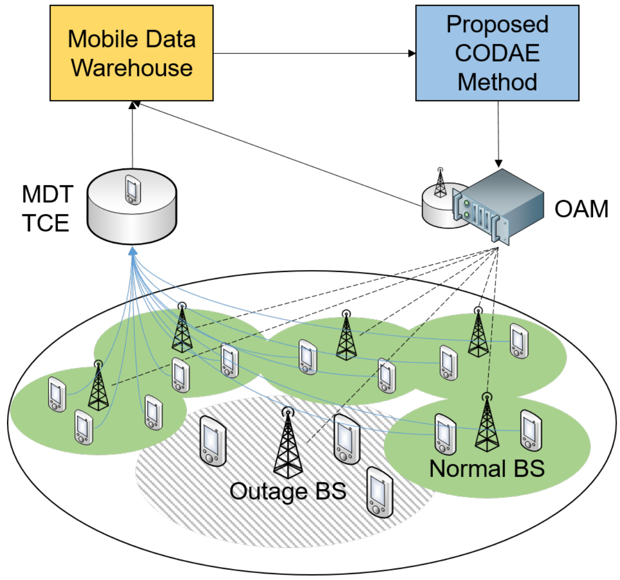

Fig. 1. The system model of cell outage detection.

The cell outage detection problem is to determine whether there exist any outage base station(s) in the mobile communication service region under consideration. This problem could be also regarded as the anomaly detection problem, i.e., the problem to determine whether the mobile communication service is normal or not.

Each mobile station measures the RSRP and RSRQ values from the neighboring base stations, and send the measurement report to the MDT Trace Collection Entity (TCE). The measurement reports in the TCE are processed and transformed to the ready-to-use dataset, and stored in the mobile data warehouse. The cell outage detection method use the dataset in the mobile data warehouse to detect the cell outage. If there exists any cell outage, the Operations, Administration and Management (OAM) would be notified, and then appropriate cell outage compensation (COC) methods could be applied.

IV. Proposed CODA Method

In order to make the proposed CODA method to be scalable, the whole mobile communication service region is divided into multiple smaller service areas, each formed by a group of n cells. Suppose that a mobile station generates a measurement report $\mathbf{x}$ in a service area, where

$$\mathbf{x} = \{\mathrm{RSRP}_1, \mathrm{RSRP}_2, \ldots, \mathrm{RSRP}_n, \mathrm{RSRQ}_1, \mathrm{RSRQ}_2, \ldots, \mathrm{RSRQ}_n\}\,. \tag{1}$$

The measurement report would be high-dimensional when n is medium to large. Since the cell outage detection problem is equivalent to the anomaly detection problem, the proposed CODA method uses the autoencoder to detect whether there exists any cell outage in a service area with n cells. We use

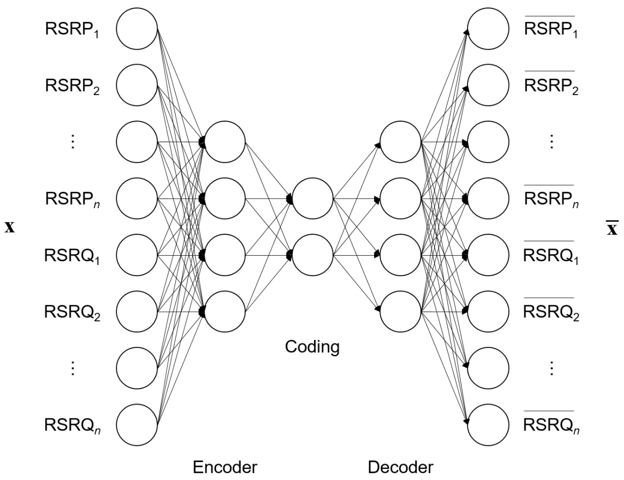

Fig. 2. The proposed cell outage detection method using autoencoder.

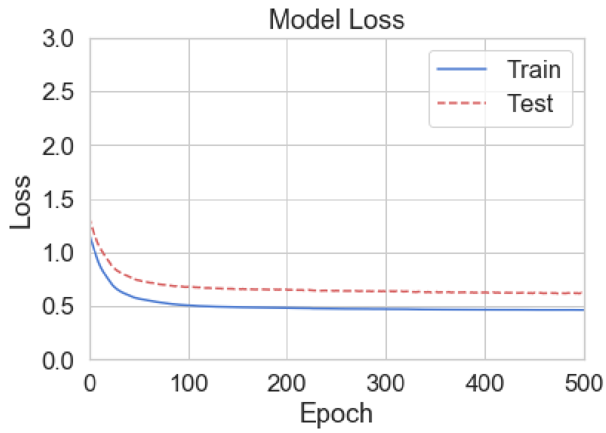

Fig. 3. The training loss and testing loss for each epoch.

the autoencoder to detect whether the measurement report $\mathbf{x}$ is normal or not.

Fig. 2 illustrates the autoencoder used in the proposed CODA method. There are two symmetrical parts in the autoencoder, including the encoder and the decoder. The objective of the encoder is to find the compressed representation of the input features which are the RSRP and RSRQ values from neighboring cells, so that the most important features could be kept. This compressed representation is called the *coding* of the input data. The high-dimensional measurement report could be represented as the low-dimensional coding. The objective of the decoder is to reconstruct the input data from the coding. The output of the decoder is the *reconstruction* $\bar{\mathbf{x}}$.

The autoencoder tries to learn the identity function:

$$\bar{\mathbf{x}} = f(\mathbf{x}) \approx \mathbf{x} \tag{2}$$

During the training of this autoencoder neural network, the well-known back propagation algorithm is adopted. The optimization objective is to minimize the reconstruction error L:

$$L(\bar{\mathbf{x}}, \mathbf{x}) = \|\bar{\mathbf{x}} - \mathbf{x}\|^2 \tag{3}$$

After the autoencoder is trained by using the normal dataset, it is used to detect the anomaly of the new inputs. An appropriate decision threshold is set depending on the precision-recall consideration of mobile communication service providers. When the reconstruction error of the new input data is below the decision threshold, this new input data is predicted as normal; otherwise, a cell outage is predicted.

V. Simulation Results and Discussions

We use the ns-3 network simulator to generate the dataset [12]. The ns-3 network simulator is a discrete-event network simulator for research and educational purpose. 9 base stations are deployed. The locations of these 9 base stations form a 3×3 squared matrix. The distance between adjacent base stations is 500 meters. In each of the 9 cells, 50 mobile stations are randomly deployed. Totally we deploy 450 mobile stations. The deployment radius is 250 meters. The transmission power of each base station is set as 46 dBm. When the cell outage occurs, the transmission power of the center base station is set as 30 dBm. 33,528 normal measurement reports and 30,860 anomaly measurement reports are generated. After the data processing and transformation, 2,466 normal data samples and 2,482 anomaly data samples form the ready-to-use dataset. The total 4,948 data samples are randomly split into two parts: the training dataset and the testing dataset. 20% of the data samples are for testing. Thus we have 990 testing data samples. The remaining 80% are for training. Among the training data samples, we use the 1,959 normal data samples to train the autoencoder. The training epoch is set as 500.

Fig. 3 shows the training loss and the testing loss for each epoch. The loss is defined as the reconstruction error, as (3) shows. Both the training loss and testing loss decrease quickly for the first tens of epochs, and then keep almost consistent after 100 epochs.

Fig. 4 shows the Receiver Operating Characteristic (ROC) curve, which is a very useful tool to understand the performance of anomaly detection. The ROC curve shows the true positive rate versus the false positive rate for different decision threshold values. The system performance is usually better when the blue ROC curve is closer to the upper left corner. When the false positive rate is 0.2, the true positive rate is almost 0.8. Based on this ROC curve, we also calculate the Area Under Curve (AUC) to be 0.86, which is good as it is close to 1.

Precision and recall are also useful tools to understand the performance of anomaly detection. Ideally, we want both high precision and high recall, since high precision relates to a low false positive rate, and high recall relates to a low false negative rate. However, there usually exists some tradeoff between precision and recall. Fig. 5 shows the tradeoff between precision and recall, which is affected by the decision threshold value. A high area under the blue curve represents both high recall and high precision. For example, the proposed CODA method is returning accurate results (precision = 0.8),

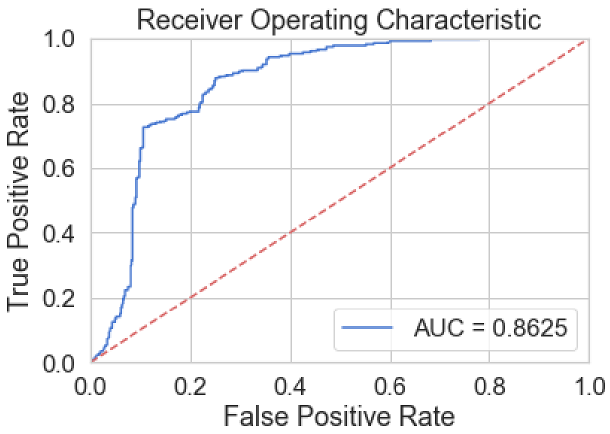

Fig. 4. The ROC curve.

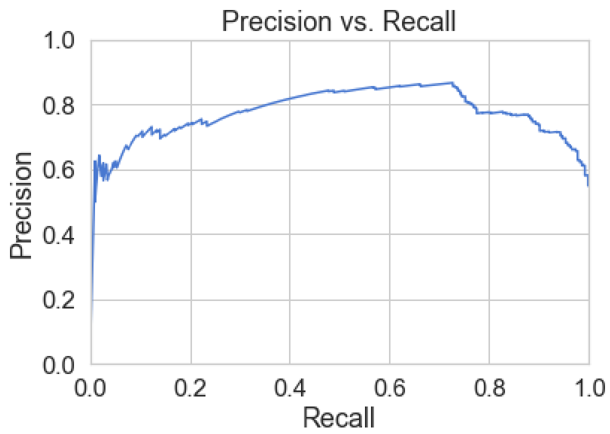

Fig. 5. Precision vs. recall.

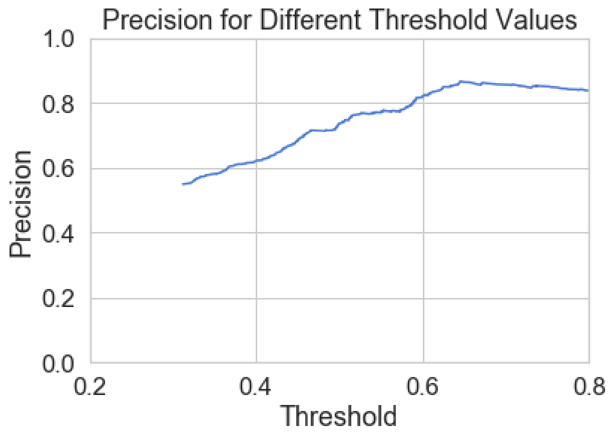

Fig. 6. Precision vs. threshold values.

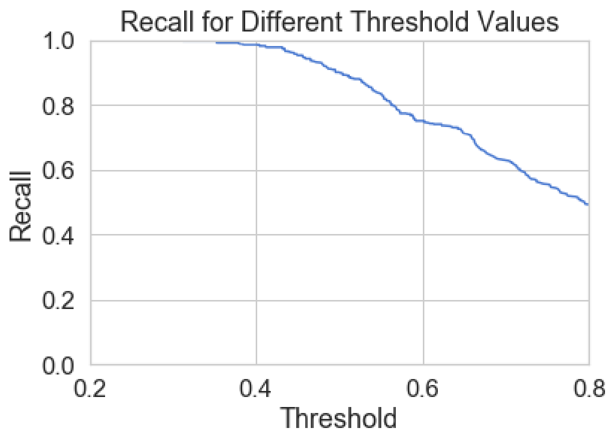

Fig. 7. Recall vs. threshold values.

as well as returning a majority of all positive results (recall $\approx$ 0.75).

Fig. 6 and 7 shows the precision and recall for different decision threshold values, individually. These two figures show the trend that as the decision threshold value increases, the precision is improved, but the recall decreases. If the decision threshold value is set as 0.6, the precision would be approximately 0.8, and the recall would be approximately 0.75.

While solving the cell outage detection problem, mobile communication service providers might want to keep the flexibility to find a suitable precision-recall tradeoff. For example, 0.75 recall mentioned above might not be good enough. The decision threshold value could be decreased from 0.6 to, say, 0.55, in order to improve the recall, with the tradeoff to slightly decrease the precision. Fig. 8 shows a scatter plot of the reconstruction errors of the 990 testing data samples. The reconstruction errors of the normal data samples are denoted as blue circles, and the reconstruction errors of the outage data samples are denoted as red crosses. This figure shows that, most of the reconstruction errors of the normal data samples locate at the lower part of the figure, and most of the reconstruction errors of the outage data samples locate at the higher part. This shows the effectiveness of the proposed autoencoder-based method. The red horizontal line represents the decision threshold value of 0.55. For those data points below the decision threshold line, the proposed CODA method predicts them as normal. On the contrary, for those data points above the decision threshold line, the proposed CODA method predicts them as outage.

Fig. 9 shows the confusion matrix of the proposed CODA method when the decision threshold value is set as 0.55. The true positive (403 at the lower-right) and the true negative (387 at the upper-left) occupy the majority of the 990 testing data samples. The precision is $403/(403 + 120) = 0.77$, and the recall is $403/(403 + 80) = 0.83$. Compared with the previous case (decision threshold value = 0.6, precision = 0.8, and recall = 0.75), the recall is improved, while the precision slightly decreases. Mobile communication service providers could adjust to appropriate precision-recall based on their requirements.

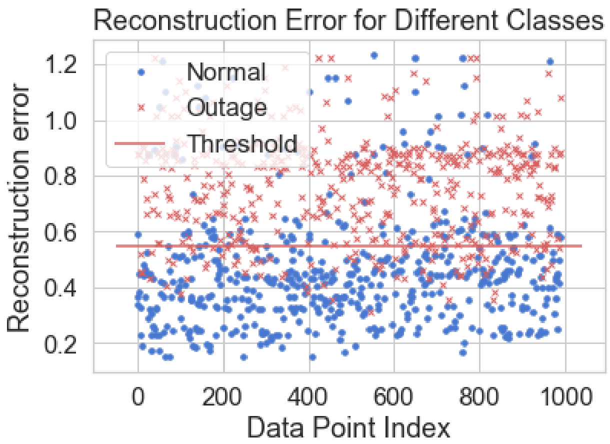

Fig. 8. Reconstruction error for normal and outage data with a decision threshold boundary.

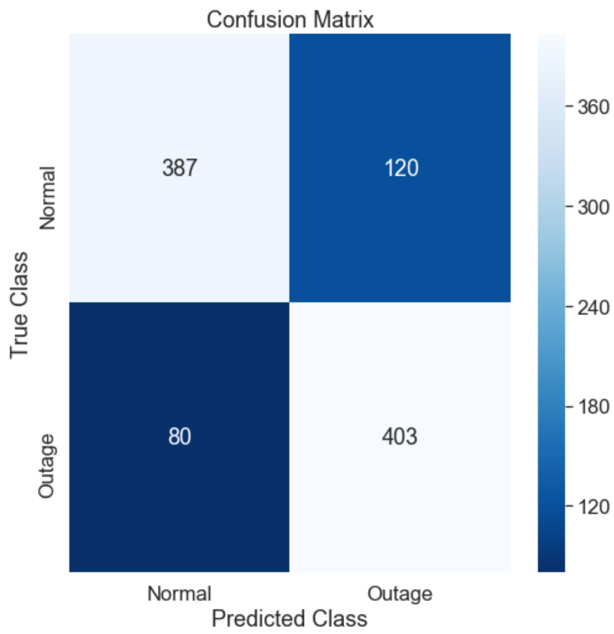

Fig. 9. The confusion matrix.

VI. Conclusion

In this paper we propose the CODA method to solve the cell outage detection problem in 5G Self-Organizing Networks. The proposed CODA method trains the autoencoder neural network by the measurement reports from mobile stations, and compares the reconstruction error of a new measurement report with the decision threshold to predict whether there exists a cell outage. Simulation results validate the effectiveness of the proposed CODA method.

In the near future, we are going to investigate the possibility to improve precision and recall of the proposed CODA method. We will also investigate both the cell outage detection and cell outage compensation solutions with deep learning / reinforcement learning techniques.

References

[1] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang, “What Will 5G Be?” *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1065–1082, June 2014.

[2] 3GPP, "Self-Configuring and Self-Optimizing Network (SON) Use Cases and Solutions," TR 36.902, V9.3.1, 2011-03.

[3] A. Imran, A. Zoha, and A. Abu-Dayya, “Challenges in 5G: How to Empower SON with Big Data for Enabling 5G,” *IEEE Network*, vol. 28, no. 6, pp. 27–33, Nov 2014.

[4] 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 Communications Surveys Tutorials*, vol. 19, no. 4, pp. 2392–2431, Fourthquarter 2017.

[5] J. Moysen and L. Giupponi, “From 4G to 5G: Self-Organized Network Management meets Machine learning,” *Computer Communications*, vol. 129, pp. 248 – 268, 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0140366418300380

[6] J. Johansson, W. A. Hapsari, S. Kelley, and G. Bodog, “Minimization of Drive Tests in 3GPP Release 11,” *IEEE Communications Magazine*, vol. 50, no. 11, pp. 36–43, November 2012.

[7] I. Akbari, O. Onireti, A. Imran, M. A. Imran, and R. Tafazolli, “How Reliable is MDT-Based Autonomous Coverage Estimation in the Presence of User and BS Positioning Error?” *IEEE Wireless Communications Letters*, vol. 5, no. 2, pp. 196–199, April 2016.

[8] Y. Kumar, H. Farooq, and A. Imran, “Fault Prediction and Reliability Analysis in a Real Cellular Network,” in *2017 13th International Wireless Communications and Mobile Computing Conference (IWCMC)*, June 2017, pp. 1090–1095.

[9] A. Zoha, A. Saeed, A. Imran, M. Imran, and A. Abu-Dayya, “A Learning-Based Approach for Autonomous Outage Detection and Coverage Optimization,” *Transactions on Emerging Telecommunications Technologies*, vol. 27, 08 2015.

[10] Y. Ma, J. Zhu, L. Liu, F. Lv, and Y. Wang, “A Dynamic Affinity Propagation Clustering Algorithm based on MDT in Self-Healing Heterogeneous Networks,” in *2016 16th International Symposium on Communications and Information Technologies (ISCIT)*, Sep. 2016, pp. 318–323.

[11] M. Alias, N. Saxena, and A. Roy, “Efficient Cell Outage Detection in 5G HetNets Using Hidden Markov Model,” *IEEE Communications Letters*, vol. 20, no. 3, pp. 562–565, March 2016.

[12] “The ns-3 Network Simulator.” [Online]. Available: https://www.nsnam.org/