Author(s): Asghar, Muhammad Zeeshan; Azhar, Farhan; Nauman, Muhammad; Ali, Nouman; Maqbool, Muaz; Ilyas, Muhammad Saqib; Baig, Mirza Mubasher

Title: Cell state prediction through distributed estimation of transmit power

Year: 2019

Version: Accepted version (Final draft)

Please cite the original version:

Asghar, M. Z., Azhar, F., Nauman, M., Ali, N., Maqbool, M., Ilyas, M. S., & Baig, M. M. (2019). Cell state prediction through distributed estimation of transmit power. In O. Galinina, S. Andreev, S. Balandin, & Y. Koucheryavy (Eds.), NEW2AN 2019, ruSMART 2019 : Internet of Things, Smart Spaces, and Next Generation Networks and Systems : Proceedings of the 19th International Conference on Next Generation Wired/Wireless Networking, and 12th Conference on Internet of Things and Smart Spaces (pp. 365-376). Springer. Lecture Notes in Computer Science, 11660. https://doi.org/10.1007/978-3-030-30859-9_31

Cell state prediction through distributed estimation of transmit power

Muhammad Zeeshan Asghar[1], Farhan Azhar[2], Muhammad Nauman[2], Muaz Maqbool[2], Nouman Ali[2], Muhammad Saqib Ilyas[2], and Mirza Mubasher Baig[2]

[1] University of Jyväskylä, Finland muhammad.z.asghar@jyu.fi
[2] National University of Computer and Emerging Sciences, Lahore, Pakistan
{l154051,l154310,l154053,l154160}@lhr.nu.edu.pk,
{saqib.ilyas,mubasher.baig}@nu.edu.pk

Abstract. Determining state of each cell, for instance, cell outages, in a densely deployed cellular network is a difficult problem. Several prior studies have used minimization of drive test (MDT) reports to detect cell outages. In this paper, we propose a two step process. First, using the MDT reports, we estimate the serving base station's transmit power for each user. Second, we learn summary statistics of estimated transmit power for various networks states and use these to classify the network state on test data. Our approach is able to achieve an accuracy of 96% on an NS-3 simulation dataset. Decision tree, random forest and SVM classifiers were able to achieve a classification accuracy of 72.3%, 76.52% and 77.48%, respectively.

Keywords: 5G cellular networks· Cell outage detection· Machine Learning.

1 Introduction

Traditionally, the deployment, operational optimization and troubleshooting have been extensive manual tasks. For large scale networks, this manual effort becomes intractable. Significant recent research focus has been directed towards Self Organizing Networks (SONs), which aim to reduce human operator involvement in the running of a network [5, 3].

Three main tasks in the realm of SON are self-configuration, self-healing and self-optimization [5, 3, 7]. Within self-healing, an important sub-task is to automatically detect faulty cells. Cell failures can result from a variety of reasons, such as component failures and mis-configuration [13, 15, 18]. Traditional techniques to detect cell failures include manual scanning of alarms from network management systems and manual drive tests. With network sizes growing, not only geographically, but also in density, this process is way too slow and ad hoc to achieve competitive network operations. Thus, automation of cell outage detection has become a necessity.

Some prior work has used insights from the way cellular networks operate to detect cell outages. For instance, [15] used neighbor cell list (NCL) reports to

construct a visibility graph. The authors observed that topology changes in this graph indicate cell outages. In [13], a weighted combination of the distribution of channel quality indicator (CQI), time-correlation of CQI differential and registration request frequency was used to detect cell outages. Incoming handover request statistics were used to detect cell outages in [4, 20]. A hidden Markov model based cell outage detection scheme was proposed in [2].

Recently, many researchers have taken a machine learning approach to cell outage detection. Clustering [14]. Bayesian networks [12]. Onireti proposed k nearest neighbors and local outlier factor based classifiers for heterogeneous cellular network in [17]. Zoha et al. applied local outlier factor and SVM based classification algorithms to the cell outage detection problem in [21]. In [6], Chernov et al. compared k nearest neighbor, self organizing map and several probabilistic data structures for sleeping cell detection in LTE networks. Gurbani et al. applied Chi-Square test and Gaussian Mixture Models trained on LTE network log data to detect cell outages in [10]. Wang et al. proposed an RBF neural network based approach to detect cell outages in [19]. Mulvey et al. applied a recurrent neural network to the sleeping cell detection in [16].

In this work, similar to works such as [17], we use Minimization of Drive Test (MDT) reports, which are transmitted by each active user in a network, periodically [1]. Each report consists of current channel and carrier characteristics. Each user transmits their location along with signal strength and quality for four best reference signals received from neighboring base transceiver stations (BTSs).

We make the simplifying assumption that the best strength signal comes from the geographically nearest BTS, the second best strength signal form the second nearest BTS and so on. Under this assumption, the received signal strength depends solely on two factors: the transmit power and the distance between the user and a BTS. Given BTS locations and a channel propagation model, from each MDT report, we can estimate the transmit power being used by the four nearest BTSs. In the present work, we only consider the BTS nearest to each user and the highest strength signal to estimate the transmit power being used by the BTS nearest to each user.

In case of normal network operation, our estimated transmit power should be somewhat similar for all users. In case of a cell outage, our estimated transmit power for the users in the vicinity of the failed BTS should be different from others. Consider a toy example, illustrated in Fig. 1. User A's nearest BTS is BTS1. Under normal conditions, user A would receive the best strength signal from BTS1[3]. If we adjust the path loss for the distance between user A and BTS1 to the best received signal strength, we get an estimated transmit power. Most users in the network would arrive at similar estimate transmit power levels. However, if BTS1 fails, user A's best received signal strength now comes from BTS2, which would be significantly weaker than before. If we adjust the path loss for the distance of user A from its nearest BTS, i.e., BTS1, we would arrive at a

[3] Under obstructions or irregular terrains, this would not hold for many users. Nevertheless, several users may still be located such that this condition holds.

significantly lower estimated transmit power. Users that are not in the vicinity of the failed BTS would estimate a higher transmit power. We hypothesize that this method can be used to detect the network's operational state.

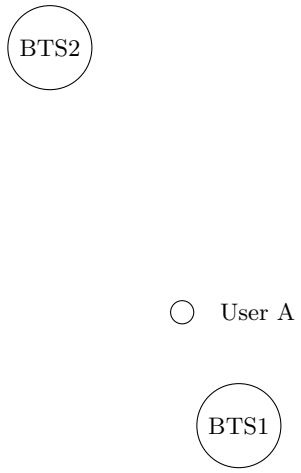

Fig. 1. A user in the proximity of two BTSs, BTS1 and BTS2. Normally, user A's best received signal strength should depend on distance from BTS1. However, if BTS1 fails, the best received signal strength would no longer be a function of geographical distance from the nearest BTS, i.e., BTS1.

We develop an algorithm based on the above technique. Our approach is able to achieve an accuracy of 96% on an NS-3 simulation dataset. Decision tree, random forest and SVM classifiers were able to achieve accuracies of 72.3%, 76.52% and 77.48%, respectively.

The rest of the paper is structured as follows. In section 2, we provide necessary background and problem formulation. In section 3, we describe our proposed classifier. In section 4, we describe our simulation setup. The results obtained through our proposed classifier on the simulation traces are discussed in section 5. We, then, conclude in section 6.

2 Background and problem formulation

In wireless communication, a signal travels through air from the transmitter to the receiver. If the signal is transmitted at a power level t (measured in dB), the strength of the signal received by the receiver r (also in dB) is typically lower than t. The difference in transmit and receive power levels, also known as power loss (l), is due to various factors such as attenuation, fading and noise.

Power loss depends on various factors such as communication signal frequency, distance between the transmitter and the receiver and the terrain. Researchers have developed various models that allow estimation of the channel

power loss given certain input parameters. Examples include free space path loss model [9] and Hata model [11]. The two most important parameters in path loss estimation are the distance between the communicating entities and the frequency of the radio signal. To emphasize this dependence, we will denote path loss as $l(d, f)$, where f is the frequency of the signal and d is the distance between the transmitter and the receiver.

A cellular network consists of a number of BTSs dispersed in a geographical area, such that each BTS provides coverage to users in its vicinity. Each BTS transmits radio signals that allow users in its vicinity to communicate using the network. One of these signals is called a reference signal - each BTS has a separate one. Every user is likely to receive the signals transmitted by several nearby BTSs, however weak some of these signals may be.

While each BTS uses several channels at different frequencies, these frequencies do not differ greatly. Furthermore, the path loss is a function of the log of the frequency. Thus, for a given distance between transmitter and receiver, the path loss is almost the same for all channels within a given operator's network. Thus, we consider path loss to be purely a function of the distance between the transmitter and receiver, denoted $l(d)$.

Consider a cellular network consisting of m BTSs $b_1, b_2, \ldots, b_m$ located at Cartesian coordinates $(x_1, y_1), (x_2, y_2) \ldots, (x_m, y_m)$, serving n users. Every active user periodically transmits an MDT report to a network entity. This MDT report consists of the user's current location, signal strength and quality of the four best reference signals the user is able to receive. In this paper, we consider a slotted time operation whereby at discrete intervals, each user transmits an MDT report.

We assume availability of training data in the form of MDT reports while operating the network in various discrete known states where in each state all BTSs are transmitting at the same power level. For instance, we may operate all BTSs transmitting for some time at 43 dBm and call this state 0 of the network. Then, for some time we may operate all BTSs transmitting at 40 dBm and call this state 1 of the network. Then, for some time we may operate all BTSs transmitting at 46 dBm and call this state 2 of the network. We label all the received MDT reports with the state label of the network at that time.

An MDT report m_i^j received from the jth user for the ith time interval consists of the following fields:

- User location (x_i^j, y_i^j): The current x and y coordinate of the user element (UE).
- RSRP1-4 $(r_i^j[1], r_i^j[2], r_i^j[3], r_i^j[4])$: The power of the four strongest reference signals received from nearby BTSs, measured in dB. $r_i^j[1]$ is the strongest and $r_i^j[4]$ is the weakest.
- RSRQ1-4 $(q_i^j[1], q_i^j[2], q_i^j[3], q_i^j[4])$: The signal quality of the above four signals, also measured in dB. The signal power does not necessarily indicate the quality of the signal. The signal quality may be thought of as its ability to carry out intelligible high-speed communication. Factors such as interference and signal to noise ratio (SNR) affect signal quality.

– Label (s_i^j): We assume that we have labelled training data. Label s_i^j represents the network state as perceived by user j in the ith interval. The label could have different integer values representing different transmit power levels, while one network state represents a failure of one of the BTSs.

3 Proposed Solution

Consider MDT report m_i^j, emanating from user j in interval i. We calculate the Euclidean distance of the corresponding user from each of the BTSs as:

$$d(u_i^j, b_k) = \sqrt{\left(x_k - x_i^j\right)^2 + \left(y_k - y_i^j\right)^2} \tag{1}$$

The user is then considered associated to the base station b_o where:

$$o = \arg\min_k d(u_i^j, b_k) \tag{2}$$

The corresponding distance from the nearest BTS is picked as: $d_i^j = d(u_i^j, b_o)$. Then, the path loss $l(d_i^j)$ is calculated corresponding to the distance d_i^j. In the present work, we use the free space path loss model, but any of the more sophisticated models could be used easily.

Since the power levels are in dB:

$$r_i^j[1] = t_i^j - l(d_i^j) \tag{3}$$

Since the left hand side in the above equation is known from the MDT reports, and the path loss is given by a path loss model, the transmit power can be estimated as:

$$t_i^j = r_i^j[1] + l(d_i^j) \tag{4}$$

The estimated transmit power for all MDT reports in a given interval can be averaged to give:

$$t_i = \frac{\sum_{i=0}^{n} t_i^j}{n} \tag{5}$$

We conducted a simulation study, which is formally described in section 4. From the simulation trace data, we have plotted the mean estimated transmit power t_i for the duration of the simulation in Figure 2. We setup the simulation such that network was in state 0 for the first 10 seconds, state 1 for the next 10 seconds, state 2 for the next 10 seconds and state 0 for the next 10 seconds. During the last ten seconds, one of the BTSs was forecefully failed, so some MDT reports were also labeled 3. State 3 represents the network state whereby one of the BTSs has failed.

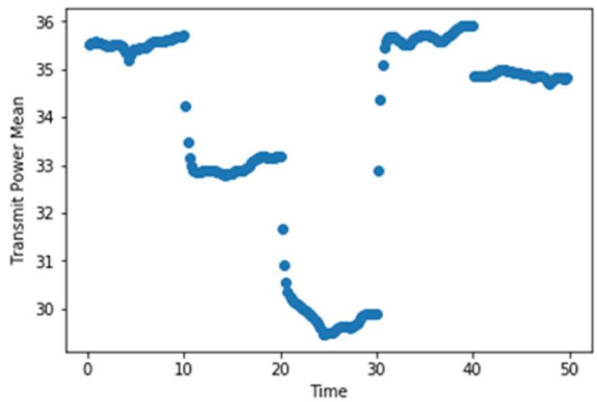

Fig. 2. Estimate transmit power vs simulation time

From Fig. 21, it appears that the average estimated transmit power is normally distributed around a different mean for each class. We may safely assume that the actual transmitted power is additively affected by a Gaussian variable with certain mean and standard deviation. Furthermore, the additive Gaussian distributions are independent for each class (different operating states vis a vis transmit power level), but have the same standard deviation. Under these conditions, we may calculate mean estimate transmit power for each class using training data and classify a test MDT report to the network state that has the closest average transmit power to the estimated transmit power for the test instance [8].

4 Experiment Setup

A simulation study has been performed in NS-3 whereby 105 mobile users were uniformly spread around 7 base stations. The base stations were spread regularly on the grid as shown in Fig. 3, where each circle represents a base station. The users were mobile and spread uniformly across the grid. The periodic (once every 0.2 ms) MDT report generated by each of the users were recorded for about 50 seconds. The data has been hand labeled so that each MDT report has been assigned a label from the set {0, 1, 2, 3}. Labels 0, 1 and 2 represent normal network conditions with distinct transmit power levels. An MDT report with label 3 indicates that the corresponding user was previously being served by a BTS that has failed. Only one BTS failure is simulated.

5 Results

We partitioned the simulation data into training and test split, where 80% of the data was taken as training data. Due to sparsity of the class 3 instances, we

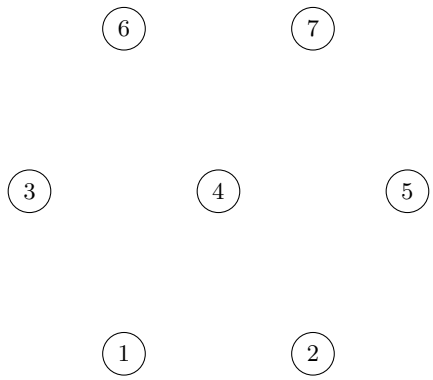

Fig. 3. Network topology for the simulation

partitioned the data such that a random 80% of each class' instances were in the training set and the rest were in the test set.

From all the t_i values from the training set that have label k, we pick the maximum and minimum values as $t^k_{\max}$ and $t^k_{\min}$, respectively. Here, $k \in \{0, 1, 2\}$. Then, for classifying a test MDT report, we selected the thresholds shown in Table 1.

Table 1. Classification rules for each class

Predicted class label	Threshold on estimated transmit power
0	$t^j_i > \frac{t^0_{\min}+t^1_{\max}}{2}$
1	$\frac{t^0_{\min}+t^1_{\max}}{2} \leq t^j_i < \frac{t^1_{\min}+t^2_{\max}}{2}$
2	$t^j_i \leq \frac{t^1_{\min}+t^2_{\max}}{2}$

The simple rules given in Table 1 form our classifier for cell state prediction. However, this simple nearest distance classifier did not prove to be effective for discriminating label 3, i.e., cell outage, from the remaining classes. The reason is that with only one BTS failing, several users that are not in the vicinity of the failed BTS are not affected by the failure. Such users would continue to perceive the network in one of states 0, 1 or 2. The few users that are affected by the failure estimate a different transmit power level. If a minority of users is affected by a failure, it skews the average estimated transmit power for network state 3 only slightly from that of a "healthy" network state. Thus, the average estimated transmit power level for class 3 differs only slightly from one of the other classes. A threshold-based classifier would not perform well.

This can be observed from Fig. 2. From time 40s to 50s, most of the BTSs are operating in state 0, but the average estimated transmit power is pulled down slightly by the few users that are in the vicinity of the failed cell.

To classify class 3 instances, we devised the following two additional rules. If either of the rules is true, the instance is classified as class 3.

Rule 1: Calculate the standard deviation of all estimated transmit power values as $\sigma = \text{stdev}(t_i^j)$. Instance m_i^j is classified as class 3, if:

$$t_i^j < \frac{t_{i-1} + t_{i-2}}{2} - \sigma \tag{6}$$

That is, if the estimated transmit power deviates from the mean estimated transmit power for the last two intervals by more than the standard deviation of estimated transmit power, an anomaly is flagged.

Rule 2: Label instance m_i^j as class 3, if:

$$r_i^j[1] < \frac{r_{i-1}^j[1] + r_{i-2}^j[1]}{2} \tag{7}$$

. That is, if RSRP 1 has dropped below the mean RSRP 1 value observed by the same user over the last two intervals, an outage is flagged. This is based on the hypothesis that a user near a failed BTS would notice a sudden drop in the best strength reference signal that they receive. For instance, in Fig. 1, if BTS1 fails, user A's RSRP1 value will drop significantly, because the best known reference signal now traverses a much greater distance, thereby suffering much greater path loss.

Based on this classifier, we were able to achieve the results given in Table 2. The average accuracy for the classifier was 96%. The confusion matrix for the classifier is given in Table 3.

Table 2. Accuracy, precision and recall for our cell state predictor

Average accuracy	Average F1 score	Average precision	Average recall
96.09%	90.33%	87.62%	94.48%

Table 3. Confusion matrix for our classifier

		Predicted Label			
		0	1	2	3
True Label	0	14536	105	105	334
	1	210	5040	0	0
	2	0	210	5040	0
	3	59	0	0	506

For comparison purposes, we trained three standard machine learning classifiers on the same feature set. We performed k-fold validation on these classifiers with k=5.

For the decision tree classifier, the average classification accuracy achieved was 72.3%. The confusion matrix for one of the folds for decision tree classifier on our dataset is given in Table 4.

Table 4. Confusion matrix for decision tree classifier

		Predicted Label			
		0	1	2	3
True Label	0	2469	359	147	20
	1	359	546	129	16
	2	161	134	710	45
	3	17	13	36	47

For the random forest classifier, the average classification accuracy achieved was 76.52%. The confusion matrix for one of the folds for decision tree classifier on our dataset is given in Table 5.

Table 5. Confusion matrix for random forest classifier

		Predicted Label			
		0	1	2	3
True Label	0	2562	268	160	5
	1	336	577	129	8
	2	145	93	794	18
	3	17	19	39	38

For the SVM classifier, the average classification accuracy achieved was 77.48%. The confusion matrix for one of the folds for decision tree classifier on our dataset is given in Table 6.

Table 6. Confusion matrix for SVM classifier

		Predicted Label			
		0	1	2	3
True Label	0	2618	216	161	0
	1	385	534	131	0
	2	130	79	836	5
	3	9	21	54	29

6 Conclusion

We used MDT reports from a simulation run to classify a cellular network's state in terms of its current transmit power and cell outage. We estimated the

transmit power from the MDT reports and applied thresholding to classify the healthy network's state. We developed two simple rules to classify the single cell outage state. Our classifier accuracy is 96%. Using the same feature set, the standard SVM classifier with 5-fold validation achieved an accuracy of 77.48%.

References

1. 3GPP: Universal Mobile Telecommunications System (UMTS); LTE; Universal Terrestrial Radio Access (UTRA) and Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Measurement Collection for Minimization of Drive Tests (MDT); Overall Description; Stage 2. Technical Specification (TS) 37.320, 3rd Generation Partnership Project (3GPP) (04 2011), version 10.1.0 Release 10
2. Alias, M., Saxena, N., Roy, A.: Efficient cell outage detection in 5g hetnets using hidden markov model. IEEE Communications Letters **20**(3), 562–565 (2016)
3. Aliu, O.G., Imran, A., Imran, M.A., Evans, B.: A survey of self organisation in future cellular networks. IEEE Communications Surveys & Tutorials **15**(1), 336–361 (2012)
4. de-la Bandera, I., Barco, R., Munoz, P., Serrano, I.: Cell outage detection based on handover statistics. IEEE Communications Letters **19**(7), 1189–1192 (2015)
5. Van den Berg, J., Litjens, R., Eisenblätter, A., Amirijoo, M., Linnell, O., Blondia, C., Kürner, T., Scully, N., Oszmianski, J., Schmelz, L.C.: Self-organisation in future mobile communication networks. In: Proceedings of ICT-Mobile Summit 2008, Stockholm, Sweden, 2008 (2008)
6. Chernov, S., Cochez, M., Ristaniemi, T.: Anomaly detection algorithms for the sleeping cell detection in lte networks. In: 2015 IEEE 81st Vehicular Technology Conference (VTC Spring). pp. 1–5. IEEE (2015)
7. Combes, R., Altman, Z., Altman, E.: Self-organization in wireless networks: A flow-level perspective. In: 2012 Proceedings IEEE INFOCOM. pp. 2946–2950. IEEE (2012)
8. Duda, R.O., Hart, P.E., Stork, D.G.: Pattern classification and scene analysis, vol. 3. Wiley New York (1973)
9. Garg, V., Wilkes, J.E.: Wireless and Personal Communications Systems. Prentice Hall, Upper Saddle Riverm NJ, 1 edn. (1996)
10. Gurbani, V.K., Kushnir, D., Mendiratta, V., Phadke, C., Falk, E., State, R.: Detecting and predicting outages in mobile networks with log data. In: 2017 IEEE International Conference on Communications (ICC). pp. 1–7. IEEE (2017)
11. Hata, M.: Empirical formula for propagation loss in land mobile radio services. IEEE transactions on Vehicular Technology **29**(3), 317–325 (1980)
12. Khanafer, R.M., Solana, B., Triola, J., Barco, R., Moltsen, L., Altman, Z., Lazaro, P.: Automated diagnosis for umts networks using bayesian network approach. IEEE Transactions on vehicular technology **57**(4), 2451–2461 (2008)
13. Liao, Q., Wiczanowski, M., Stańczak, S.: Toward cell outage detection with composite hypothesis testing. In: 2012 IEEE International Conference on Communications (ICC). pp. 4883–4887. IEEE (2012)
14. Ma, Y., Peng, M., Xue, W., Ji, X.: A dynamic affinity propagation clustering algorithm for cell outage detection in self-healing networks. In: 2013 IEEE Wireless Communications and Networking Conference (WCNC). pp. 2266–2270. IEEE (2013)

15. Mueller, C.M., Kaschub, M., Blankenhorn, C., Wanke, S.: A cell outage detection algorithm using neighbor cell list reports. In: International Workshop on Self-Organizing Systems. pp. 218–229. Springer (2008)
16. Mulvey, D., Foh, C.H., Imran, M.A., Tafazolli, R.: Cell coverage degradation detection using deep learning techniques. In: 2018 International Conference on Information and Communication Technology Convergence (ICTC). pp. 441–447. IEEE (2018)
17. Onireti, O., Zoha, A., Moysen, J., Imran, A., Giupponi, L., Imran, M.A., Abu-Dayya, A.: A cell outage management framework for dense heterogeneous networks. IEEE Transactions on Vehicular Technology **65**(4), 2097–2113 (2015)
18. Wang, W., Zhang, J., Zhang, Q.: Cooperative cell outage detection in self-organizing femtocell networks. In: 2013 Proceedings IEEE INFOCOM. pp. 782–790. IEEE (2013)
19. Wang, Y., Long, P., Liu, N., Pan, Z., You, X.: A cooperative outage detection approach using an improved rbf neural network with genetic abc algorithm. In: 2018 10th International Conference on Wireless Communications and Signal Processing (WCSP). pp. 1–6. IEEE (2018)
20. Zhang, T., Feng, L., Yu, P., Guo, S., Li, W., Qiu, X.: A handover statistics based approach for cell outage detection in self-organized heterogeneous networks. In: 2017 IFIP/IEEE Symposium on Integrated Network and Service Management (IM). pp. 628–631. IEEE (2017)
21. Zoha, A., Saeed, A., Imran, A., Imran, M.A., Abu-Dayya, A.: A learning-based approach for autonomous outage detection and coverage optimization. Transactions on Emerging Telecommunications Technologies **27**(3), 439–450 (2016)