N-GRAM ANALYSIS FOR SLEEPING CELL DETECTION IN LTE NETWORKS

Fedor Chernogorov[1,2], *Tapani Ristaniemi*[1], *Kimmo Brigatti*[1], *Sergey Chernov*[1]

[1]University of Jyväskylä, Department of Mathematical Information Technology, Jyväskylä, Finland
[2]Magister Solutions Ltd., Jyväskylä, Finland

ABSTRACT

Sleeping cell detection in a wireless network means to find the cells which are not working properly due to various reasons. The research in the area has mostly focused on cell outage detection, e.g. due to hardware failures at the base station antennas or non-optimal network planning. In this paper we extend the research into a more challenging setting which is overlooked in the literature: the case where no outages occur in the network. The essence of the proposed method for detection of problematic cells is to analyze the sequences of the events reported by the mobile terminals to the serving base stations. The suggested n-gram analysis includes dimensionality reduction and classification of the data and ends up with providing a set of abnormal users, which at the end reveal the location of the problematic cell. We verify the proposed framework with simulated LTE network data and using the minimization of drive testing (MDT) functionality to gather the training and testing data sets.

1. INTRODUCTION

Self-healing, which is a part of self-organizing network concept, means automated detection of problems or malfunctioning in the radio network elements and actions to automatically recover from these problematic situations [1]. Most of the works considered so far have focused on cell outage detection (see e.g. [2–4] and references therein) and management [5, 6]. Reasons for outage situations are many, but the usual ones are hardware problems in base station antennas, improper radio network planning, erroneous antenna tilt or transmit power. Hence the usual approach for cell outage detection is to analyze several key performance indicator (KPI) measurements from both base stations and mobile terminals.

Latest works in this line of research were recently published by the authors in [7] and [8]. The approach in [7] was to analyze the data set of signal strength and quality measurements reported by mobile terminals. These measurements contained both serving and neighboring cell measurements in LTE network according to minimization of drive testing (MDT) functionality specified by 3GPP. The main finding was that advanced data mining and machine learning techniques, which rely on autonomous learning of network behavior, were able to reveal latent abnormal behavior in the high dimensional data set of RF measurements and can thus be used to pinpoint a problematic cell in the network. In [8] this approach was extended by targeting to find similarities between periodical measurement reports and reports related to failures happened before at the radio link. By this way one was able to substantially increase the number of samples (in addition to true failure reports) which indicated the existence of a problem in specific cells, resulting in more reliable and faster detection.

All the above-mentioned approaches rely on the measurements of the radio environment, which however, are able to reveal only radio related problems. In this paper we extend the scope of problem detection in radio networks by considering the case where radio coverage outages do not exist. This is a relevant case in practice e.g. when there exists hierarchical cells (pico/micro/macro) in the same area or when the real problem of a particular cell is not radio related at all. An example of the latter case is a software bug or a malfunctioning protocol. Detecting a cell having such problems is no longer doable by analyzing RF measurements, but calls for another approach. A relevant solution where the problematic cell was detected by investigating graphs constructed from the reported neighboring cell patterns can be found in [4]. The essence of this paper is to employ more generic approach by analyzing the sequences of events reported by mobile terminals to the serving base stations. Subsequently, the approach will end up with providing a set of abnormal users (or calls) in the networks, which can be utilized at the end to reveal the location of the problematic cell.

2. SLEEPING CELL PROBLEM

Sleeping cell is a special kind of cell degradation. A cell is called degraded in case if it is not 100% functional - its services are suffering in terms of quality what affects user ex-

This work was financially supported by Tekes (the Finnish Funding Agency for Technology and Innovation) under grant 40040/08 "Real-Time Management of Industrial Processes and Communication Networks (MIPCOM)" and Celtic-Plus (an industry-driven European research initiative) project "Cognitive Network Management Under Uncertainty (COMMUNE)". Authors would like to thank colleagues from University of Jyväskylä and Magister Solution Ltd. for their professional advices and support.

perience. There exist a vague classification of degraded cells depending on how much they affect the network operation (partly based on [9]). The first type is *impaired* cell - which still carries some traffic, but the performance characteristics are slightly lower expected. The second kind of degradation is *crippled* cell, characterized by a severely decreased capacity. The last, clearly most critical type of sleeping cell is *catatonic* cell - kind of outage which leads to complete absence of service in the faulty area and cell does not carry any traffic and for that reason it is important to timely detect such cells and apply recovery actions.

Usual degraded cell produces fault alarms which are available to mobile network operator. In opposite, in sleeping cells degradation appears seamlessly and no direct notification to the service provider is given.

Different hardware or software failures can cause appearance of a sleeping cell and due to that it is considered to be a complex umbrella term. In this research we investigate catatonic sleeping cells with RACH (Random Access Channel) problem described in further details in Section 4.

3. DETECTION FRAMEWORK

3.1. N-Gram Analysis

An n-gram is defined as a subsequence of n terms. These terms can be e.g. letters or words from a sequence. The analysis results in statistics regarding the frequency of occurrence of n-grams within string sequence. Thus, feature vector of n-gram frequencies can be assembled from the string sequence.

N-gram analysis is widely used in spheres concerning data processing. It has been utilized e.g. for the analysis of whole-genome protein sequences [10], computer virus detection [11] and also in a wide variety of natural language processing applications.

In our research the terms are network events reported by the mobile terminal to the base station (in total 10 events listed in Table 1). [1] The data used for sleeping cell detection is a K by 10^n matrix containing the n-gram frequencies of each of K individual users (or call), where n is the number of terms in considered subsequences.

3.2. Dimensionality reduction and classification

The goal for data analysis here is first to identify abnormal calls. As a next step this information is used for the detection of the sleeping cell. To do that, one usually performs reduction of the dimensionality for the data and clusters the data in low dimensional space. Here we performed standard principal component analysis for dimensionality reduction and applied the FindCBLOF [14] algorithm for clustering and outlier detection part. The advantage of FindCBLOF is in its ability to find local outliers based on the clustering solution for training data.

[1] RSRP = Reference Signal Received Power; RSRQ = Reference Signal Received Quality; A2 = an event which triggers when the serving cell becomes worse than threshold; A3 = an event which triggers when a neighboring cell becomes an offset better than the serving cell.

Table 1. Network events triggering MDT log entry

PL PROBLEM - Physical Layer Problem [12].
RLF - Radio Link Failure [13].
RLF REESTABLISHMENT - Connection reestablishment after RLF.
A2 RSRP ENTER - RSRP goes under A2 enter threshold.
A2 RSRP LEAVE - RSRP goes over A2 leave threshold.
A2 RSRQ ENTER - RSRQ goes over A2 enter threshold.
A3 RSRP - A3 event, according to spec.
HO COMMAND RCVD - handover command received [13].
HO COMPLETE RCVD - handover complete received [13].
HO TO VOID - handover is done to one of the cells in outer tier.

3.3. Symmetry Analysis of 2-Gram Subsequences

Under symmetry we mean the following: if the first event of a 2-gram is located in cell A and the second event is located in cell B, we are interested in how many of those 2-grams originate from A and how many originate from B. In simulations, where the user movement is random, one expects any 2-grams to be somewhat balanced. Hence, the deviation from learned balance is to be used as an indication of problem in a particular cell.

4. SIMULATION ASSUMPTIONS AND GENERATED DATA

Dynamic system level LTE simulator with step resolution of one OFDM[2] symbol has been used as a platform for data generation in this research. The simulator is designed in accordance to specifications 3GPP E-UTRAN Release. 8 and beyond. Methodology for mapping link level SINR to system is presented in [15].

Network scenario utilized in the simulations for this study and shown on Fig. 1, is an extended version of 3GPP macro case 1, described in [16]. Scenario setup is such that outer tier of cells is used only for interference generation to make radio link conditions more realistic. On the other hand 21 center cells are utilized for statistical data collection. The main simulation parameters applied in this research are presented in Table 2.

[2] OFDM - Orthogonal Frequency-Division Multiplexing

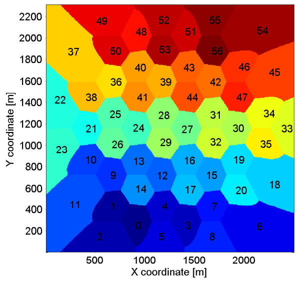
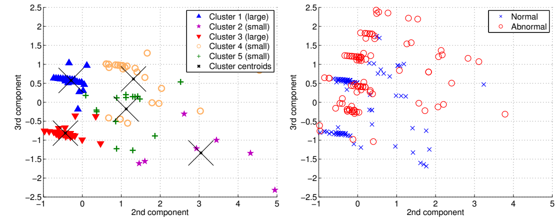


Fig. 1. Macro 57 network scenario layout



(a) FindCBLOF training (b) FindCBLOF testing

Fig. 2. FindCBLOF testing and training results

Table 2. General Simulation Parameters

Parameter	Value
Cellular layout	Homogeneous Macro 57
Number of cells	21 active and 36 interfering
Inter-Site Distance	500 m
Link direction	Downlink
Maximum BS TX power	46 dBm
Initial cell selection criterion	Strongest RSRP value
Simulation length	142 s (2000000 steps)
Simulation resolution	1 time step = 71.43 μs
Max number of UEs/cell	20
UE velocity	30 km/h
Duration of calls	Uniform 30 to 140 s
Traffic model	Constant Bit Rate 256 kbps
Reference case	Simulation without sleeping cell
Problematic case	Simulation with RACH problem in cell 28

In this paper the sleeping cell was modelled through malfunctioning of the Random Access Channel (RACH). RACH is a channel used in connection establishment in the beginning of a call when establishment procedure is initiated, during handover to another cell or connection re-establishment after handover failure or RLF.

By simulating LTE network operation we generate a performance monitoring dataset using the principle of drive test minimization reporting. This principle implies addition of log entry by the mobile terminal to a global MDT log either periodically or at occurrence of a specified network events, presented in Table 1. Usually one sample includes values of different performance indicators, time stamp and location fingerprint. Depending on the type of the sleeping cell we might need different amount of information from the log. As far as in this research we are doing identification of random access sleeping cells the required MDT data limits to event type and location of this event occurrence. In more details the procedure of drive test minimization, variables of the MDT log utilized for detection of other kinds of sleeping cells are presented in [7, 8].

5. RESULTS

5.1. Analysis of Abnormal User Calls

In accordance to our detection framework the first step is the construction of n-gram subsequences as described in Section 3.1. Using network events shown in Table 1, we chose $n = 2$ for simplicity and generate the full set of 2-gram subsequences.

Reference data were used for creation of a normal network operation model. There were 264 users with sequences longer than 20 event-triggered MDT log entries, while shorter user sequences were filtered. On the basis of these reference user sequences corresponding matrix of 2-gram occurrences was constructed. After that same procedure was done with the problematic data, Resulting occurrence matrix of 2-grams was compared to the corresponding reference matrix in order to find anomalous users.

Reference data were clustered to five groups, among which there were two large and dense clusters (1 & 3) and three small (clusters 2, 4 & 5), as shown in Fig. 2a. In problematic data, users were clustered into normal and abnormal groups; red markers shown on Fig. 2b represent abnormal user, while the rest belong to normal users' group. Decision whether a certain point is abnormal or not was based on the value of CBLOF, where higher likelihood sample abnormality corresponds to high score values. From Fig. 2a and Fig. 2b it can be seen that points from the problematic dataset which belong to the areas of compact clusters in training data are marked as normal, while samples which are outside of dense clusters or are in low density areas are clustered as abnormal. In total 113 users were marked as abnormal and 205 users as normal.

After having detected the abnormal users, their movement can be traced by locating the events through dominance maps like in [7]. As can be seen from Fig. 3, abnormal users tend to

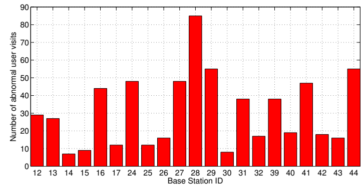


Fig. 3. Number of cell visits by abnormal users

be those who have visited the sleeping cell 28. Naturally due to mobility they visit other cells as well. In these simulations abnormal users on average camp in 6 different cells, while the range varies from 3 to 11 cells. From the histogram on Fig. 3, it can be observed that 85 abnormal users (75.2%) have visited cell 28. Next 8 most visited cells get from 38 to 55 visits from which 6 are neighbors of cell 28 (cells 24, 27, 29, 39, 41 & 44). On the basis of this information we can claim that there is some anomalous behavior in the area of cell 28.

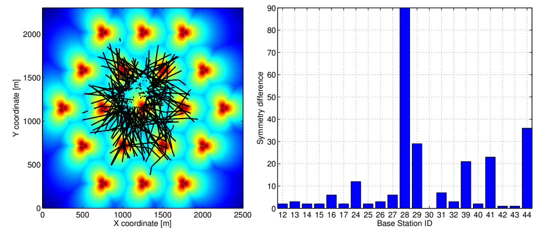
5.2. Analysis of Abnormal 2-Gram Sequences

Observation of the abnormal users gives only a rough idea of possible problem in the cell of interest. For that reason more detailed analysis of 2-gram subsequences of the abnormal users' calls needs to be employed. As further results demonstrate, this approach gives more reliable indication of problem existence. In particular, the knowledge of the most descriptive 2-grams, meaning that over 50% of abnormal users have this 2-gram occurred at least once, is taken into account. Nine 2-grams met this condition and as an example the characteristics of two of them are shown on Figs. 4 and 5.

Sequence "A2 RSRP LEAVE - A3 RSRP" is a common 2-gram all over the network and it should occur within all users who are on the move. In the group of abnormal users it exists for all users and the total number of occurrences is 869. However, in the dominance area of cell 28 this sequence occurs far less frequently than for the rest of the network, as it can be seen from Fig. 4a.

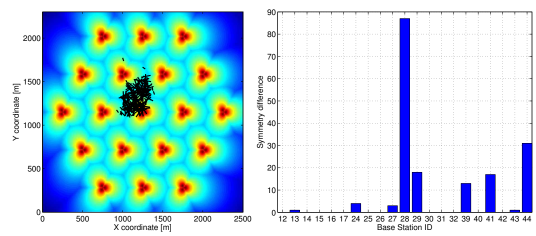
Sequence "HO COMMAND - A2 RSRP ENTER", on the other hand, is a direct consequence of Random Access problem. In normal network behavior "HO COMMAND" should be followed by "HO COMPLETE", but as it can be seen "A2 RSRP ENTER" appears instead. Thus this sequence happens only in the area of problem (in total 126 times), as it can be seen from Fig. 5a.

The described examples of abnormal 2-gram sequences are the only ones among the nine selected 2-grams. To select these subsequences in automatic manner, thus being able to detect sleeping cell, symmetry analysis based on their locations is employed, as described in Section 3.3. As it can be seen from Figs. 4b and 5b there exists a clear unbalance for each of these 2-grams in the dominance areas of cell 28 and also in its neighbor cells 29, 39, 41 and 44. Elsewhere



(a) Locations of this 2-gram sequence
(b) Deviation from the 2-gram symmetry in each cell

Fig. 4. Characteristics of abnormal 2-gram sequence "A2 RSRP LEAVE - A3 RSRP", which is a common 2-gram for all the abnormal calls.



(a) Locations of this 2-gram sequence
(b) Deviation from the 2-gram symmetry in each cell

Fig. 5. Characteristics of abnormal 2-gram sequence "HO COMMAND - A2 RSRP ENTER", which is a common 2-gram for all the abnormal calls.

in the network these 2-grams are more or less in better balance. In fact, the sequence "A2 RSRP LEAVE - A3 RSRP" completes in cell 28 more often than it starts from there and more often than it ends in one of its neighbor cells. Regarding "HO COMMAND - A2 RSRP ENTER" subsequence, it starts more often from cell 28 ending up in one of its neighbors than vice versa. Thus, the symmetry analysis demonstrates that the behavior of cell 28 is clearly abnormal.

6. CONCLUSIONS

In this article advanced data mining framework for the network performance monitoring automation was presented. The considered problem of sleeping cell detection, is among highly complex identification problems as far as there is no direct alarm sent to the operator. A validation of the framework was given in this setting using the random access malfunction as an example for the sleeping cell root cause.

Suggested detection framework is based on such techniques as n-gram analysis, association-based clustering algorithm and dimensionality reduction. Altogether application of these methods in the proposed way on top of MDT data leads to a reliable detection of random access sleeping cell.

References

[1] *Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Self-configuring and self-optimizing network (SON) use cases and solutions (Release 11)*, 3GPP Std. TR 36.902, 2011.

[2] O. Sallent, J. Pérez-Romero, J. Sánchez-González, R. Agustí, M. Díaz-guerra, D. Henche, and D. Paul, "A roadmap from UMTS optimization to LTE self-optimization," *Communications Magazine, IEEE*, vol. 49, no. 6, pp. 172–182, June 2011.

[3] J. Sánchez-González, O. Sallent, J. Pérez-Romero, R. Agustí, M. Díaz-guerra, J. Moreno, and D. Paul, "A new methodology for RF failure detection in UMTS networks," in *Network Operations and Management Symposium, NOMS 2008, IEEE*, April 2008, pp. 718–721.

[4] C. M. Mueller, M. Kaschub, C. Blankenhorn, and S. Wanke, "A cell outage detection algorithm using neighbor cell list reports," *Lecture Notes in Computer Science, Springer Berlin/Heidelberg*, 2008.

[5] M. Amirijoo, L. Jorguseski, T. Kürner, R. Litjens, M. Neuland, L. C. Schmelz, and U. Türke, "Cell outage management in LTE networks," in *Proceedings of the 6th international conference on Symposium on Wireless Communication Systems*, ser. ISWCS'09. IEEE Press, 2009, pp. 600–604.

[6] M. Amirijoo, L. Jorguseski, R. Litjens, and R. Nascimento, "Effectiveness of cell outage compensation in LTE networks," in *Consumer Communications and Networking Conference (CCNC), 2011 IEEE*, January 2011, pp. 642–647.

[7] F. Chernogorov, J. Turkka, T. Ristaniemi, and A. Averbuch, "Detection of sleeping cells in LTE networks using diffusion maps," in *Vehicular Technology Conference (VTC Spring), 2011 IEEE 73rd*, May 2011, pp. 1–5.

[8] J. Turkka, F. Chernogorov, K. Brigatti, T. Ristaniemi, and J. Lempiäinen, "An approach for network outage detection from drive-testing databases," *Journal of Computer Networks and Communications*, 2012.

[9] B. Cheung, S. Fishkin, G. Kumar, and S. Rao, "Method of monitoring wireless network performance," Patent US 2006/0 063 521 A1, 2006.

[10] M. Ganapathiraju, D. Weisser, R. Rosenfeld, J. Carbonell, R. Reddy, and J. Klein-Seetharaman, "Comparative n-gram analysis of whole-genome protein sequences," in *Proceedings of the Human Language Technologies conference*, 2002.

[11] J. Choi, H. Kim, C. Choi, and P. Kim, "Efficient malicious code detection using n-gram analysis and SVM," in *Proceedings of the 2011 14th International Conference on Network-Based Information Systems*, ser. NBIS '11. Washington, DC, USA: IEEE Computer Society, 2011, pp. 618–621.

[12] *Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Resource Control (RRC); Protocol specification (Release 10)*, 3GPP Std. TS 36.331, 2011.

[13] S. Sesia, M. Baker, and I. Toufik, *LTE - The UMTS Long Term Evolution: From Theory to Practice*. John Wiley & Sons, 2011.

[14] Z. He, X. Xu, and S. Deng, "Discovering cluster based local outliers," *Pattern Recognition Letters*, vol. 2003, pp. 9–10, 2003.

[15] K. Brueninghaus, "Link performance models for system level simulations of broadband radio access systems," in *PIMRC-2005*, Berlin, Germany, 2005.

[16] *Technical Specification Group Radio Access Network; Further Advancements for E-UTRA Physical Layer Aspects (Release 9)*, 3GPP Std. TR 36.814, 2010.