Can you hear me now? A call detail record based end-to-end diagnostics system for mobile networks

Anssi Tauriainen
Department of computing science
Aalto University
Helsinki, Finland
anssi.tauriainen@aalto.fi

***Abstract*—Automation of mobile network fault diagnostics and troubleshooting is critical for successful transformation to new network technologies such as 5G and core Network Function Virtualization (NFV). This paper presents a decision tree-based call detail record (CDR) labeling process, which is used to construct an automated end-to-end diagnostics system for mobile network faults. The presented diagnostics system will enable the utilization of automated troubleshooting systems, and the execution of automated corrective actions in third party systems such as Self-Organizing Network (SON) and NFV domain orchestrator.**

***Index Terms*—Mobile network monitoring, service assurance, automated end-to-end diagnostics, decision tree, machine learning, call detail record**

I. Introduction

Automation of the mobile network fault diagnostics and troubleshooting process in mobile networks has recently been recognized as an important topic because of network evolution to 5G [1] and core Network Function Virtualization (NFV) [2]. Increasing mobile traffic and growing complexity of the networks make the Quality of Service (QoS) challenging to monitor and assure. Currently, network fault diagnostics and troubleshooting is mostly manual work and therefore time consuming, error prone and requires substantial domain expertise [3].

Probably the most researched area of network automation is Self-Organizing Networks (SON), which aims to automate Radio Access Networks (RAN) troubleshooting process, containing fault detection, diagnosis, compensation and recovery phases [4]. It also includes network self-healing and self-optimization functionality [5]. However, SON does not encompass diagnostics or self-healing functionality for Extended Packet Core (EPC), IP Multimedia Subsystem (IMS) or Core Networks (CN), which provide essential functionality for voice and other mobile services. SON is based on summarised Key Performance Indicators (KPIs), making it impossible to automatically diagnose, troubleshoot and eventually perform corrections to individual usage events.

Call Detail Records (CDR) contain much more information than KPIs and enable Quality of Service (QoS) diagnostics on a single subscriber level. CDRs can be collected from various standardized interfaces of mobile network, making it possible to construct an end-to-end view to service quality. The challenges of working with CDRs are related to data volumes, data labeling and ambiguity of the information. Assuming that the challenges can be solved, the use of CDRs enable effective utilization of supervised learning techniques to automatically diagnose user QoS issues with much more detailed granularity than with using KPIs [6].

This paper presents a novel method for accurate labeling and resolution of ambiguity of CDR data in mobile networks for voice service. This labeling mechanism is used to construct an automated diagnostics system, which groups the labeled CDR information to QoS KPIs while preserving the relation to individual call details. This will enable further automated troubleshooting and corrective actions by third party systems.

The rest of the paper is organized as follows. Related work is described in Section 2. Section 3 describes the utilization of decision tree for labeling of CDRs, and Section 4 illustrates how automated diagnostics system is constructed. Section 5 presents the training scenarios, Section 6 the test results and conclusions are presented in Section 7.

II. Related work

Most of the literature on automated fault diagnostics and troubleshooting in mobile networks is based on utilization of KPIs and radio networks. However, a little research has focused on use of CDRs.

ARCD [7] is an example of automated diagnostics system that is based on supervised learning process enabled by labeled CDRs. However, the described labeling is rather basic and is based only on standardized success/failure labels of CDRs, and the suggested ARCD system would greatly benefit from more granular CDR labeling.

CDRs are utilized for analysis of user activities and detection of anomalies [8], traffic prediction [9], understanding the calling patterns [10], characterizing mobile application usage [11], [12], modeling metro density [13] and predicting user location [14]. None of these studies discusses CDR-based automated diagnostics or troubleshooting functionality.

An automated root cause system can be also based on call traces [15], which are structurally very similar to CDRs. The main difference is that call traces store only radio measurements and signaling messages, whereas CDRs contain also Circuit and Packet Switched, EPC, IMS and Policy &

978-3-903176-24-9

Charging events, i.e., end-to-end view across the multiple types of networks. The article presents a rule-based classification system, which is based on threshold values and therefore is not suitable for more granular classification.

The combination of measurements from User Equipment (UE) and related radio conditions are used for automated root-cause analysis [16]. The proposed system uses binary classification tree to identify most common radio causes that may impact user throughput. The diagnosed data sessions are aggregated at cell level and compared to reference values to find anomalous cells. This article considers only radio and UE measurements, and does not discuss end-to-end diagnostics.

Another example of automated troubleshooting is a KPI-based, unsupervised root cause analysis system [17], in which the diagnostics process emulates the manual process to ensure accuracy and reliability. This article suggests that the automatic root cause analysis system can utilize supervised learning processes to learn the behavior of faults if accurate information of known network faults is available. The automated diagnostics system presented in this article is able to produce such accurate network fault information and would therefore improve the automated learning of the network fault behavior. Other examples of similar automated diagnostics and troubleshooting systems are presented in [18], [19], [20], [21] and [22].

Decision tree algorithm is utilized for detection of QoS anomalies [23] and network anomaly detection [24]. Neither of the suggested systems consider the classification of CDRs.

III. Labeling of CDRs with decision trees

A. CDR data

CDRs are data records used to log all communications transactions, such as calls or location updates in all parts of the mobile network. A monitoring device creates a separate CDR for each transaction transferred in the connected network interface. A single CDR may contain over one hundred fields of information per transaction. Such information includes event data, such as calling parties, call location, user equipment identification, time, date, call duration and call quality, and network data such as call termination cause, mobile network cell information and identifiers of serving network elements. The structure of CDR is not standardized, and field naming and contents can vary depending on the interface, network vendor and the monitoring device, complicating the correlation of CDRs from different interfaces.

Real-time identification, i.e. labeling of various scenarios within CDR flow is a common challenge in LTE networks [15]. Current monitoring devices usually contain rule-based labeling mechanisms to identify simple fault scenarios, such as handover failures or radio network coverage issues, but the rules cannot cover more complex scenarios since labeling process is executed in real-time. The rules are slow and time-consuming to maintain, and labeling of any new scenario in the network will require code changes to the monitoring device. Labeling of more complex scenarios is usually done off-line in post-processing stage and requires specific analytics functionality.

Adjustable real-time CDR labeling would be highly beneficial for automating the network monitoring and assurance process, because it would enable utilization of supervised learning methods in diagnostics, troubleshooting and definition of next best actions. CDRs also can be used to construct end-to-end view to service quality across various types of networks, which will improve the accuracy of the diagnostics process.

B. Decision tree

Decision tree is a supervised learning algorithm which is well-suited for both regression and classification problems.

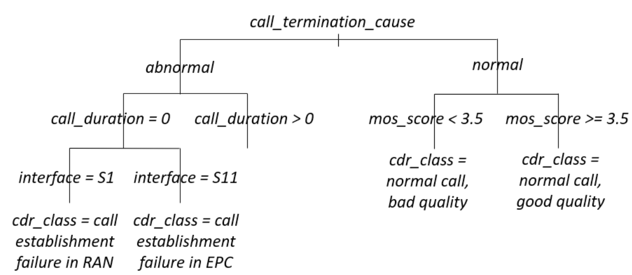

Fig. 1. Example of decision tree for labeling of CDRs.

Decision trees are often used to predict a qualitative response, and it predicts that each observation belongs to the most commonly occurring class of training observations in the region to which it belongs [25]. There are several benefits which make decision tree a good fit for labeling of CDRs:

- Trees can easily handle qualitative predictors such as CDR field values
- The tree construction logic is transparent and easy to explain
- Trees can be displayed graphically and are easily interpreted

Decision tree has a tree-like model structure where each internal node represents a test, each branch represents an outcome of the test, and each leaf node holds the class label. For labeling of CDRs, the tree is trained with the attributes of manually labeled CDR classes, tests will be done on CDR field values and leaf nodes are the CDR classes, i.e., CDR labels.

Example of a decision tree for CDR labeling is presented in Figure 1 above.

CDR labels can be based also on a combination of multiple attributes and contain diagnostic information. For example, a call with normal release and low voice quality score could be labeled as a *successful call with low speech quality*, or the call termination cause and location can be identified in the label (i.e. *call drop due to insufficient IMS capacity*). Some attributes for different CDRs classes are listed later in Section 5.

Once trained, the decision tree uses the attribute information to perform recursive binary splitting to determine the growth of the tree. Either *Gini index* or *entropy* measurement is used to evaluate the quality of the split. The *Gini index* measures the probability of misclassification, and it is defined by a measure of total variance across K classes [25]:

$$G = \sum_{k=1}^{K} p_{mk}(1 - p_{mk} \tag{1}$$

Where p_{mk} represents the proportion of training observations in the mth region that are from the kth class.

Entropy measures the impurity of the node and is defined by

$$D = -\sum_{k=1}^{K} p_{mk} \log p_{mk} \tag{2}$$

As described later in Section 6, the decision trees constructed with the *Gini index* and *entropy* are very similar.

Once the tree is constructed, the prediction accuracy of the decision tree can be tested with a confusion matrix. It is a table which describes the performance of a classification model on a set of test data for which the true values are known. The decision tree *accuracy* is defined by

$$Classification\ accuracy = \frac{correct\ predictions}{total\ predictions} \tag{3}$$

A typical challenge with decision trees is over-fitting, which means that algorithm reduces the errors in classification of the training set at the cost of error rate in classification of the test data. Non-necessary branches are created to reduce the impurity of the samples in each leaf, and in the worst case a leaf node is created for every sample in the data set. This reduces the performance and accuracy of the decision tree, and the algorithm may lose the ability to generalize well to new data sets. Over-fitting can be avoided by pruning the tree after the classification has been done, as discussed later in Section 6.

IV. AUTOMATED DIAGNOSTICS SYSTEM

The goal of network diagnostics is to identify the most likely fault cases among the total traffic. For this, a system utilizing the previously described CDR labeling method is presented. Such system contains CDR collection, tree construction, training, labeling and accuracy analysis phases as presented in Figure 2 below.

This study focuses primarily on voice calls in LTE networks and secondarily on voice call service inter-operability between LTE and UMTS / GSM networks. Voice calls within UMTS or GSM network are not considered.

A. Call types in LTE networks

There are three types of calls in LTE networks

- **Voice over LTE (VoLTE)** call where both users are in LTE network
- **Circuit Switched Fallback (CSFB)** where call is initiated in LTE network is handed over to Circuit Switched 3G or 2G network due to missing VoLTE service [26]
- **Single Radio Voice Call Continuity (SRVCC)** where call is initiated as VoLTE call but then handed over to legacy 3G / 2G network due to missing VoLTE capacity or coverage [27]

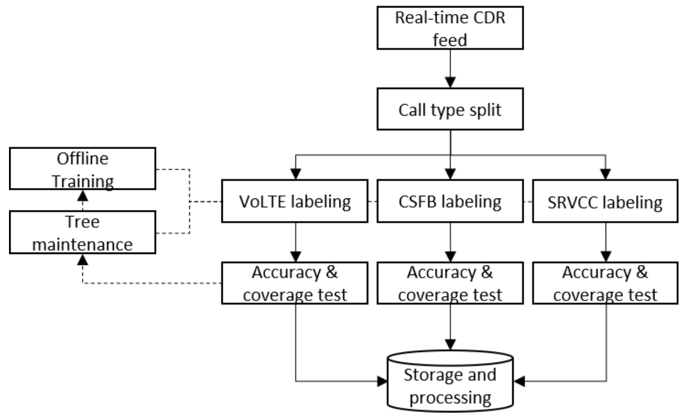

Fig. 2. CDR labeling based diagnostics system.

B. System overview

In order to avoid over-fitting of the decision tree, it is practical to split the CDR labeling phase according to call types and this way reduce the number of trained CDR classes per tree.

As presented later on in next sections, CDR collection and labeling are different for each type of the call.

C. CDR collection

A **VoLTE call** is the simplest scenario to reconstruct and diagnose from CDR information. Figure 3 below describes a simplified mobile originated VoLTE call architecture with potential CDR collection interfaces.

CDRs from S1-MME interface contain control plane VoLTE call related signaling information, such as call result and call termination cause, which are sufficient to determine whether the call was successful. VoLTE call utilizes the capabilities of EPC and IMS, and therefore call quality indicators can be seen only in CDRs containing S11 or Session Initiation Protocol (SIP) messaging.

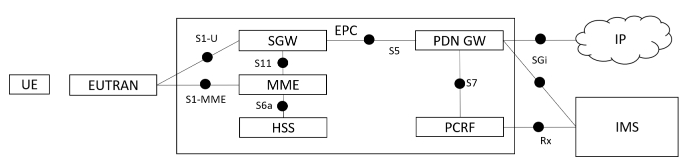

Fig. 3. Simplified LTE call architecture with potential CDR collection points.

When **CSFB call** is triggered, the network moves the User Equipment (UE) temporarily to a legacy circuit switched

TABLE I
SUMMARY OF INTERFACES AND CDR CONTENTS

Interface	CDR contents
S1-MME	Signaling protocols between eNodeB and MME
S1-U	User plane Protocol Data Unit (PDU) details between eNodeB and S-GW
S6a	Authentication information such as location updates and authentication requests
S11	Establishment and deletion of bearer sessions within EPC
S5/S8	User plane tunneling and tunnel management between S-GW and PDN GW
A	2G signalling between GERAN and CS Core
IuCS	3G signalling protocols between UTRAN and CS core
SGi	Between PGw and external network such as Internet

GSM / UMTS network to perform the call. Once the call is completed, the UE moves back to LTE network.

In a similar way than in VoLTE call, the initiation of the call can be analysed from S1-MME CDRs, but once the call has been handed over to GSM / UMTS network, also analysis of corresponding A (GSM) or IuCS (UMTS) CDRs is needed.

Example of CSFB call architecture is presented in Figure 4 above.

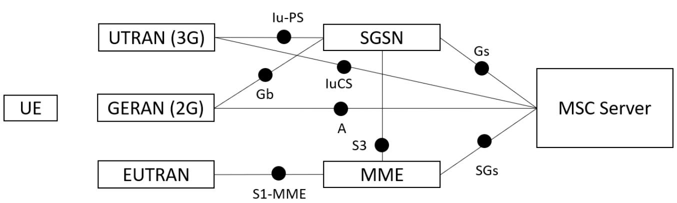

Fig. 4. CSFB call architecture with potential CDR collection points.

A **SRVCC call** is initiated as VoLTE call but then handed over to legacy GSM / UMTS network when it is already active in LTE network. So in addition to CS connection, also EPC and IMS signaling should be analysed in order to diagnose the entire call flow. Example of SRVCC call flow is presented in Figure 5 below.

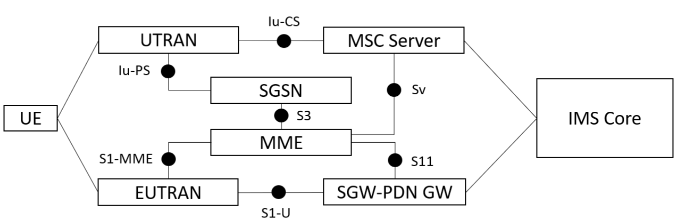

Fig. 5. SRVCC call architecture with potential CDR collection points.

Table 1 above summarizes the CDR collection interfaces and main contents of CDR types.

V. TRAINING OF CDR LABELING SYSTEM

The service quality of mobile network is assured by monitoring KPIs that are grouped to higher level categories such as accessibility, retainability, integrity, availability, mobility, utilization and energy efficiency [28]. The names of KPI categories already indicate the aspect of service quality that they address, and similar grouping is used in this article to group CDR training rules.

The presented list of training scenarios is not meant to be exhaustive and for practical reasons all various combinations cannot be listed. Some examples for accessibility, retainability, integrity and mobility classes are provided instead.

- **Voice accessibility** measures how often the requested network or service can be accessed normally. Call blocking may take place radio or core network and can be identified by looking at the CDR termination cause codes. For example, 'Cell not available' cause code in S1-MME CDRs, or 'EPC bearer termination' cause code in S11 CDRs are such termination causes. IMS blocking can be seen in SIP signaling by looking for failed SIP session initiation.
- **Voice retainability** measures how often voice service is interrupted while call is ongoing. Abnormal interruption in all types of radio networks are similar, and can be identified by looking at cause codes such as 'Radio connection to UE lost' or 'Fail in radio interface' either in S1, IuCS or A interface CDRs.
- **Voice integrity** measures the quality of the service. Mean Opinion Score (MOS) is the most common quality metric in VoLTE, where MOS can be measured from Real-Time Transport (RTP) stream in IMS Mb interface. Another quality measurement for is call setup time, which can be seen directly in IuCS or A interface. For IMS calls is has to be measured from SIP signaling. Integrity can can be measured also indirectly, for example looking for calls with abnormally short duration.
- **Voice mobility** measures how successfully the user sessions are handed over from one cell to another within same network, or from one network to another. In addition to standard RAN handovers, there are also previously mentioned special voice mobility scenarios CSFB and SRVCC.
  In order to correctly diagnose the completion of a CSFB call, the diagnostics system has to analyse the successful completion of call setup steps in each of the CSFB steps listed in Section 4 above. In practice the correlation between S1-MME and either A or IuCS CDRs, depending on the target legacy network type, is needed. It should be noted that even if some of the call mobility events are completed unsuccessfully, it is possible that CS call is still

TABLE II
EXAMPLE TRAINING RULES FOR CDR LABELING

Category	RAN	EPC	IMS	CS Core
Accessibility	Cell not available	S11 bearer creation failure	Failure in SIP initiation	IuCS reloc protocol error
Retainability	Radio connection to UE lost	Abnormal PDN termination	SIP abnormal termination	Release due to CN failure
Integrity	N/A	N/A	Bad quality: MOS score < 3.5	Abnormally short calls
Mobility / CSFB	A or IuCS reloc failure	N/A	N/A	CN acc. or retain. failure
Mobility / SRVCC	S1 relocation to 3G/2G failure	EPC acc. or retain. failure	IMS acc. or retain. failure	CN acc. or retain. failure

complete successfully and end user does not experience any quality degradation.

In order to diagnose SRVCC process correctly, the call flow steps has to be reconstructed from multiple CDRs from S1-U, IuCS / A and Sv interfaces. It should be also noted that single SRVCC call may consist of tens of individual CDRs and each call setup step should be analysed to properly diagnose fault situations. Session Initiation Protocol (SIP), which is used for initiating and terminating VoLTE calls, is fault tolerant and any failed initiation or registration attempt is followed by another attempt which may lead to success. Therefore it is common that a single successful SRVCC call contains also a number of failed SIP events. Usually a call where *call duration > 0*, *MOS > 4* and *termination cause = normal* can be considered successful. In case of unsuccessful calls, it is important to label in which part of the network (RAN, EPC, IMS, CS etc.) the failure happens.

Due to a special nature of these two mobility scenarios, the diagnostics system should contain two separately trained decision trees for analysing CSFB and SRVCC calls. As presented later on in Section 6, this will effectively prevent tree over-fitting and improve overall performance of the diagnostics process.

Table 2 lists the examples of training rules for end-to-end accessibility, retainability, integrity and mobility analysis.

VI. TEST SYSTEM

A test system was constructed to test the accuracy of presented labeling mechanism and diagnostics system. It contains following functions:

1) Input and data preparation
2) CDR labeling
3) Accuracy and coverage testing
4) Output

A. Input and data preparation

The method was tested with anonymized data set that contains 4.35 M CDRs from commercial mobile network. The data set consisted of CDRs from 89 interfaces between LTE, 3G and 2G network elements.

Training data was prepared by labeling manually the extracted VoLTE, CSFB and SRVCC CDRs. The calls were randomly selected and training data for each call type contained approximately 200 - 300 CDRs.

Test data was prepared in a similar way by extracting random samples of each call type, and the test data was manually labeled to test the accuracy of the labeling afterwards.

B. Tree construction

The training was done separately for each call type by feeding the training data to the tree algorithm. In all of the scenarios the labeling was done based on *interface*, *cause code*, *root cause*, *CDR type* and *CDR subtype* fields. For VoLTE and SRVCC calls *MOS indicator* was available and included to the labeling.

The depth of the decision tree was not constrained to see how large the tree will become. Splitting of the tree was based on $entropy$ to reduce the randomness of data in CDR categories. The same trees were also constructed with using $gini$ and it was noticed that while the prediction accuracy remained the same, the depth of the tree grew. This reduces the performance of the classification because more splits are needed.

To analyse the risk of over-fitting, a 4th scenario that combined all 3 calling scenarios was tested. As expected, the decision tree grew larger.

C. Accuracy and coverage testing

The tested CDR samples were all labeled perfectly, and the accuracy remained 1.0 regardless of the test sample size.

The accuracy of labeling is monitored by analysing $entropy$ values of leaf nodes. Increased $entropy$ value indicates that decision tree is not able to classify all provided CDRs properly and as a result, leaf nodes are not homogeneous and contain multiple types of CDRs. As a rule of thumb, $entropy$ values close to 0 with more than 1 sample per leaf is an indication of a well performing tree. Any leaf node with increased $entropy$ value should be analysed further.

In the analysis it was noticed that the test CDRs contained minor variations in the *CDR subtype* field, and the leaf nodes containing such events had increased $entropy$ value. However, the CDRs were correctly labeled based on other CDR field values.

Another explanation for misclassification is previously mentioned over-fitting, in which the algorithm has over-optimized the tree based on training data, ended up reducing the accuracy, performance and adaptability of the classification. The best mechanism to detect over-fitting in this scenario is to monitor the number of leaf nodes and sample size. As a second rule of thumb, if the number of leaf nodes is substantially higher than trained CDR classes, the tree is over-fitted and it needs

TABLE III
CDR DIAGNOSTICS TEST

Call type	CDR Classes	Nodes	Leafs	Depth (entropy)	Depth (gini)	Accuracy
VoLTE	17	41	21	8	12	1.0
CSFB	18	51	26	8	10	1.0
SRVCC	28	67	34	8	13	1.0
Combined	55	121	61	10	13	1.0

to be pruned. Another sign of over-fitting is a high number of leaf nodes with just 1 sample. In case over-fitting is detected, the negative impact can be mitigated with two methods: pre- and post-pruning the tree.

Pre-pruning means that the growth of the tree is restricted to prevent the creation of non-necessary branches. All pre-pruning methods require prior analysis of data sets to stop the tree construction exactly when all CDR classes are learned. This is challenging in real-time CDR classification where the number and the content of the samples are not known beforehand.

In **post-pruning** the decision tree is allowed to classify the training set, and afterwards the non-significant branches are removed to reduce the amount of needed splits in classification. In all of the previously tested labeling scenarios the number of leaf nodes was close to trained CDR classes and no post-pruning is needed.

D. Output

The end result was a CDR test set which was labeled successfully to 55 voice CDR categories according to trained sample values. The results of the labeling tests are presented in Table 3 and a subset of the decision tree is presented in Figure 6 below.

The output provides a number of benefits compared to current monitoring systems:

- Real-time recognition of user impacting call accessibility, retainability and integrity issues in RAN, EPC, IMS or CS Core network. Once identified, such issues can be sent further to SON, NFV domain orcherstrator or other automated management systems for self-optimization and self-healing purposes.
- Automated end-to-end analysis of CSFB and SRVCC call scenarios. Previously the reconstruction of CSFB and SRVCC call scenarios from various signaling events has been mostly manual work. With the proposed diagnostics system, the analysis can be automated and quality issues reported further to SON or NFV orchestrator, or for manual troubleshooting.
- Accurately labeled CDR data enables utilization of more advanced, supervised learning based algorithms for further automated diagnostics and troubleshooting. For example, combined analysis of labeled CDR information and CDR sequence can be used to discover previously unknown quality issues in the network.

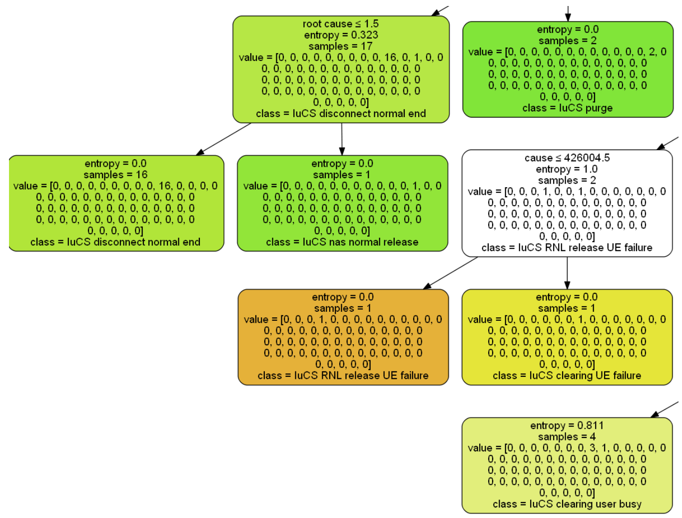

Fig. 6. Subset of CDR labeling tree.

VII. Conclusion

Lack of end-to-end automated diagnostics and troubleshooting system is a major issue in the management of mobile network quality. Existing research is mostly based on RAN quality and summarized KPIs, which do not enable diagnostics or troubleshooting on individual session level.

CDRs are a rich data source which is currently underutilized due to large data volumes and the complexity of the data structure. To solve this, a decision tree based CDR labeling method and a system for automated diagnostics were presented. Decision tree is a simple yet effective method for classification, and manual CDR labeling based training is easy to understand, configure and maintain.

The accuracy of the system was high, and there were no signs of over-fitting when the number of trained CDR classes is kept small. The proposed solution to use separate decision tree for each call type is an effective way to manage the accuracy of the labeling system. Methods for monitoring the prediction accuracy and optimization of the algorithm were also presented and evaluated.

The main benefits of presented automated diagnostics system are enabling of automated corrective actions and utilization of supervised learning based methods for troubleshooting and definition of next best actions.

As a next step, utilization of supervised learning based diagnostics and troubleshooting methods for CDR data should be researched further.

REFERENCES

[1] "Technical Specification Group Services and System Aspects; Service requirements for the 5G system," 3GPP, TS22.261 version 16.7.0 Release 16, 2019.

[2] R. Mijumbi, J. Serrat, J.-L. Gorricho, N. Bouten, F. De Turck and R. Boutaba. "Network Function Virtualization: State-of-the-Art and Research Challenges." IEEE Communications & Tutorials, Vol. 18, No. 1, pp. 236-262, 2016.

[3] A. Celdrán, M. Pérez, F. Clemente and G. Pérez. "Automatic monitoring management for 5G mobile networks," Procedia Computer Science, Vol. 110, pp. 328-335, 2017.

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

[5] "Telecommunication management; Self-Organizing Networks (SON); Concepts and requirements," 3GPP, TS 32.500 version 15.0.0 Release 15, 2018.

[6] C. Zhang, P. Patras and H. Haddadi. "Deep Learning in Mobile and Wireless Networking: A Survey, " IEEE Communications Surveys & Tutorials, March 2019.

[7] M. Mdini, G. Simon, A. Blanc and J. Lecouvre. "ARCD: a Solution for Root Cause Diagnosis in Mobile Networks, " 14th International Conference on Network and Service Management (CNSM), 2018.

[8] S. Parwez, D. Rawat and M. Garuba. "Big data analytics for user-activity analysis and user-anomaly detection in mobile wireless network," IEEE Transactions on Industrial Informatics, Vol. 13, no. 4, pp. 2058-2065, Aug. 2017.

[9] K. Sultan, H. Ali and Z. Zhang. "Call Detail Records Driven Anomaly Detection and Traffic Prediction in Mobile Cellular Networks," IEEE Access, Vol. 6. pp 41728 -41737, 2018.

[10] Z. Aziz and R. Bestak. "Analysis of Call Detail Records of International Voice Traffic in Mobile Networks," International Conference on Ubiquitous and Future Networks, ICUFN, pp. 475 -480. 2018.

[11] Y. Leoa, A. Bussona, C. Sarraute and E. Fleurya. "Call detail records to characterize usages and mobility events of phone users," Computer communications, Vol. 95, pp. 43 -53. 2016.

[12] J. Wu, M. Zeng, X. Hen, Y. Li and D. Jin. "Characterizing and Predicting Individual Traffic Usage of Mobile Application in Cellular Network," Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers, pp. 852 -861. 2018.

[13] V. Liang, R. Ma, W. Ng, L. Wang, M. Winslett, H. Wu, S. Ying, and Z. Zhang. "Mercury: Metro density prediction with recurrent neural network on streaming CDR data, " IEEE 32nd International Conference on Data Engineering, pp. 1374 -1377, 2016.

[14] N. Chen, W. Xie, R. Welsch, K. Larson, and J. Xie. "Comprehensive predictions of tourists' next visit location based on call detail records using machine learning and deep learning methods, " IEEE International Congress on Big Data (BigData Congress), pp. 1–6, 2017.

[15] A. Gomez-Andrades, R. Barco, I. Serrano, P. Delgado, P. Caro-Oliver and P. Muñoz. "Automatic root cause analysis based on traces for LTE self-organizing networks," IEEE Wireless Communications, Vol. 23, pp. 20-23, 2016.

[16] P. Muñoz, R. Barco, E. Cruz, A. Gómez-Andrades, E. J. Khatib, N. Faour. "A method for identifying faulty cells using a classification tree-based UE diagnosis in LTE, " EURASIP Journal on Wireless Communications and Networking, 2017.

[17] A. Gómez-Andrades, P. Muñoz, I. Serrano and R. Barco. "Automatic Root Cause Analysis for LTE Networks Based on Unsupervised Techniques," IEEE Transactions on Vehicular Technology, Vol. 65, No. 4, April 2016.

[18] R. Barco, V. Wille and L. Diez. "System for automated diagnosis in cellular networks based on performance indicators," European Transactions on Telecommunications,Vol. 16, pp. 399–409. 2005.

[19] P. Szilagyi and S. Novaczki. "An Automatic Detection and Diagnosis Framework for Mobile Communication Systems," IEEE Transactions on Network and Service Management, Vol. 9, No. 2, June 2012.

[20] R. Khanafer, B. Solana, J. Triola, R. Barco, L. Moltsen, Z. Altman and P. Lázaro. "Automated Diagnosis for UMTS Networks Using Bayesian Network Approach," IEEE Transactions on Vehicular Technology, Vol. 57, No. 4, July 2008.

[21] R. Khanafer, L. Moltsen, H. Dubreil, Z. Altman and R. Barco. "A Bayesian Approach for Automated Troubleshooting for UMTS Networks," IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC, art. no 4022501, 2006.

[22] R. Barco, L. Diez, V. Wille and P. Lazaro. "Automatic diagnosis of mobile communication networks under imprecise parameters," Expert Systems with Applications, 36(1), pp. 489-500, 2009.

[23] G. Zhu, J. Zan, Y. Yang and X. Qi. "A Supervised Learning Based QoS Assurance Architecture for 5G Networks," IEEE Access, Vol. 7. pp 43598 - 43606, 2019.

[24] "Network anomaly detection by cascading K-means clustering and C4. 5 decision tree algorithm," Procedia Eng., Vol. 30, pp. 174–182, 2012

[25] J. Gareth, D. Witten, T. Hastie, R. Tibrishani. "An Introduction to Statistical Learning: with Applications in R," New York: Springer. pp. 303–336. 2017.

[26] "Digital cellular telecommunications system (Phase 2+); Universal Mobile Telecommunications System (UMTS); LTE; Circuit Switched (CS) fallback in Evolved Packet System (EPS); Stage 2," 3GPP, TS 23.272 version 10.4.0 Release 10, 2011.

[27] "Digital cellular telecommunications system (Phase 2+); Universal Mobile Telecommunications System (UMTS); LTE; Single Radio Voice Call Continuity (SRVCC); Stage 2, " 3GPP TS 23.216 version 11.7.0 Release 11, 2013.

[28] "Technical Specification Group Services and System Aspects; Telecommunication management; Key Performance Indicators (KPI) for Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Requirements (Release 15), " 3GPP TS 32.451 version 15.1.0 Release 15, 2018.