A Knowledge-Based Model for Analyzing GSM Network Performance

Pasi Lehtimäki and Kimmo Raivio

Helsinki University of Technology Laboratory of Computer and Information Science P.O. Box 5400, FIN-02015 HUT, Finland

Abstract. In this paper, a method to analyze GSM network performance on the basis of massive data records and application domain knowledge is presented. The available measurements are divided into variable sets describing the performance of the different subsystems of the GSM network. Simple mathematical models for the subsystems are proposed. The model parameters are estimated from the available data record using quadratic programming. The parameter estimates are used to find the input-output variable pairs involved in the most severe performance degradations. Finally, the resulting variable pairs are visualized as a tree-shaped cause-effect chain in order to allow user friendly analysis of the network performance.

1 Introduction

The radio resource management in current mobile communication networks concentrates on maximizing the number of users for which the services can be provided with required quality, while using only limited amount of resources [4]. Once the network is designed and implemented, the goal is to find a network configuration parameters that use the existing resources as efficiently as possible from the user point of view. In practice, this means that a reasonable tradeoff between the coverage and capacity of the network must be found. Good coverage allows users to initiate services at any location with acceptable service quality, while high capacity allows many network subscribers to use services simultaneously. However, improving the coverage tends to diminish the capacity and vice versa. A good tradeoff between coverage and capacity is obtained when the number of service denials (blocking) and abnormal service interruptions (dropping) are at the minimum, i.e the performance of the network is well optimized.

In this paper, the performance of a GSM network is analyzed based on massive data records and application domain knowledge. Next, the GSM network infrastructure is shortly outlined. In Section 3, a hierarchical model for describing the network performance is proposed. Then, the usage of the proposed model as a part of an analysis process is presented. In Section 5, results of the experiments are presented.



Fig. 1. GSM network architecture.

2 The GSM Network

A GSM network consists of high number of sites, each usually having three base station transceivers (BTS) positioned to cover separate sectors around the site (see Fig. 1). Each BTS has one or more transceiver/receiver pairs (TRX), each allocated on a single physical radio frequency. Base station controller (BSC) manages the operation of several BTSs connected to it through the Abis interface. A single mobile services switching center (MSC) is connected to several BSCs through the A interface.

The performance of the mobile network is measured based on thousands of counters, describing the numbers of the most important events over a measurement period (typically one hour). Due to the high number of counters, a set of high-level key performance indicators (KPIs) are defined by the network manufacturers in order to allow more efficient performance monitoring. Typically, the KPIs describe the success/failure rates of the most important events such as service blocking, service dropping and handovers.

Such indicators are traditionally used in resource management [2] and are well suited for performance monitoring, but there are several drawbacks when they are used in fault diagnosis [3]. Also, the most widely used performance indicators describe the operation of the network at the BTS level. As a result, the performance degradations originating from interaction between several BTSs become very difficult to observe. In many cases, however, the operation of the close-by BTSs are highly dependent on each other. Examples of operation in which several BTSs interact are handovers between close-by BTSs and interference between BTSs having TRXs on the same physical frequency.

3 A Model for GSM Network Performance

The basic idea in the developed model is to measure performance in terms of the number of failed operations in the network. Examples of such operations are an attempt to allocate a signaling channel, an attempt to allocate a traffic channel for a call, or to perform a handover between two neighboring BTSs. In order to obtain good performance, the number of failing operations within the network must be minimized.



Fig. 2. (a) The data set contains 101 variables (x-axis) and 33 subsystems (y-axis). This plot shows the set of input variables (gray) and output variables (black) that belong to each of the subsystems. (b) The subsystem hierarchy. The solid lines indicate that the input of the upper level system contains outputs of the lower level system from the same BTS only. The dashed line indicates, that the input of the upper level system contains input signals from lower level systems of the other BTSs also.

3.1 Model structure selection

The set of available measurements seems to consists of counter groups. Within a counter group, the counters are clearly connected while the counters from different groups are independent on each other. Therefore, the modeling problem is more easily solved by identifying a separate subsystem for each counter group separately. In Figure 2(a), the memberships of the variables in different counter groups (subsystems) are shown. The subsystems tend to form a hierarchical structure (see Figure 2(b)), i.e the outputs of a subsystem describing some lowlevel phenomena can be an input to a higher-level subsystem. Next, the principles used to divide the data generating system into subsystems is described.

User perceived quality The main purpose of the analysis is to locate bottlenecks in the network performance that have direct impact on user perceived quality. The system $S_{1,1}$ describes the number of user perceived quality problems by summing the problems from four different categories, each focusing on different part of the transaction. This model is of type I (see Table 1), in which the number of user perceived quality problems in the whole network is the output variable y(t) and the set of input variables $x_i(t)$ consist of number of blocked channel requests, the number of call setup failures, the number of calls dropped during transaction, and the number of failed handovers, each computed over the whole analyzed network. The contribution of each problem type into the overall user perceived quality is described by the parameter a_i , describing the percentage of the failures of type x_i to the total number of failures y.

-			
Type	Systems	Model	Parameter estimation
Ι	$S_{1,1}, S_{2,1}, S_{2,2}, \\S_{2,3}, S_{2,4}, S_{3,1}, \\S_{3,2}, S_{3,3}$	$y(t) = \sum\nolimits_i x_i(t)$	$a_i = \sum\nolimits_t x_i(t) / \sum\nolimits_t y(t)$
II	$S_{5,7},\ S_{5,8},\ S_{5,9}$	$ \begin{aligned} y(t) &= ax(t) + b \\ &= \begin{bmatrix} a & b \end{bmatrix} \begin{bmatrix} x(t) \\ 1 \end{bmatrix} \\ &= \mathbf{\tilde{a}}^T \mathbf{\tilde{x}}(t) \end{aligned} $	$ \min_{\tilde{\mathbf{a}}} \ \frac{1}{2} \tilde{\mathbf{a}}^T \tilde{\mathbf{X}}^T \tilde{\mathbf{X}} \tilde{\mathbf{a}} - \mathbf{y}^T \tilde{\mathbf{X}} \tilde{\mathbf{a}} \\ \tilde{a}_1 \in [0, 1] \\ \tilde{a}_2 \ge 0 $
III	$egin{array}{llllllllllllllllllllllllllllllllllll$	$y(t) = \mathbf{a}^T \mathbf{x}(t) + b$ = $\begin{bmatrix} \mathbf{a} \ b \end{bmatrix} \begin{bmatrix} \mathbf{x}(t) \\ 1 \end{bmatrix}$ = $\tilde{\mathbf{a}}^T \tilde{\mathbf{x}}(t),$	$\min_{\tilde{\mathbf{a}}} \frac{1}{2} \tilde{\mathbf{a}}^T \tilde{\mathbf{X}}^T \tilde{\mathbf{X}}^T \tilde{\mathbf{X}} \tilde{\mathbf{a}} - \mathbf{y}^T \tilde{\mathbf{X}} \tilde{\mathbf{a}}$ $\sum_{i \neq N} \tilde{a}_i = 1,$ $\tilde{a}_{i \neq N} \in [0, 1]$ $\tilde{a}_N \ge 0$
IV	$egin{array}{llllllllllllllllllllllllllllllllllll$	$\mathbf{y}(t) = \mathbf{x}(t)\mathbf{A},$	$\min_{\mathbf{A}} \sum_{i} \frac{\frac{1}{2} \mathbf{A}_{i}^{T} \mathbf{X}^{T} \mathbf{X} \mathbf{A}_{i} - \mathbf{y}_{i}^{T} \mathbf{X} \mathbf{A}_{i}}{\sum_{a_{ij} \in [0, 1]} a_{ij} \in [0, 1]}$
v	$S_{5,4},S_{5,5},S_{6,8}$	$ \begin{aligned} y(t) &= a_1(c_1(t) + c_2(t) + c_3(t))x_1(t) \\ &+ a_2(c_1(t) + c_2(t) + c_3(t))x_2(t) \\ &+ a_3(c_1(t) + c_2(t) + c_3(t))x_3(t), \end{aligned} $	$\min_{\mathbf{a}} \frac{\frac{1}{2} \mathbf{a}^T \mathbf{X}^T \mathbf{X} \mathbf{a} - \mathbf{y}^T \mathbf{X} \mathbf{a}}{\sum_{i} a_i = 1, \\ a_i \in [0, 1]}$

Table 1. The different types of models for the subsystems.

Blocking The number of blocked channel requests (rejects) measure the networks ability to satisfy the demand generated by the network users. System $S_{2,1}$ with a model of type I describes how many blocked requests y(t) in the network originate from BTS i of the network (the variable $x_i(t)$ is the number of blocked requests in BTS i at time t). The contribution of BTS i to the number of all blocked requests in the analyzed network is described by the parameter a_i .

The system $S_{3,1}$ is described by a model of type I, in which the proportions of stand-alone dedicated control channel (SDCCH) rejects $x_1(t)$, full rate traffic channel (FR-TCH) rejects $x_2(t)$ and half rate traffic channel (HR-TCH) rejects $x_3(t)$ to the total number of rejected channel requests y(t) is computed. This model is estimated for each BTS separately. That is, the parameter a_i describes the contribution of the channel type i to the total number of blocked channel requests.

Finally, the systems $S_{5,7}$, $S_{5,8}$ and $S_{5,9}$ describe the above mentioned channel type rejects due to congestion vs. other possible reasons for blocking. These subsystems are described by a model of type II, in which the output variable y(t) describes the number of blocked requests and the input variable $x(t) = C(t)R_{tot}(t)$ describes the proportion of channel requests assumed to have occurred during congestion (C(t) denotes the percentage of time in congestion in time period tand $R_{tot}(t)$ is the total number of channel requests). These models include a bias b since it is not expected that all request rejects are due to congestion, but also other causes may exist (but measurements are not available). The minimization of the mean square prediction error $\frac{1}{T}\sum_{t} e^2(t) = \frac{1}{T}\sum_{t} (y(t) - \hat{y}(t))^2$ with the corresponding constraints (see Table 1) leads to a standard quadratic programming problem. For more information about algorithms to solve such problems, see [1]. **Call setup failures** It is possible, that the user request is not served due to problems in the resource allocation phase (call setup) of the transaction. As in the case of service blocking, a model describing the contributions of each BTS to the total number of call setup failures in the network is defined (model $S_{2,2}$). Basically, the call setup phase includes allocation of a signaling channel in which the negotiation for the actual traffic channel is performed. Model $S_{3,2}$ of type II divides the call setup failures in a single BTS into SDCCH and TCH setup failures separately.

Both the SDCCH and TCH signaling may fail due to problems in different network elements (or interfaces between them) involving in the signaling procedure. The number of SDCCH and TCH signaling failures due to problems in different network elements are described by the models $S_{5,10}$ and $S_{5,11}$, respectively. These models are of type *III* and include a bias since it is possible, that call setup failures are caused by other reasons for which measurements are not available. Also, an equality constraint is introduced since it is necessary to require that a single failure in a network element causes the failure of the call setup phase of exactly one transaction.

The most common reasons for failures during the call setup phase or actual service are the inadequate radio signal propagation conditions (problems in radio channel in the air interface). The failures in radio channel are usually due to bad signal quality, i.e the transmitted data includes too many bit errors. The models $S_{6,9}$ and $S_{6,10}$ of type *III* describe the number of SDCCH and TCH radio channel failures due to bad signal quality vs. other reasons.

The radio signal quality is mostly affected by two components. Firstly, the propagation environment causes attenuation to the transmitted radio signal due to path loss, shadow fading and multipath fading. Secondly, the radio signal may be attenuated by the other radio signals originating from other BTSs having a TRX on the same physical frequency (interference). The purpose of the models $S_{7,2}$ and $S_{7,3}$ of type *III* is to compute how many bit errors in uplink (from MS to BTS) and downlink (from BTS to MS) traffic are due to difficult propagation conditions in the BTS's coverage area and how many are likely the result of interference.

Call dropping When the call setup phase is successfully completed, the actual service (usually speech in GSM networks) is started. However, the service may be abnormally interrupted (dropped) due to several reasons. The purpose of the model $S_{2,3}$ is to describe how many calls are dropped in BTS *i* w.r.t the number of dropped calls in the whole analyzed network. The model is of type *I* and the contributions of each BTSs to the total number of dropped calls is described by the parameter a_i (similarly to the blocking and call setup failures).

The call may be dropped due to internal failures in the network elements or interfaces between network elements. The purpose of the model $S_{5,12}$ is to describe the contributions of the possible causes (very similar to the causes of call setup failures). This model is of type *III*, having a bias since a call may be dropped due to reasons for which measurements are not available. As in the case of call setup failures, most of the dropped calls are expected to be due to radio channel (air interface) problems. Therefore, the same models explaining the number of call setup failures due to bad signal quality also describe the number of dropped calls due to radio channel problems.

Handover failures As in the previous cases, also the (outgoing) handover failures (HO) are divided into network level variable and BTS level variables. The model $S_{2,4}$ is used to compute the value for parameters a_i describing the percentage of handover failures originating from BTS *i*. The model $S_{3,3}$ divides the handover failures according to the handover type (within BTS HO, BSC controlled HO and MSC controlled HO), and a model of type *I* is used to obtain the contributions of the different handover types into the total number of handover failures.

Both the BSC and MSC controlled outgoing handovers may fail due to problems in the source (serving) BTS or the target BTS. The serving BTS problems can be various BSS problems (very rare in practice) and the target BTS problems may be due to lack of resources, BSS level problems (rare) or problems with the connection (radio link) to the target BTS. Models $S_{4,2}$ and $S_{4,3}$ explicitly describe the dependencies between these failures in different BTSs, i.e the cause for failed outgoing handover may be in lack of resources or connection in any of the BTSs around the same operation area. Model $S_{4,1}$ describes the within BTS handover failures due to BSS problems or lack of resources. These three models are of type IV in which both the equality constraint and box constraints for the parameters are used.

Model $S_{5,2}$ describes the causes for the target BTS radio channel failures and the model $S_{5,3}$ describes the reasons for BSS problems in the target BTS that caused the failed outgoing handover from the serving BTS. Models $S_{5,4}$ and $S_{5,5}$ describe the causes for the failing BSC and MSC controlled outgoing handovers due to lack of resources in the target BTS, respectively. Similarly, model $S_{6,8}$ describes the causes of the lack of resources in within BTS HO attempts. The purpose of these three latter models is to analyze the number of HO failures per HO type (SDCCH-SDCCH, SDCCH-TCH, TCH-TCH) due to SDCCH, HR-TCH and FR-TCH congestion vs. other unmeasured causes. These models are of type V, and contain signals $c_1(t)$, $c_2(t)$ and $c_3(t)$ that describe the percentages of SDCCH, FR-TCH and HR-TCH congestion w.r.t the length of the measurement period (one hour in our case) and $x_i(t)$ denotes the number of HO attempts.

Since both the BSC and MSC controlled handovers may be of one of the three above mentioned types, there are six different kind of handovers involving distinct target and source BTS. Models at the level six all describe the percentage of incoming handovers originating from different source BTSs, one model per each HO type. These models can be used to find out which close-by BTSs are generating the major portion of the handover load to the target BTS during handovers. Finally, the model $S_{7,1}$ describes the causes for BSC controlled TCH-TCH handovers (most typical type of handover).

4 Model Based Analysis Process

4.1 Preprocessing

In order to estimate all subsystem models, the data must be carefully preprocessed. In this work, the preprocessing phase includes outlier removing, data segmentation and constant variable pruning. The number of outliers (points clearly differing from other measurements) is quite high in this type of application. Such samples are generated during network reconfigurations or hardware breakdowns, typically lasting only few hours. Such time points should be removed since they do not help in finding the major bottlenecks in network performance. In TRX level model construction, it is also necessary to segmentate the data into subsegments (time periods) during which the number of TRXs on a certain physical frequency do not change.

After outlier detection and data segmentation, the model parameters can be estimated from the data. However, in some network elements rarely suffering from any types of failures, some or most of the signals in the model are nearly constant. In such a case, the data is not rich enough for estimating a model. Instead, the (nearly) constant input signals are pruned before the model is estimated. In the case of constant signal being an output variable, the model is not estimated at all.

4.2 Visualization of the dependencies

After the data is cleaned, the models can be estimated using standard quadratic programming techniques. After the parameters of the subsystem models have been estimated, an item to a dependency list is generated per each input-output variable pair. Each item in the dependency list include the strength of the dependency between the input and output variable (the value of the parameter a), a measure of model accuracy (root mean square prediction error (RMSE) of the model), and a measure of models importance in overall network performance analysis (the average number of failures stored in the output variable of the input-output variable pair).

After all the models have been estimated and the properties of each inputoutput variable pairs are stored in the dependency list, a tree-shaped graph is constructed in order to analyze the cause-effect chains generating the major performance degradations of the network. Since the number of theoretically possible dependencies is extremely large, only the most important dependencies are included to the dependency tree.

Three criteria are used to prune uninteresting dependencies from the tree. Firstly, the model accuracy from which the dependency originates must be at a reasonable level. Otherwise, the analysis might be mislead by very inaccurate models having large values for parameter a (which is forced in several models due to the equality constraints for the parameter vector). Secondly, the output variable of the dependency must be interesting enough (i.e relatively large number of failures must be observed in the output variable). Finally, only the dependencies that belong to the cause-effect chains contributing most to the overall network performance degradations are included into the dependency tree. For each subsystem, different minimum and maximum values for strength of dependency, model accuracy and model interestingness are defined.

5 Experiments

The analyzed GSM network data contained 120 BTSs, in which 101 most important variables (counters) were measured during a two-month time period. Since the variables were divided into 33 subsystems consisting of 5 network level systems, 24 BTS level systems, 4 TRX level systems and 143 frequency related segments with non-changing frequency plan, we have $5 + (120 \times 24) + (143 \times 4)$ models with a corresponding parameter vector **a** estimated using quadratic programming.

Figure 3(a) shows the root mean square errors (RMSE) of the models. Note, that the RMSE of the models can be interpreted as the number of user perceived quality problems that could not be explained by the model. Models in which the RMSE is below ~ 50 failures can be regarded as accurate enough in order to make useful inferences about the data. Clearly, there are lots of models that are accurate enough, but also many models are not accurate enough to allow any justified conclusions to be made. Also, three models seem to be very inaccurate.

Figure 3(b) shows the means of the output variables of the models. These values measure the number of user perceived failures of the subsystems. Therefore, it can be regarded as a measure of interestingness or importance of the subsystem in the performance analysis.



Fig. 3. (a) The prediction errors (RMSE) of the subsystems. (b) The interestingness (cost) of the subsystems.

Figures 4(a)-(d) show the pruned dependency trees in four separate cases. In Figure 4(a), the cause-effect chains of the most significant blocking problems are shown. Clearly, there are 4 BTSs (6,11,18,85) that suffer from lack of resources. BTSs (6,11,18) suffer from lack of half rate traffic channels and BTS 85 suffers

from lack of full rate traffic channels. Only BTS 6, the causes for blocking can be said to result regularly from congestion.

In Figure 4(b), the results of the corresponding analysis for the call setup failures are shown. Here, four BTSs (17,52,66,74) seem to suffer from call setup failures regularly. In all these four BTSs, the failures tend originate during SD-CCH signaling and fail due to radio link problems. In BTSs 17 and 74 the radio link failures can be said to result from bad downlink signal quality and in BTSs 52 and 66 they are due to bad downlink signal quality.

Figure 4(c) shows the corresponding results for call dropping problems. Again, the cause-effect chains for describing the reasons for dropped calls in four BTSs (6,7,69,74) are shown. The reasons for call dropping seem to be radio link failures. In BTSs 6, 7 and 69 the radio failures are likely due to bad signal quality in uplink. In two TRXs of BTS 6 and in one TRX of BTS 69 the number of bit errors seem to correlate with the amount of traffic in both the own BTS as well as in interfering TRXs on the same radio frequency.

Finally, in Figure 4(d) the analysis of the handover problem sources are shown. The results for three BTSs (7,9,90) having the worst handover performance are shown, indicating that the problems are in BSC controlled outgoing handovers. In BTSs 7 and 9, the problems seem to be in lack of resources in the target BTSs (6,11,85). In target BTS 6 suffering from lack of resources, there seems to be high amount of incoming TCH-TCH handover attempts from BTS 7. This same BTS tend to cause problems also for target BTS 11 suffering from lack of resources during handover attempts. For target BTS 11, two other BTSs (12,18) are found that can be said to generate high number of handover attempts. The handover attempts of these two BTSs are due to very similar reasons: the quality of the uplink and downlink radio connection and the uplink and downlink signal strength are not reasonable in these two BTSs, and some of the users are switched into more appropriate BTSs. Also, significant number of users are switched to another BTS in order to minimize the energy consumption of the MSs (power budget).

6 Conclusions

In this paper, a knowledge-based model for analyzing the performance of the GSM network is presented. The presented model is based on subsystem hierarchy that was developed during an explorative data analysis process involving iterative literature research and data visualization. As a result, a logical division of the system into subsystems with the corresponding input and output variables were established. Also, the type of the model per subsystem required a priori knowledge about the semantics of the subsystem variables.

A data record from an operational GSM network was used to estimate the parameters of the subsystems. The estimated parameters were interpreted to describe the strength of dependency between input-output variable pairs. After parameter estimation, the most important input-output variable pairs were analyzed further by constructing a hierarchical dependency tree. The depen-



Fig. 4. The analysis of the four main components of the user perceived quality. (a) Cause-effect chains for blocking of services, (b) call setup failures, (c) call dropping and (d) handover failures.

dency tree was constructed for four major problem types in order to analyze the cause-effect chains generating the user perceived quality problems. The provided information can be used to enhance the current radio resource usage in the network.

Acknowledgment

Nokia Foundation is gratefully acknowledged for their financial support.

References

- 1. S. Mokhtar Bazaraa, D. Hanif Sherali, and C. M. Shetty. *Nonlinear Programming:* theory and algorithms. John Wiley and Sons, Inc., 1993.
- Sofoklis A. Kyriazakos and George T. Karetsos. Practical Radio Resource Management in Wireless Systems. Artech House, Inc., 2004.
- 3. Pasi Lehtimäki and Kimmo Raivio. A SOM based approach for visualization of GSM network performance data. In *Proceedings of the 18th Internation Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems (to appear)*, 2005.
- 4. Jens Zander. Radio Resource Management for Wireless Networks. Artech House, Inc., 2001.