Publication VI

© Pasi Lehtimäki.
Combining Measurement Data and Erlang-B Formula for Blocking Prediction in GSM Networks

Pasi LEHTIMÄKI a, 1, Kimmo RAIWIO a

a Adaptive Informatics Research Centre, Helsinki University of Technology, Finland

Abstract. In cellular network performance optimization, it is important to be able to predict the amount of user-experienced quality problems such as call blocking with alternative modifications to the network configuration. In this paper, a method based on mathematical optimization and knowledge representation to predict the amount of blocking in GSM networks is presented. The method is based on exploiting the available statistical measurement information describing the individual characteristics of the network elements. In addition, application domain knowledge about blocking is included to the model by using the well known Erlang-B formula to establish a mapping between blocking and existing network measurements. The results of the experiments show that the proposed method performs better than the comparison method based on basic Erlang-B formula and raw measurement data.

Keywords. Blocking prediction, Erlang-B formula, GSM network, measurement data, radio resource optimization.

Introduction

In cellular networks, the optimization of radio resources becomes necessary when the number of subscribers increases while the amount of resources remains limited. When optimizing the radio resource usage, the number of subscribers for which the quality of service requirements are met is maximized in order to gain maximal profit for the operators with minimum costs [7]. One of the most important quality of service indicators in GSM networks is service (call) blocking. The maximization of the number of subscribers that can be served corresponds to the minimization of service blocking, meaning that as many subscribers as possible should be able to access the network without service denial.

The amount of blocking can be reduced by changing the capacity of the network in a way that more communication channels are available in locations with higher user density. In addition, it is possible to adjust the size of the Base Transceiver Station (BTS) coverage areas (cells) and therefore, to control the amount of input traffic in the cells. In both cases, it would be appropriate to be able to predict the amount of blocking if the cell sizes (the amount of input traffic) or the number of communication channels are modified.

1Corresponding Author: Pasi Lehtimäki, Adaptive Informatics Research Centre, Helsinki University of Technology, P.O. Box 5400, FIN-02015 HUT, Finland; E-mail: Pasi.Lehtimaki@hut.fi
In this paper, an artificial intelligence (AI) based approach to predict the amount of blocking is presented. The proposed approach combines the benefits of data analysis methods and knowledge representation techniques. Past measurements made from the BTSs are an important source of information describing the unique properties of the BTSs such as the propagation environment and network configuration. In order to make predictions in cells without past capacity problems, the well-known Erlang-B formula is included to the model that allows the establishment of mapping between blocking, number of channel requests and the number of busy communication channels for which good statistical data records are available.

Next, the use of traditional Erlang-B formula in blocking prediction is presented and its drawbacks are demonstrated. In Section 2, the proposed method combining the Erlang-B formula with the measurement data is presented. Finally, the results of the experiments with artificial and real data sets are presented in Section 3.

1. Traditional Blocking Modeling

1.1. Erlang’s loss system

Traditionally, the Erlang-B formula is used as a model when computing the amount of blocking with different number of channels and demand [2]. When the incoming transactions follow the Poisson arrival process with arrival rate $\lambda$, transaction length is exponentially distributed with mean $1/\mu$ and the number of channels $N_c$ is finite, the probability that $n$ channels are busy at random point of time can be computed using the Erlang-distribution

$$p(n|\lambda, \mu, N_c) = \frac{(\lambda/\mu)^n / n}{\sum_{k=0}^{N_c} (\lambda/\mu)^k / k}.$$

Using this formula, the expected value for the number of blocked requests is obtained by multiplying the number of arriving requests with the blocking probability, leading to $B = \lambda p(N_c|\lambda, \mu, N_c)$. The expected value for the congestion time is $C = p(N_c|\lambda, \mu, N_c)$ and the expected value for the number of busy channels is $M = \sum_{n=0}^{N_c} np(n|\lambda, \mu, N_c)$.

The Erlang-B formula can be applied in long-term blocking prediction if reasonable predictions for the number of requests, average channel hold time and number of channels are available. The most straightforward method would be based on hourly averages of number of requests and channel hold time in past observations. More precisely, the average value $\mu_{R,h}$ for the number of requests $R$ at hour $h$ could be estimated from $N$ past observations from hour $h$ using $\mu_{R,h} = \frac{1}{N} \sum_{i=1}^{N} R_{h,i}$ and similarly, the mean $\mu_{T,h}$ for the channel hold time $T$ at hour $h$ is estimated using $\mu_{T,h} = \frac{1}{N} \sum_{i=1}^{N} T_{h,i}$.

1.2. Drawbacks of using Erlang-B formula in long-term prediction

The main problem with the use of the previously discussed approach in capacity optimization is that the theoretical predictions for the number of blocked requests are not well describing the blocking measurements made in operative GSM networks. The reason may be in stochastic nature of the capacity due to the different priorities of the trans-
actions [6] or variating propagation conditions [3]. In addition, each BTS has unique software and hardware configuration as well as resource allocation algorithms having an influence on blocking behavior.

The amount of variation in transaction arrival rate during the measurement period has a strong impact on blocking. Figure 1 demonstrates the connection between the number of requests $R$ in hour, the number of blocked requests in hour and the blocking percentage. The number of requests $R$ per hour takes values from 500 to 8000. The time period of one hour was divided into 180 segments of length 20 seconds. For each value of $R$, different distributions for the number of channel requests per segment were simulated. Then, the Erlang-B formula was used to compute the segmentwise values for the number of blocked requests, and finally, the sum of the segmentwise values were obtained. The number of channels was 16 and the average channel hold time was 5 seconds. Therefore, this is a typical setting for the analysis of the capacity of a BTS with 16 Standalone Dedicated Control CHannels (SDCCHs).

![Figure 1](image)

Figure 1. The effect of variability in arrival rate to blocking. Clearly, the more variation in arrival rate is present, the more blocking occurs.

The solid line shows the cumulative values over one hour for the number of requests and blocked requests when the arrival rate of incoming channel requests during the segment $i$ of each hour is $\lambda(i) = R/180$. In other words, this means that the arrival rate is constant during the one hour time period. The dashed line shows the corresponding results in the case when the arrival rate during the segments is Normally distributed around the mean $\lambda(i) \sim N(\mu = R/180, \sigma = 5)$. This leads to the approximately same number of requests in hour, but now, there are also segments during which the arrival rate exceeds the mean of the arrival rate and are more likely generating also blocked requests. The dot-dashed line and the dotted lines show the results for the cases when the arrival rate during segments are Normally distributed with standard deviations $\sigma = 10$ and $\sigma = 15$. It is clear, that the more variation exists in arrival rates between segments, the more blocking can be expected to occur. The typical target value for SDCCH blocking probability is usually 0.2%.

2. Extended Blocking Model

2.1. The model

Now, suppose that the time period of one hour is divided into $N_s$ segments of equal length. Also, assume that we have a vector $\Lambda = [0 \ 1\Delta_{\lambda} \ 2\Delta_{\lambda} \ \ldots \ \ (N_s - 1)\Delta_{\lambda}]$ of $N_{\lambda}$ possible arrival rates per segment with discretization step $\Delta_{\lambda}$. Let us denote the number of blocked requests during a segment with arrival rate $\Lambda_i$ with $B_i =$
\( p(N_c|\Lambda_i, \mu, N_c)\Lambda_i, \) where \( p(N_c|\Lambda_i, \mu, N_c) \) is the blocking probability given by the Erlang-distribution. Also, the congestion time and the average number of busy channels during a segment with arrival rate \( \Lambda_i \) are denoted with \( C_i = p(N_c|\Lambda_i, \mu, N_c) \) and \( M_i = \sum_{n=0}^{N_c} np(n|\Lambda_i, \mu, N_c) . \) In other words, the segment-wise values for blocked requests, congestion time and average number of busy channels are based on basic Erlang-B formula. For short time segments, the assumptions related to Erlang-B formula can be said to hold.

Now, assume that the number of segments with arrival rate \( \Lambda_i \) is \( a_i \) and \( \sum_i a_i = N_s \). Then, predictions for the cumulative values over one hour for the number of requests \( R \), blocked requests \( B \), congestion time \( C \) and average number of busy channels \( M \) can be computed with

\[
\begin{bmatrix}
\Lambda_1 & \Lambda_2 & \ldots & \Lambda_{N_s} \\
B_1 & B_2 & \ldots & B_{N_s} \\
C_1/N_s & C_2/N_s & \ldots & C_{N_s}/N_s \\
M_1/N_s & M_2/N_s & \ldots & M_{N_s}/N_s \\
\end{bmatrix}
\begin{bmatrix}
a_1 \\
a_2 \\
\vdots \\
a_{N_s} \\
\end{bmatrix}
= 
\begin{bmatrix}
\hat{R} \\
\hat{B} \\
\hat{C} \\
\hat{M} \\
\end{bmatrix}
\]

(2)

or in matrix notation \( X a = \hat{Y} \). Note, that it is necessary to divide the last two rows of matrix \( X \) with \( N_s \), since the congestion time calculation in matrix \( X \) is based on time unit corresponding to the length of the segment instead of the whole time period consisting of \( N_s \) segments. Similarly, \( M \) denotes the average number of busy channels during the whole time period of \( N_s \) segments, while \( \sum_i M_i a_i \) would indicate the sum of segment-wise averages of number of busy channels.

2.2. Estimation

Now, the problem is that the vector \( a \) is unknown and it must be estimated from data using the observations of \( Y \) and matrix \( X \) which can be constructed by applying the Erlang-B formula to the number of channels \( N_c \), observed average channel hold time \( \mu \) and the selected \( \Lambda \) vector. The most common procedure for parameter estimation in regression problems is the minimization of prediction error. Let us denote the prediction error between the observed output vector \( Y \) and the corresponding predictions \( \hat{Y} \) with \( e = Y - \hat{Y} = Y - X a \). Since the output vector \( Y \) includes variables that are measured in different scales, it is necessary to include weighting of variables into the cost function. Now, the error vector becomes

\[
e_w = 
\begin{bmatrix}
w_1 & 0 & \ldots & 0 \\
0 & w_2 & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & w_n \\
\end{bmatrix}
\begin{bmatrix}
Y - \hat{Y} \\
\end{bmatrix}
= \text{We}
\]

(3)

and the cost function to be minimized is

\[
E_w = \frac{1}{2} e^T_w e_w = \frac{1}{2} Y^T W^T WY - Y^T W^T WXa + \frac{1}{2} a^T X^T W^T WXa.
\]

(4)
Since the first term does not depend on variable $a$ to be estimated, it can be ignored from the cost function. Denoting $f = -X^T W^T W Y$ and $H = X^T W^T W$, the optimization problem can be represented in the form

$$\min_a \left\{ \frac{1}{2} a^T H a + f^T a \right\}, \quad \text{w.r.t} \quad 0 \leq a_i \leq N_s, \sum_i a_i = N_s. \quad (5)$$

In other words, the goal is to find the vector $a$ that provides the smallest prediction errors for variables $R$, $B$, $C$ and $M$. This is the form of standard quadratic programming problem and there are several algorithms and programming packages that can be used to solve the parameter estimate [1].

2.3. A probabilistic model for arrival rate

Suppose that we have $N_d$ observations for variables $R$, $B$, $C$ and $M$, all measured at hour $h$ during different days. The optimization problem described in the previous section could be solved for each of the $N_d$ observation vectors separately, leading to $N_d$ solution vectors $a$ for hour $h$. Let us denote the $i$th solution vector for hour $h$ with $a^{(i)}_h$ and the $j$th element of the corresponding solution vector with $a^{(i)}_{j,h}$. Since $a^{(i)}_{j,h}$ described the number of segments with arrival rate $\Lambda_j$ during $i$th observation vector at hour $h$, the probability for a random segment during $i$th observation period to have an arrival rate $\Lambda_j$ can be computed from $a^{(i)}_{j,h}$ with $p^{(i)}_{j,h} = a^{(i)}_{j,h}/N_s$, where $N_s$ is the number of segments.

Now, if we are interested in occurrences of $\Lambda_j$ at hour $h$ in the long run, it would be straightforward to sum the occurrences of the $\Lambda_j$ during the $N_d$ observations. In other words, the probability for observing a segment with arrival rate $\Lambda_j$ at hour $h$ would become $p_{j,h} = \frac{1}{N_d N_s} \sum_{i=1}^{N_d} a^{(i)}_{j,h}$. Now, the probabilities for each arrival rate $\Lambda_j$ at hour $h$ form a probabilistic model for arrival rate.

The above probabilistic model can be used to compute the elements of the average model $\bar{a}$ using $\bar{a}_{j,h} = p_{j,h} N_s$. In other words, the average model includes the expected number of occurrences of $\Lambda_j$ at hour $h$ during $N_s$ segments. Now, this average model can directly be used in prediction of the output variables of $Y$ given the selected $X$ matrix. In particular, we are interested in prediction of number of blocked requests $B$ in the output vector $Y$. From now on, we refer to this method as the Erlang-B with varying arrival rate (EBV$\lambda$) and the method presented in Section 1.1 as the Erlang-B with constant arrival rate (EBC$\lambda$).

3. Experiments

3.1. Artificial data sets

In order to study the performance of the proposed method, four different data sets were generated. The number of channel requests per hour takes values $R = [1000, 8000]$, the number of channels is 16 and the average channel hold time was 5 seconds. In the first data set, all the requests appeared with constant arrival rate during each of the $N_s = 180$ segments for all values of $R$. In the remaining three data sets, the arrival rates of the segments are drawn from Normal distributions with mean $\mu_\lambda = R/N_s$ and standard de-
Table 1. The prediction errors of the EBC_λ and EBV_λ methods with the artificial data sets.

<table>
<thead>
<tr>
<th>p(λ)</th>
<th>EBC_λ</th>
<th>EBV_λ</th>
<th>EBV_λ</th>
<th>EBV_λ</th>
<th>EBV_λ</th>
<th>EBV_λ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>0</td>
<td>2.54</td>
<td>0.436</td>
<td>0.201</td>
<td>0.126</td>
<td>0.124</td>
</tr>
<tr>
<td>N(R/N_s, 1)</td>
<td>0.739</td>
<td>2.50</td>
<td>0.411</td>
<td>0.182</td>
<td>0.114</td>
<td>0.100</td>
</tr>
<tr>
<td>N(R/N_s, 3)</td>
<td>5.41</td>
<td>1.92</td>
<td>0.12</td>
<td>0.0668</td>
<td>0.0666</td>
<td>0.0666</td>
</tr>
<tr>
<td>N(R/N_s, 5)</td>
<td>14.9</td>
<td>1.07</td>
<td>0.0751</td>
<td>0.0739</td>
<td>0.072</td>
<td>0.0725</td>
</tr>
</tbody>
</table>

N(λ) = 5, 10, 15, 20, 25, 30 then, the Erlang-B formula was used to compute the segmentwise values of R, B, C and M. Finally, the sum of the segmentwise values for R, B, C and M were obtained to correspond data from one hour time periods. This hourwise data was used as input data for the EBV_λ method. The predictions for B of the EBV_λ were compared with the corresponding predictions given by the EBC_λ method for which the total number of requests R during N_s segments was used as the only input.

Table 1 demonstrates the performance of EBC_λ as well the EBV_λ method with different values of N_λ. Clearly, the prediction errors of the EBC_λ method that assumes constant arrival rate get worse as the amount of variation in arrival rate increases. Secondly, the prediction errors of the EBV_λ method assuming varying arrival rate performs better as the number of values for arrival rate N_λ increases. Although, no significant improvements for N_λ ≥ 15 are achieved. Thirdly, the prediction accuracies of the EBV_λ for certain value of N_λ become better as the amount of deviation in arrival rate increases. In other words, data with small variation in arrival rate seems to be difficult to model for the EBV_λ method. Most likely, this is a result of bad matching between the true arrival rate in the data and the closest candidate λ_j in the arrival rate vector Λ.

3.2. Real GSM network data

In addition to the self-generated data sets, we have made experiments with measurement data from an operational GSM network consisting of hundreds of measurements over 40 day time period in 2500 BTSs. For this study, we have chosen BTSs with different capacities that include the most serious blocking problems. Before the raw data can be used in model estimation, some preprocessing was performed in order to prevent the use of data from abnormal operational conditions. The preprocessed data set was divided into two proportions. The first half of the data was used as a training data for model estimation. The second half of the data was used as a validation data.

3.3. Error measures

In this section, the performance measures used to compare the EBC_λ and EBV_λ methods in long-term prediction of blocking are introduced. Let us assume, that we have the observations of R, B, C and M from a single time period of one hour. The quadratic programming techniques can be used to find the instantaneous parameter estimate a that provides the smallest prediction error with the selected X matrix.

The ability of the proposed method to explain the instantaneous values of the output variable B in vector Y in the training data with 24 × N_d samples can be measured with

$$E^{trn}_{a_h} = \frac{1}{24N_d} \sum_{h=0}^{23} \sum_{i=1}^{N_d} \left| B_h^{(i)} - \hat{B}_{a_h}^{(i)} \right|$$

where the subscript a_h(i) of the error measure is used to denote that the prediction \( \hat{B} \) is obtained with a estimated from the i-th observa-
tion from hour $h$ using the EBV$_{\lambda}$ method. $E_{ebc}^{trn}$ indicates the performance of the EBC$_{\lambda}$ method in which the Erlang-B formula is directly applied to the observations of number of requests and average channel hold time in the training data. In other words, it is similar to the above error measure given that the number of blocked requests is given by $\hat{B}_h^{(i)}$.

Since we are interested in long-term prediction and the cumulative behavior of the prediction errors, the hourwise validation error $H_{ebc}^{val} = \frac{1}{24N_d} \sum_{h=0}^{23} \left| \sum_{i=1}^{N_d} B_h^{(i)} - N_d \hat{B}_h \right|$ was also computed. This error measure allows the analysis of how the prediction $\hat{B}$ fits the sum of blocked requests given the hour $h$. The same measure can be used with EBC$_{\lambda}$ by replacing the output variable predictions with $\hat{B}_h$, and is denoted with $H_{EBV}^{val}$.

Finally, the total prediction error of the EBV$_{\lambda}$ in the whole validation set can be measured with $T_{ebc}^{val} = \frac{1}{24N_d} \sum_{h=0}^{23} \sum_{i=1}^{N_d} B_h^{(i)} - N_d \sum_{h=0}^{23} \hat{B}_h$. The corresponding measure for EBC$_{\lambda}$ is denoted by $T_{EBV}^{val}$.

### 3.4. Prediction of SDCCH and TCH blocking

<table>
<thead>
<tr>
<th>Type</th>
<th>Id</th>
<th>$N_c$</th>
<th>$P_{B&gt;0}$</th>
<th>$\bar{B}_{B&gt;0}$</th>
<th>$E_{ebc}^{trn}$</th>
<th>$E_{EBV}^{trn}$</th>
<th>$H_{ebc}^{val}$</th>
<th>$H_{EBV}^{val}$</th>
<th>$T_{ebc}^{val}$</th>
<th>$T_{EBV}^{val}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>SDCCH</td>
<td>052</td>
<td>16</td>
<td>0.1071</td>
<td>28.14</td>
<td>1.34</td>
<td>2.06</td>
<td>3.25</td>
<td>3.96</td>
<td>3.22</td>
<td>3.96</td>
</tr>
<tr>
<td>SDCCH</td>
<td>053</td>
<td>16</td>
<td>0.1007</td>
<td>23.89</td>
<td>1.50</td>
<td>2.93</td>
<td>0.82</td>
<td>1.87</td>
<td>0.37</td>
<td>1.87</td>
</tr>
<tr>
<td>SDCCH</td>
<td>011</td>
<td>32</td>
<td>0.0357</td>
<td>36.36</td>
<td>1.02</td>
<td>1.62</td>
<td>1.14</td>
<td>1.06</td>
<td>0.43</td>
<td>0.96</td>
</tr>
<tr>
<td>SDCCH</td>
<td>032</td>
<td>32</td>
<td>0.0903</td>
<td>26.30</td>
<td>0.42</td>
<td>1.01</td>
<td>3.24</td>
<td>3.73</td>
<td>3.23</td>
<td>3.73</td>
</tr>
<tr>
<td>TCH</td>
<td>002</td>
<td>11</td>
<td>0.4647</td>
<td>105.11</td>
<td>3.17</td>
<td>9.88</td>
<td>12.00</td>
<td>26.19</td>
<td>5.48</td>
<td>26.19</td>
</tr>
<tr>
<td>TCH</td>
<td>018</td>
<td>11</td>
<td>0.5719</td>
<td>262.71</td>
<td>4.86</td>
<td>14.28</td>
<td>35.49</td>
<td>56.95</td>
<td>21.60</td>
<td>56.93</td>
</tr>
<tr>
<td>TCH</td>
<td>007</td>
<td>25</td>
<td>0.4562</td>
<td>443.78</td>
<td>6.14</td>
<td>23.54</td>
<td>40.85</td>
<td>61.41</td>
<td>21.01</td>
<td>60.88</td>
</tr>
<tr>
<td>TCH</td>
<td>063</td>
<td>70</td>
<td>0.1433</td>
<td>312.77</td>
<td>2.12</td>
<td>13.24</td>
<td>15.58</td>
<td>44.59</td>
<td>13.06</td>
<td>44.59</td>
</tr>
</tbody>
</table>

In the experiments with real GSM network data, we have experienced that the number of possible arrival rate values $N_\lambda = 7$ provides good results. However, one of the $\lambda$ values was always selected to be $\lambda' = K_h / N_s$, i.e. the hourly average of number of requests divided by the number of segments, in order have at least one arrival rate value that is close to the most expected value. Also, the weights for the output variables were $w_i = 1 / \sigma_{Y_i}$, i.e. the greater the standard deviation $\sigma_{Y_i}$ of the variable $Y_i$ in the measurement data, the smaller weight was used in order to avoid them to dominate in parameter estimation.

In Table 2, the results of the experiments with SDCCH and TCH (Traffic CHannel) data are shown. We have selected four BTSs with serious SDCCH blocking problems and four BTSs with serious TCH blocking problems. In the table, the number of channels $N_c$, the frequency of observations with blocking $P_{B>0}$ and average size of the blocking peak $\bar{B}_{B>0}$ are shown. Then, the instantaneous prediction errors $E_{ebc}^{trn}$ for the number of blocked requests $B$ in the training data are shown for the EBV$_{\lambda}$ (left) and EBC$_{\lambda}$ (right) methods. Also, the hourwise prediction errors $H_{ebc}^{val}$ and the total prediction errors $T_{ebc}^{val}$ over the validation data are shown.

According to the table, the proposed EBV$_{\lambda}$ method provides smaller instantaneous prediction errors than the EBC$_{\lambda}$ method for each of the analyzed BTSs. In other words, the flexibility of the EBV$_{\lambda}$ method allows more accurate explanation of the training data.
Also, the EBV\(\lambda\) provides better hourwise prediction errors for all BTSs except one (BTS 011). In this BTS, blocking is more rare than in the other analyzed BTSs. It is likely that there are some differences in blocking patterns between the training and validation sets of this BTS, and therefore, more data should be used in estimation and validation when comparing these two methods with this data.

However, the most important comparison is the one in which the prediction errors for the sum of blocked requests during a time period of approximately 20 days are analyzed. As can be seen from the table, the proposed EBV\(\lambda\) provides clearly smaller total prediction errors than the EBC\(\lambda\) method. These results indicate that a more flexible method that is able to characterize cell-specific traffic patterns is able to provide better predictions of blocking in operative GSM networks.

4. Conclusions

In this paper, a method that exploits the GSM network measurement data and Erlang-B formula in blocking prediction is presented. The results of the experiments show that the proposed EBV\(\lambda\) method is able to give better predictions than the use of the basic Erlang-B formula. This is due to the fact that the sampling frequency used in BTS level data collection is too low in order to be directly applicable to the Erlang-B formula. During a time period of one hour, the arrival rate of the incoming transactions most likely faces some variation, causing the basic Erlang-B formula to give underestimated blocking predictions. We used the measurement data to estimate the distribution for the hourly arrival rates during short time segments and applied the Erlang-B formula for the different segmentwise arrival rates. The proposed method can be used to predict the amount of blocking if traffic balancing or cell size adjustments must be made in operative GSM networks [4]. The proposed method can also be easily modified to prediction of blocking in other types of communication channels if the Erlang-B formula with suitable modifications or extensions are used in the model.

References