A Channel Recovery Method in TDMA Wireless Systems

Yue Ma
Dept. of Computer Science
Duke University, Box 90129
Durham, NC 27708-0129
USA
yuem@cs.duke.edu

James J. Han
Corporate Software Tech. Center
Motorola Inc.
Schaumburg, IL 60196
USA
James_Han-
CJH048@email.mot.com

Kishor S. Trivedi
Duke University, Box 90291
Durham, NC 27708-0291
USA
kst@ee.duke.edu

Abstract: A single base repeater failure in TDMA wireless systems causes all active calls on this base repeater to be dropped. In order to increase system end-to-end availability, an RF channel recovery method for TDMA wireless systems is proposed in this paper. By applying the method, when a base repeater fails, the channels of active calls carried by the base repeater are replaced by working channels in the channel pool of the base site. All the active calls continue without the intervention of end users. The method deals not only with the failure and recovery of base repeaters, but also with the channel failure recovery in handoff processes, transient channel failures, and new call setup processes. To predict system availability and performance, hierarchical stochastic reward net (SRN) models are developed for analyzing channel allocation, base repeater failure/repair, and channel recovery. The results show that the recovery method can nearly eliminate dropped calls when the traffic load is light, and can dramatically reduce dropped calls when the traffic load is normal. The price we pay is a slightly increased blocking probability, nearly transparent to the end users.

1 Introduction

In wireless communications systems, the base site controller (BSC) handles radio channel allocation and many other functions. To date, BSC only handles channel allocation for new calls and handoff calls; no channel failures of ongoing calls are recovered. Whenever a channel fails, the call carried by the channel will be dropped. There are many factors that cause channel failures, for example, power failure, radio frequency (RF) amplifier failure, shadowing, etc. Some of the failures can be categorized as permanent, while the others are transient. To reduce dropping probability, a channel recovery method is proposed in this paper. Both permanent and transient failure recoveries for time division multiple access (TDMA) wireless systems are considered.

The permanent channel failures are caused by equipment failures of base repeaters, such as power failure or RF amplifier failure, etc. In this paper, we assume that a base repeater (BR) can be found in failure status under two situations:

1. While there are some ongoing calls on the BR, the BR is down.

2. A BR may go down while it is in idle status (no on-going calls on the BR). Thus when a call attempting to access a channel on that BR will acquire a faulty channel.

The transient channel failures are caused by many factors in the RF transmission, such as fading, shadowing and interference.

In our recovery method, a failed channel is automatically switched to an idle channel, if one is available. Otherwise, the call with a failed channel is queued until an idle channel is available. Our method is quite similar to the Automatic Protection Switching (APS) [13] scheme, which is widely used to enhance the network integrity in ATM networks. In APS systems, a failed network component is switched to an identical spare component when the protection switch detects a failure. Since the spectrum resource is limited, no spare channels are reserved exclusively for the calls with failed channels. However, failed calls are treated with the same priority as the handoff calls in the sense that both of them can access any available channel in the reserved channel pool.

For numerical evaluation of the recovery method, a hierarchical stochastic reward net (SRN) [2, 3] model is developed. For comparison purposes, an SRN model is also developed for describing the behavior of a base site without recovery mechanism. The numerical results show that the recovery method can nearly eliminate dropped calls under light traffic; it can also dramatically reduce dropped calls under normal traffic. Even under heavy traffic, the method still improves the system performance by decreasing the dropping probability considerably. At the same time, the blocking probabilities are increased slightly. This is because some of the idle channels are used to recover the failed calls. In the non-recovery method, these idle channels would be otherwise used for accepting new calls.

This paper is organized as follows. In Section 2, we give an introduction to some of the features of the TDMA systems which are related to the recovery scheme. In Section 3, a brief introduction to SRN is given. In Section 4, we describe the channel recovery scheme. In Section 5, numerical results are presented and discussed. Finally, we make our conclusions in Section 6.
2 System Description

TDMA systems divide each radio channel into N time slots. Each time slot can be assigned to a different mobile subscriber (MS). The TDMA system that we are investigating consists of a base controller and many base repeaters. In this paper, an IS-136 [9] TDMA system is assumed. Six subscribers (N = 6) can share a single radio channel. In the remainder of this paper, when we speak of a channel, we mean a time slot on a single radio channel.

A TDMA base repeater failure can cause multiple permanent channel failures. In this paper, it causes 6 permanent channel failures, which can be either in use or idle. The BR failure recovery has to restore all the active calls carried by the failed BR. There are some scenarios in which single channel failure occurs, for example, a call can be blocked by a moving/permanent object. This kind of call failure is referred to as the transient failed call (TFNC) in our paper. An active call carried by a failed BR is denoted as BFC. A failed call (FC) can be either a BFC or a TFNC.

The dropping of a handoff call (HC) is considered more severe than the blocking of a new call (NC). One method [10, 11] to reduce the dropping probability of HC is to reserve a number of channels exclusively for HC. For example, if the total number of channels is C and the number of the channels in the reserved channel pool is g, then the number of channels available for NC is C - g.

3 Introduction to SRN

Stochastic reward net (SRN) [3] is an extension of Petri net (PN) [7], which is a high level description language for formally specifying complex systems. A PN is a bipartite directed graph with two types of nodes: places and transitions. Each place may contain an arbitrary (natural) number of tokens. For a graphical presentation, places are depicted as circles, transitions are represented by bars and tokens are represented by dots or integers in the places. Each transition may have zero or more input arcs, coming from its input places; and zero or more output arcs, going to its output places. A transition is enabled if all of its input places have at least as many tokens as required by the multiplicities of the corresponding input arcs. When enabled, a transition can fire and will remove from each input place and add to each output place the number of tokens corresponding to the multiplicities of the input/output arcs. A marking depicts the state of a PN which is characterized by the assignment of tokens in all the places.

Generalized stochastic Petri nets (GSPNs) [1] extend the PNs by assigning a firing time to each transition. Transitions with exponentially distributed firing times are called timed transitions while the transitions with zero firing times are called immediate transitions. A marking in a GSPN is called vanishing if at least one immediate transition is enabled; otherwise it is called a tangible marking. Under the condition that only a finite number of transitions can fire in finite time with non-zero probability, it can be shown that a given GSPN can be reduced to a homogeneous continuous time Markov chain (CTMC) [1].

In order to make more compact models of complex systems, several extensions are made to GSPN, leading to the SRN. One of the most important features of SRN is its ability to allow extensive marking dependency. In an SRN, each tangible marking can be assigned one or more reward rate(s). Parameters such as the firing rate of the timed transitions, the multiplicities of input/output arcs and the reward rate in a marking can be specified as functions of the number of tokens in any place in the SRN. For an SRN, all the output measures are expressed in terms of the expected values of the reward rate functions. To get the performance and reliability/availability measures of a system, appropriate reward rates are assigned to its SRN. In this paper, we use the tool SPNP [4, 5] to specify and solve the SRN models.

4 Model Description

4.1 A Hierarchical SRN Model for the Channel Recovery Method

Because of the interactions among the base repeaters in a base site, an SRN model for a base site would instigate a huge state space for the underlying Markov reward process. Due to the nature of the system to be modeled, we employ a hierarchical approach to model the channel allocation, failure and recovery behavior in a base site. A two-level hierarchical SRN model is constructed:

1. The higher level (Figure 1) models the overall behavior (channel allocation, channel failure and recovery) of a base site (BS).

2. The lower level (Figure 2) models the detail of working and failure status of a base repeater (BR).

Figure 1: Higher level for the recovery scheme: model for a base site.

In Figure 1, Place CP is the channel pool for the remote site. The number of channels assigned to the site is C = N · M, where N is the number of channels a BR can support and M is the number of base repeaters that a base site has. Transitions t_0, t_a, t_b and t_f represent the arrivals of NCs, HCAs, BFCs and TFNCs, respectively. Transition t_M is disabled if there are less than g + 1 channels in Place CP, where g is the number of channels reserved for ongoing calls, which include HC and FC.
On the left side of Figure 1, at Place $P_o$, a NC has already obtained an idle channel, which is about to be tested for its working status. There are two possible outcomes of the test:

1. With probability $c_1$ and the immediate Transition $t_{4}$, the channel obtained is in working condition, and the NC is set-up in Place T.

2. With probability $c_2 = 1 - c_1$ and the immediate Transition $t_5$, the channel obtained is in non-working condition. From our assumptions in Section 1, this also implies that the BR carrying the channel is in bad condition. Through Transition $t_6$, $N - 1$ channels from the same BR are removed from the channel pool, and the bad BR is deposited into the BR repair pool $R_{BR}$. The NC is deposited into the waiting queue $Q_1$. If there are more than $g$ idle channels in the channel pool, the NC needs not wait. The immediate transition $t_7$ is fired instantly. Again the idle channel is tested in Place $P_n$. If there are less than $g + 1$ channels in the channel pool, the NC needs to wait in the queue $Q_1$.

On the right side of Figure 1, at Place $P_1$, an HC/FC has already grabbed an idle channel. Transitions $t_{1}$ and $t_2$ have similar functions as their counterparts $t_4$ and $t_5$, respectively. If the idle channel obtained is in non-working condition, the HC/FC stays in Place $Q_1$ before it fetches another idle channel through Transition $t_6$.

A channel is released under one of the following four situations:

1. Normal termination of a call, represented by Transition $t_{4}$.

2. Handoff departure to another cell, represented by Transition $t_{5}$.

3. Transient failure, represented by Transition $t_{4}$.

Transitions $t_{2}$, $t_{3}$, and $t_4$ have marking dependent firing rates. Their firing rates are proportional to the number of tokens in Place T. This is represented by the sharp sign $\#$ beside the transitions.

4. BR failure, represented by Transition $t_{5}$. When a BR fails while there are some active calls on this BR, the BFCs will be recovered by being put into the queue $Q_1$. The number (denoted by $N_{\text{T}}$) of active calls on a failed BR is obtained from the model (Figure 2) for a generic BR. At the same time, $N_{\text{IC}}$, the number of idle channels will be removed from the channel pool CP. $N_{\text{IC}} = N - N_{\text{T}}$ is the expected number of idle channels when a generic BR fails. The firing rate $\lambda_{\text{T}}$ of Transition $t_{5}$ should equal to $M_{a} \cdot \Lambda_{\text{T}}$, where $M_{a}$ is the expected number of working BRs in a base site, i.e., $M_{a} = M - E[\#(R_{BR})]$. $\Lambda_{\text{T}}$ is the throughput of Transition $t_{5}$ for a generic BR. In the rest of the paper, we use $\Lambda$ to denote the throughput of a transition.

When a BR is recovered through Transition $t_{5}$, $N$ channels are added back into the channel pool. This is shown by the arc from Transition $t_{5}$ to Place CP.

![Figure 2: Lower level (BR) model for the recovery scheme.](image)

Figure 2 models a generic BR in a base site. Transition $t_{5}$ represents the actual job arrivals for each BR. We assume that the job handling is fairly distributed among the working BRs in a base site. Under this assumption, the firing rate $\lambda_{\text{T}}$ equals to $(1/M_{\text{T}}) \cdot (N_{\text{T}} \cdot \lambda_{\text{B}} + \Lambda_{\text{B}} + \Lambda_{\text{C}} + \Lambda_{\text{F}})$. Transition $t_{\text{FR}}$ represents the release of a single channel. Its rate equals to $\lambda_1 + \lambda_{\text{B}} + \Lambda_{\text{F}}$.

The failure of a generic BR is represented by the Transition $t_{5}$. When a BR fails, all the tokens from Places $BR_{a}$ and $T_{1}$ will be flushed out. The dashed arc from Transition $t_{5}$ to Place $Q_{1}$ represents the relationship between the base site and a generic BR. It shows that the expected number of talking channels will be recovered through Place $Q_{1}$. Transition $t_{\text{BR}}$ represents the expected repair time for a generic BR. We assume that at most three BRs (an ultra-rare event) can be down at the same time in a base site, the average repair time for each BR is $k$ hours. Then the expected weighted repair time for a generic BR is calculated through the following formula:

$$
\frac{1}{\mu_{\text{BR}}} = k \cdot P_{k} + 1.5k \cdot P_{k} + 2k \cdot P_{k},
$$

where $\mu_{\text{BR}}$ is the expected repair rate for a generic BR, $P_{k}$ is the probability that there are $j$ BRs in Place $R_{BR}$, $j \in [1, 3]$.

For comparison purposes, a similar hierarchical SRN model for channel allocation without recovery is also developed. Because of the space limitation, the SRN figures are omitted here.

### 4.2 Fixed-Point Iteration and Numerical Measures

With different NC arrival rates, the parameters $N_{B}, \Lambda_{B}, \Lambda_{C}, \Lambda_{F}$ vary accordingly. To capture their dynamic behavior, a fixed-point iteration scheme [6] is applied to determine the above parameters. The values of these parameters are calculated as following,

$$
N_{B} = \sum_{j \in \Pi_{a}} \left( \#[(RBR)], \right) \pi_{j} \left( N_{B}, \Lambda_{B}, \Lambda_{C}, \Lambda_{F}, N_{\text{T}}, \Lambda_{\text{F}}, \right),
$$

$$
\Lambda_{B} = \sum_{j \in \Pi_{b}} \left( \#(T_{j}) \right) \Lambda_{\text{B}} \pi_{j} \left( N_{B}, \Lambda_{B}, \Lambda_{C}, \Lambda_{F}, N_{\text{T}}, \Lambda_{\text{F}}, \right),
$$

$$
\Lambda_{C} = \sum_{j \in \Pi_{c}} \left( \#(T_{j}) \right) \Lambda_{\text{C}} \pi_{j} \left( N_{B}, \Lambda_{B}, \Lambda_{C}, \Lambda_{F}, N_{\text{T}}, \Lambda_{\text{F}}, \right),
$$

$\Pi_{a}, \Pi_{b}, \Pi_{c}$.
\[ N_T = \sum_{j \in \Omega_h} \lambda_j^{(N)}(N_R, \lambda_R, \lambda_T, N_T, \lambda_T^{(j)}) \]

\[ \lambda_T^{(j)} = \sum_{j \in \Omega_h} \lambda_j^{(N)} \pi_j^{(N)}(N_R, \lambda_R, \lambda_T, N_T, \lambda_T^{(j)}) \]

where \( \Omega_h \) and \( \Omega_l \) are the sets of tangible markings of the higher (Figure 1) and the lower (Figure 2) level models, respectively. \( \#(BR_B) \), \( \#(T) \) and \( \#(BR_R) \) respectively denote the number of tokens in Place \( BR_B \), \( T \) or \( BR_R \) in marking (state) \( j \). \( \pi_h \) is the steady-state probability vector of the higher model and \( \pi_l \) is the steady-state probability vector of the lower model. Because of the interdependency between the higher and the lower models, \( \pi_h \) and \( \pi_l \) are functions of \( N_T, \lambda_R, \lambda_T, N_T, \lambda_T^{(j)} \) and \( \lambda_T^{(j)} \). In [6], Equations (1)-(5) are defined as the fixed-point equations. According to Theorem 2 in [6], a fixed point will exist if:

- The functions of \( \pi_h \) and \( \pi_l \) are weighted sums of state probabilities and the weights are constants;
- The CTMCs underlying the SRNs are irreducible with more than one state.

It is easy to verify that the two SRN models that we have developed satisfy the above two conditions. Therefore, fixed-points exist for (1)-(6). For the SRN models of the non-recovery scheme, we also applied the iteration method to decide the related parameters.

In this paper, we mainly concentrate on calculating the dropping and blocking probabilities. We denote the dropping and the blocking probability for the recovery method as \( P_{dr} \) and \( P_{br} \), respectively. To calculate \( P_{dr} \), the reward rate assignment is:

\[ r_d^{(j)} = \begin{cases} 1 & \text{if } \#(CP) = 0, \\ 0 & \text{otherwise}. \end{cases} \]

Here \( r_d^{(j)} \) is the reward rate for state \( j \) in the CTMC of SRN, and \( \#(CP) \) represents the number of channels in Place \( CP \) in marking (state) \( j \). Thus a reward rate of 1 is assigned to the states where the channel pool is empty [12], and a reward rate of 0 is assigned to the other states. Then \( P_{dr} \) is calculated as

\[ P_{dr} = \sum_{j \in \Omega_h} r_d^{(j)} \pi_j^{(N)}. \]

To calculate \( P_{br} \), a reward rate of 1 should be assigned to the states where the channel pool has less than \( g + 1 \) channels, that is

\[ r_b^{(j)} = \begin{cases} 1 & \text{if } \#(CP) \leq g, \\ 0 & \text{otherwise}, \end{cases} \]

and

\[ P_{br} = \sum_{j \in \Omega_h} r_b^{(j)} \pi_j^{(N)}. \]

The dropping and blocking probabilities for the non-recovery method are derived in a similar way.

5 Numerical Results and Discussions

Under the assumption that all the neighboring cells are statistically identical and behave independently, the characteristics of the overall system can be captured by focusing on a single cell. In our system, all the cells are treated equivalently so that we can concentrate our attention on the performability aspects of the channel recovery method. By varying the traffic parameters and developing additional submodels, the presented modeling technique can also be applied to the non-homogeneous cases where cells are located at the border of the covered area.

For the purpose of discussion, we assume that a set of \( M = 8 \) BRs are assigned to a base site. At the base site, \( g = 1 \) out of \( C \) channels is reserved exclusively for the HC/FC. The average handoff rate (1/\( \lambda_h \)) is once per five minutes. The expected call holding time (1/\( \lambda_i \)) is 1.7 minutes. The mean time between transient failure (1/\( \lambda_f \)) is 8 hours. The average failure rate (1/\( \lambda_c \)) for each BR is once every 22,000 hours. The expected repair time (1/\( \mu_b \)) is 2 hours. The failure probability (\( \mu_f \)) for an idle BR is assumed to be \( \lambda_c/\mu_b \).

The grade of service (GOS) [8] is a measure of congestion, which is generally given as the probability of a call being blocked (for Erlang B) or as the probability of a call experiencing a delay greater than a certain queueing time (for Erlang C). Based on the blocking probabilities, we define the following:

- Light traffic: \( \text{GOS} < 0.2\% \),
- Normal traffic: \( 0.2\% \leq \text{GOS} \leq 2\% \),
- Heavy traffic: \( \text{GOS} \geq 2\% \).

The numerical results from the recovery and the non-recovery schemes under different traffic loads are shown.
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>20.4</td>
<td>98.73% ↓</td>
<td>21.03% ↓</td>
</tr>
<tr>
<td>22.1</td>
<td>97.23% ↓</td>
<td>8.11% ↓</td>
</tr>
<tr>
<td>23.8</td>
<td>94.41% ↓</td>
<td>1.23% ↓</td>
</tr>
<tr>
<td>25.5</td>
<td>89.30% ↓</td>
<td>1.21% ↑</td>
</tr>
<tr>
<td>27.2</td>
<td>82.02% ↓</td>
<td>2.70% ↑</td>
</tr>
<tr>
<td>28.9</td>
<td>71.76% ↓</td>
<td>5.39% ↑</td>
</tr>
<tr>
<td>30.6</td>
<td>60.51% ↓</td>
<td>5.20% ↑</td>
</tr>
<tr>
<td>32.3</td>
<td>49.13% ↓</td>
<td>4.79% ↑</td>
</tr>
<tr>
<td>34.0</td>
<td>38.84% ↓</td>
<td>4.33% ↑</td>
</tr>
<tr>
<td>35.7</td>
<td>30.30% ↓</td>
<td>3.87% ↑</td>
</tr>
<tr>
<td>37.4</td>
<td>23.53% ↓</td>
<td>3.53% ↑</td>
</tr>
<tr>
<td>39.1</td>
<td>19.37% ↓</td>
<td>2.22% ↑</td>
</tr>
<tr>
<td>40.8</td>
<td>15.63% ↓</td>
<td>2.0% ↑</td>
</tr>
</tbody>
</table>

Table 1: Recovery vs. non-recovery: percentage change in dropping/blocking probability. ↓/↑ respectively represents that the recovery scheme decreases/increases the dropping and/or blocking probability.

in Figures 3 and 4. In Table 1, we compare the performance of the recovery and the non-recovery schemes through the percentage changes in dropping and blocking probabilities.

When the traffic load is below 25.5 Erlangs, the network is undergoing light traffic. The recovery method reduces the dropped calls by at least 89.3%. Under most scenarios when the traffic is light, the blocked calls are reduced as well. When the traffic load is between 27.2 and 32.3 Erlangs, the network is undergoing normal traffic. The dropped calls are reduced with an average of 65.86%. In the recovery method, some of the idle channels are used to recover the failed calls. In the non-recovery method, these channels might be used for accepting new calls. Consequently, the blocking probability might increase in the recovery scheme. Under normal traffic, this increase is fairly small, with an average of 4.52%. The traffic is heavy when the traffic load is above 32.3 Erlangs. From Table 1, we notice that the recovery method can still reduce the dropped calls by 15.63% when the traffic load is as high as 40.8 Erlangs. However, since the blocking probability is relatively high when the traffic is heavy, increasing the bandwidth is highly recommended for enhancing the overall system performance.

6 Conclusion

In this paper, a channel failure recovery method and the related algorithms of channel allocation with retries and repair are introduced for TDMA wireless systems. SRN models of the channel recovery method with repair process, and the original channel allocation method without recovery are developed. The results show that the recovery method can nearly eliminate dropped calls under light traffic. It can also dramatically reduce dropped calls under normal traffic. Even under heavy traffic, the method still improves the system performance by considerably decreasing the dropping probability. For heavily loaded wireless systems, increasing the bandwidth is recommended to operating companies for larger revenue, more profit, and improved customer satisfaction.

Since some of the idle channels which would otherwise be used for new calls are used to recover the failed ongoing calls, the blocking probability with the recovery scheme is increased in most situations, especially under normal and heavy traffic. This price is well justified. The numerical results indicate that with the recovery method, the dropped calls can be reduced significantly. At the same time, the blocked calls are increased insignificantly.

References


