Radio resource management in WCDMA-based networks in emergency situations

Krzysztof Kordybach, Sami Nousiainen

VTT Information Technology*
Tekniikantie 4B, FIN-02044 Espoo, Finland
Tel: +358 9 456 4415, email: sami.nousiainen@vtt.fi

The paper addresses a situation, when universal mobile telecommunication system (UMTS) network gets congested as a result of a catastrophe or an accident involving many people. The scenario of the situation and objectives for the network during the congestion are defined. Finally, two simple resource management algorithms useful in this case are compared. Their performance, when used separately and together, is evaluated with a UMTS system-level simulator and results presented. They suggest that usage of the techniques offers significant advantages, which surpass drawbacks. The advantages were more substantial when the techniques were combined.

1. INTRODUCTION

Mobile networks are designed to operate at a certain load level. There is always a margin left for load fluctuations, but when a serious congestion happens, the network, or its part, is very likely to fail. This is frustrating for users and dangerous for operators, for it jeopardizes their market position (by both, limiting income during the congestion and questioning their reliability). A classification of congestion situations was presented in [1]. There are two most important types of the situations:

- Predictable, occurring either periodically (e.g. new year's eve), or announced beforehand (e.g. sport events, or open-air concerts); in such a case the operator has time to prepare for the congestion and thus minimize its negative results.
- Unpredictable, usually caused by accidents affecting many people (e.g. railway catastrophes) or natural disasters (e.g. earthquakes); situations of this kind happen suddenly, leaving the operator unprepared.

Especially dangerous are unpredictable events. Then, it is important, that the network does not fail completely, but is able to provide services for the victims and rescue groups. This

* This work has been partially performed in the framework of the projects IST CAUTION and IST CAUTION++, which are partly funded by the European Community. The Authors would like to acknowledge the contributions of their colleagues from the projects.

objective can be achieved only if special algorithms, which can alter normal resource management procedures, are implemented.

In the chapter 2 of the paper we describe more precisely the situation we consider and objectives for radio resource management techniques. The techniques we analyze are presented in chapter 3. Chapter 4 contains parameters of simulations and results. In the last chapter we summarize the paper and comment on the results.

2. EMERGENCY SITUATION

At normal load conditions, one objective for the network is to keep probability of being unsatisfied ($P_{unsatisfied}$) as low as possible, prioritizing ongoing calls. The probability is calculated as probability of unsuccessful call termination, so when either a block (probability of blocking, P_{block}), or a drop (probability of dropping, P_{block}; drop can take place only if the call was admitted beforehand) occurs. A simple formula to calculate it is presented in (1). Appropriate values for network parameters are selected in the network planning phase to meet that objective.

$$P_{unsatisifed} = P_{block} + P_{drop} \cdot (1 - P_{block}) \quad (1)$$

However, during catastrophes the objective changes. A resource management technique (RMT) designed to deal with congestion situations caused by a disaster event must fulfill certain requirements:

- Minimize blocking; this is important, since all people should be able to call for help or to their loved ones.
- Guarantee certain quality level; if it is necessary, a connection must be kept long enough, so that all important information can be exchanged.
- Using available resources as efficiently as possible to maximize network capacity (this is always important, but especially in case of congestion).

When a catastrophe occurs, it influences individual users' behavior. First, their mobility is usually strongly constrained by the disaster. Calls tend to be shorter, but if a user is not satisfied with the connection (because of a block or an enforced termination), he redials until eventual success. Separate category of connections, emergency calls, which normally play little role in a network, form a substantial part of the traffic during a catastrophe.

Indicators used to assess an RMT designed to tackle serious congestions, can be considered either separately for emergency calls (distinguished by e.g. calling or called party's number) and all other calls, or uniformly for whole traffic.

I the analysis we consider a network with wideband code division multiple access (WCDMA) radio interface, which brings up additional problems related to the interference. It increases when load becomes high and connections can be dropped, when required signal to interference ratio (SIR) level can't be maintained. To protect ongoing calls, load threshold must be used for an admission control [2] — after the load in a cell exceeds certain, predefined level, no new call is admitted. The load indicator can take into account many factors, e.g. number of ongoing calls, interference or downlink transmission power.

3. THE TECHNIQUES

3.1. General

In the paper we compare two simple techniques, which can be implemented in any cellular system:

- Time-limited calls
- Pre-emption

A detailed description of the techniques is presented in [3]. Here we present only their brief characteristics.

3.2. Time-limited calls

The time-limited calls RMT consists in establishing a connection only for a limited period of time — if a user has not finished the call after the timer expires, the connection is dropped by the network (there may be voice information for the user about the limited duration when the connection is being established). This has a direct impact on number of users being active in the network, so in case of high arrival rate in the congested area, limited call duration can improve admittance. Simultaneously, calls lasting shorter time are less likely to be dropped due to bad quality in radio interface. It is important to note, however, that the limit must be long enough to enable minimum conversation. Otherwise it will only increase user's frustration and make him try to establish the connection again, thus causing additional traffic.

3.3. Pre-emption

Pre-emption is based on using varied priorities. For the purposes of the considered case, only two levels are sufficient: higher level for the emergency calls and lower for all other. When a new call of high priority requests an access to the network but can't be admitted normally, one of lower priority calls is dropped to free resources. In UMTS and GSM the technique can be implemented with eMLPP (enhanced Multi-Level Precedence and Pre-emption service) framework.

4. SIMULATIONS

4.1. The simulator

To analyze the problem we use the UMTS system-level simulator described in [4]. It has been designed for evaluation of RMTs in various conditions, including high congestion. The main focus of the simulation is radio interface. The simulator enables using many mobility models, including real scenarios (based on digital maps). All important UTRAN (UMTS terrestrial radio access network) procedures, namely power control, measurements, admission control, load control and soft/softer handovers have been implemented. Each user is simulated separately and therefore it is possible to record very detailed output data.

4.2. Optimization of parameters

We assume simple scenario, in which the UMTS network consisting of 7 cells has a fixed admission threshold (admission control is applied to only new calls). As a service we used only the speech, modeled in the simulator as recommended in [5]: on/off model, activity in either direction is 0.5, "on" period lasts on average 1s and average call duration is 120s.

Popularity of data services in emergency situations in the near future is questionable and therefore we neglect them. To facilitate analysis of the results we do not model redials caused by a block or a drop.

Probability of being dropped requires some discussion. It generally represents a probability that an admitted user would be dropped without considering the reasons. When RMTs are in use forced drop can take place in addition to drop due to bad connection quality. Moreover, this enforced drop can be applied to a user without regarding his call history (pre-emption RMT) or taking it into account (time-limited calls RMT). Therefore we use following nomenclature:

- General probability of being dropped, $P_{drop\_general}$, represents probability of any drop.
- Probability of being dropped due to bad quality (when SIR in the uplink or downlink direction remains below required SIR target for 2 s), $P_{drop\_bad\_quality}$.
- Probability of controlled drop, $P_{drop\_controlled}$, when the drop occurs with the duration control.
- Probability of uncontrolled drop, $P_{drop\_uncontrolled}$, when the call duration was not considered before the drop.

In reality the load threshold for admission control is selected by the operator at the network planning stage. In the simulator the same result can be obtained by performing several simulations. We minimized the probability of being unsatisfied, as calculated in (1), because the planning aims at objectives for normal conditions. From the optimal values for admission threshold we chose the one that provides the lowest dropping probability. As the load indicator for the admission threshold we used the number of users, calculated as the fraction of allocated orthogonal variable spreading factor (OVSF) codes in the downlink. When current load in a cell exceeds given threshold, all call establishment requests are rejected. The probabilities of being unsatisfied and dropped (the general probability of being dropped, because there are no RMTs used) for two different arrival rates (calculated as a number of new requests per second in a cell) vs. load threshold for admission control are presented in Figure 1. Only arrival rates below 0.4 can be regarded as normal conditions, though it still generates a high load. The probability of being unsatisfied stabilizes when the threshold is higher than 0.3. Probability of dropping is increasing within that range, so the threshold of 0.3 was the best for assumed requirements and we took it for the RMT evaluation.

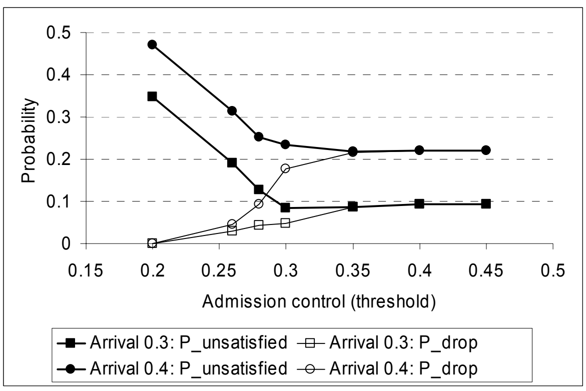

Figure 1. P_unsatisfied and general P_drop for different arrival rates.

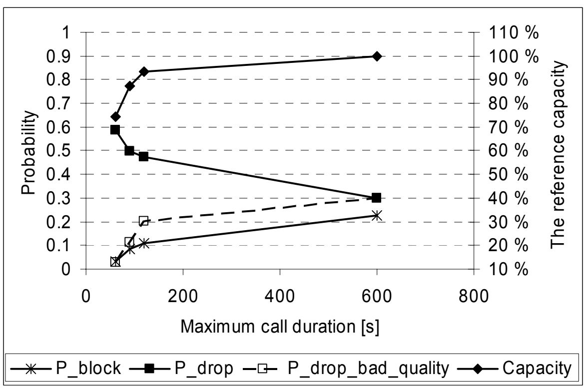

Figure 2. Performance of limited call duration.

In case of the limited duration technique, there is one more parameter to fix: the maximum duration. This is more difficult, since it does not depend only on performance indicators, but it must take into account minimum time required to say everything that is important to enable rescue action. Figure 2 presents several performance indicators vs. max time duration. Since the time-limited calls is a technique for emergency situations, we do not use probability of being unsatisfied, but instead probability of being dropped because of bad quality (which is here the same as the probability of the uncontrolled drop). The limit of 600 s (duration of the simulation) had almost no impact on the calls, so it can be regarded as "no limit" case (capacity of the simulated network when 600 s limit was applied is taken as reference capacity: 100 %; the capacity is calculated as the number of ongoing calls averaged after the simulation has stabilized).

It is clear, that the lower the limit is, the better performance (lower probability of block) can be achieved in emergency situation. At 60 s blocking probability (P_block) and probability of being dropped because of bad quality (P_drop_bad_quality) of the connection are close to zero. General probability of being dropped (P_drop) is high, but since this is enforced dropping, certain call duration can be guaranteed. However, 60 s seems the minimum duration if any communication is to be enabled. Therefore, for the comparison we took 2 cases, 60 s and 90 s.

4.3. Scenario and evaluation criteria

To simulate an emergency situation we used two kinds of calls, normal (arrival rate per cell 0.4 calls/s) and emergency (arrival rate per cell 0.2 calls/s). When pre-emption is used emergency calls are assigned higher priority. We assume the users are stationary.

Evaluation criteria must reflect the objectives defined in chapter 2 for the emergency situations. We have taken two most important parameters: blocking and dropping probabilities. They are calculated for all calls and separately for normal and emergency calls. Additionally, the capacity of the simulated network was observed and compared to the reference case when no RMT was used (100 % capacity) in order to assess the efficiency of resource usage.

4.4. Evaluation

Figure 3 presents results obtained from six simulations (randomness in the simulator causes minor variations in the results in repeated simulations). First was performed without any RMTs. In the next one only pre-emption was used. In the 3rd and 4th simulation time-limited calls RMT was evaluated, with the limit of 60 s and 90 s, respectively. Two last simulations were performed with both RMTs enabled: pre-emption and time-limited calls (limit of 60 s in the 5th and 90 s in the 6th simulation). As it has already been mentioned, in case of the time-limited calls RMT dropping probability can represent either all calls ended prematurely (the general probability of drop), or only those dropped due to bad quality or pre-emption action (the uncontrolled drop). Since in the definition of the time-limited calls RMT it is assumed that the limit is long enough to enable satisfactory call, we took the probability of being dropped without duration control (depicted on the figure as P_drop).

Leaving the network without any RMT is the worst solution: blocking rate is high for all calls, including emergency calls, and uncontrolled dropping due to bad quality is also high. Pre-emption significantly improves quality of emergency calls: there are nearly no blockings and droppings are at the same level as in case of no RMT. This is achieved at the cost of worse general performance: uncontrolled dropping increased (because of pre-emption of

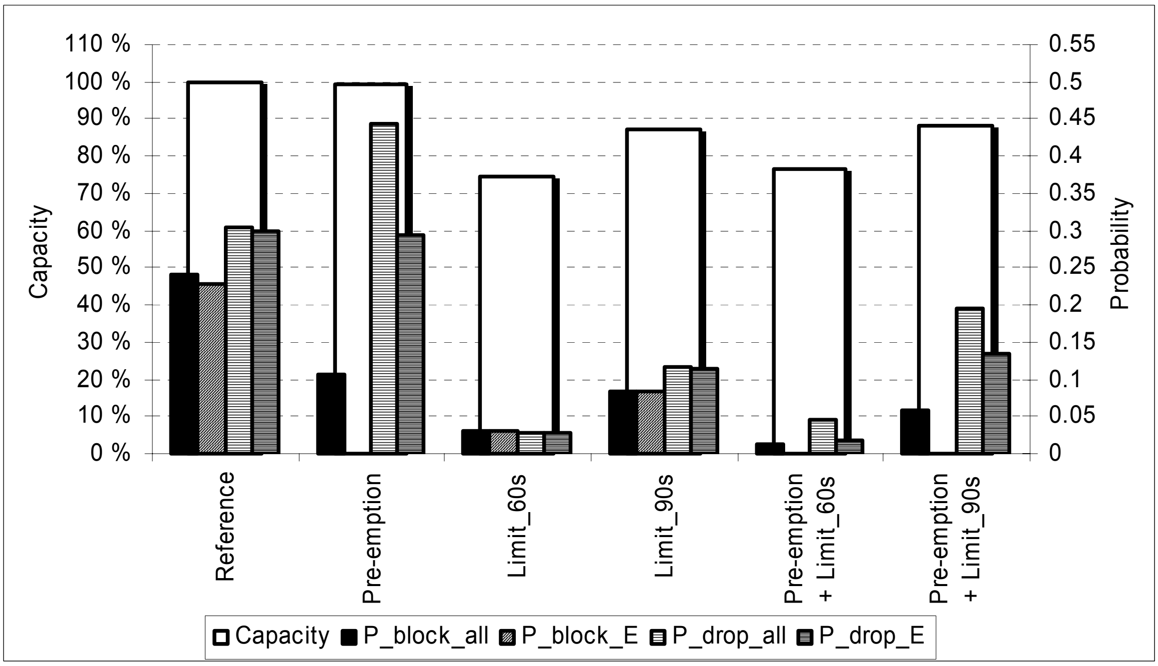

Figure 3. Performance of the network in emergency situation.

normal calls), but at the same time blocking probability is still lower. The emergency calls contributed to the lower value of the parameter, but the decrease is higher than the portion of the emergency traffic (due to the pre-emption it formed about half of all traffic). It means high dropping rate caused shorter duration of normal calls, therefore at the same cell load more calls could be admitted. The capacity of the network is used almost as efficiently as in the reference case. The most important drawback of the technique is the fact, that call duration in case of dropped calls is not controlled, which may result in further increase of arrival rate in real situation (users dropped too early are likely to retry to establish the connection). Detailed results of the pre-emption RMT are presented in Figure 4. In the x-axis first number of each pair defines presented indicator (1: P_{block}, 2: $P_{drop\_general}$, 3: $P_{drop\_uncontrolled}$, 4: Capacity) and the second one described traffic (0: all calls, 1: normal calls, 2: emergency calls). In this case all droppings are uncontrolled, thus the general and uncontrolled probabilities are the same. The favoring of the emergency calls is even better visible on this figure.

Controlled call duration can be partially achieved when limit of duration is implemented (time-limited RMT). Usage of the RMT helps decrease probability of blocking, the more the shorter the maximum duration is. This is obtained at the cost of high the general dropping rate, but the duration of calls is controlled and can be guaranteed with high probability (77 % of all droppings were controlled in the case of 90 s and 95 % in the case of 60 s limit) — probability of uncontrolled dropping is very low, as seen on Figure 3 and Figure 5. This technique, however, has drawbacks, too. The most important is uniform treatment of all calls, normal and emergency ones. Another is less efficient utilization of network resources (expressed as lower capacity). Detailed results of the time-limited calls RMT are presented in Figure 5 (the x-axis is described as in Figure 4), which visualizes great difference of the controlled droppings and those due to bad quality.

A solution may be simultaneous use of both RMTs. The simulations proved that it enables different treatment of normal and emergency calls, while uncontrolled dropping indicators

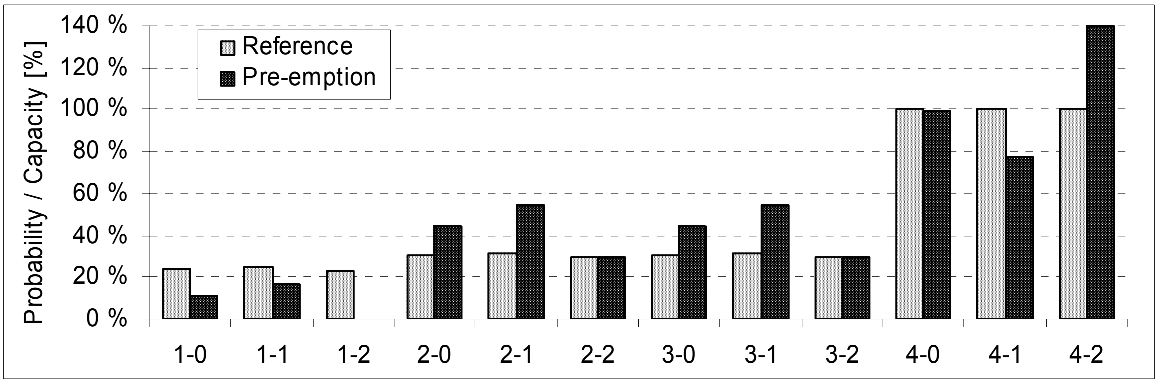

Figure 4. Detailed performance of the pre-emption RMT (x-axis described in the text).

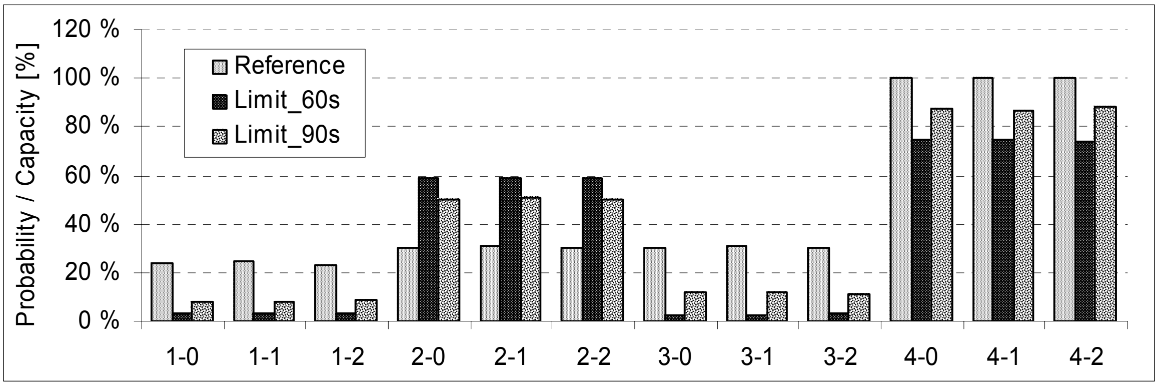

Figure 5. Detailed performance of the time-limited calls RMT (x-axis described in the text).

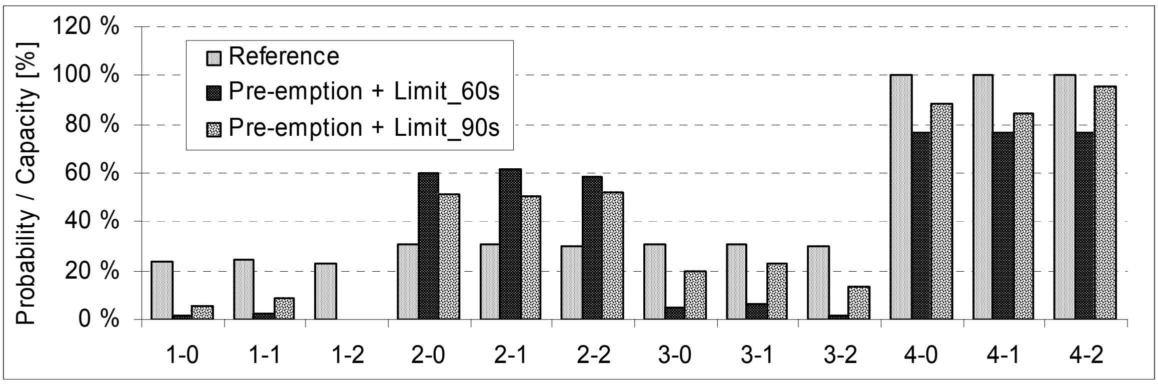

Figure 6. Detailed performance of the mixed solution (x-axis described in the text).

remain still below the reference level. The problem of less efficient network utilization is, however, coupled with time-limited RMT and therefore exists in mixed solution. Detailed results of the mixed solution are presented in Figure 6 (the x-axis is described as in Figure 4). They enable to draw further conclusion, that the shorter maximum call duration, the more final performance depends on the time-limited calls part of the mixed solution. When the duration limit is longer, the role of the pre-emption manifests itself stronger (it is well visible in the capacity comparison).

5. CONCLUSION

Four solutions to congestions in UMTS networks caused by catastrophes have been compared. One is just leaving the network to cope with the congestion (only basic mechanisms are enabled, e.g. admission control), whereas three others consist in using advanced resource management techniques, either separately, or together.

Both techniques, time-limited calls and pre-emption, can improve network performance when the specified conditions occur. They are both simple and possible to implement in all cellular systems. The pre-emption is better when there are relatively few emergency calls and the traffic is still low enough to provide decent quality. However, when number of emergency calls increases and probability of being dropped because of bad quality is high, time-limited calls are better solution. It is also possible to use successfully both the techniques together, however in this case both advantages and drawbacks are inherited. The selection of the most effective approach should therefore depend on the situation and operator's specific objectives.

REFERENCES

1. Kyriazakos S., Karetsos G., Gkroustiotis E., Kechagias C., Fournogerakis P.: "Congestion Study and Resource Management in Cellular Networks of present and Future Generation". Proc. IST Mobile Summit 2001, September 2001, Barcelona, Spain.
2. Holma H., Toskala A.: "WCDMA for UMTS", revised edition, chapter 9.5. John Wiley & Sons, Ltd., England, 2001.
3. CAUTION IST-2000-25352, Deliverable 2.1: "Requirements Analysis and Functional Specification", chapter 5.1, project deliverable, 2001.
4. Nousiainen S., Kordybach K., Kemppi P., Kröger V.-P.: "Dynamic UMTS Simulator for Congestion Studies and Evaluation of Resource Management Techniques", Proc. Personal Wireless Communications conference, October 2002, Singapore.
5. ETSI TR 101.112 v. 3.2.0.: "Selection procedures for the choice of radio transmission technologies of the UMTS". 1998.