Trial Results of Intelligent Paging in GERAN

J. Wang, H. Zhang, M. Toril, V. Wille

***Abstract*—In cellular networks, location management has a strong impact on signaling load. To minimise mobility-based signaling traffic, Location Areas (LAs) have to be defined correctly. Large LAs increase paging traffic, while small LAs increase location updates. Thus, operators have to trade-off these two constraints to maximize network performance. Several authors have proposed selective paging to reduce paging load. This paper describes an intelligent paging algorithm based on most-recent interaction data. Trial results in a live GSM-EDGE Radio Access Network (GERAN) show that the algorithm can reduce paging traffic significantly, while keeping paging success rates almost unaltered.**

***Index Terms*—Mobile communication, Optimization methods, Load management.**

I. Introduction

In cellular networks, location management is key to offering a high quality-of-service. The main procedures involved in location management are *paging* and *location update* (LU). While the former aims to inform a target Mobile Station (MS) of an incoming call, the latter aims to keep the system updated about the location of MSs. To reduce the number of LUs, Base Stations (BSs) are grouped in *Location Areas* (LAs). Thus, an LU attempt is only triggered when an MS moves to a BS belonging to a different LA. When an incoming call arrives, a paging message needs to be broadcast in all cells of the LA where the MS is known to be located. Hence, a trade-off exists regarding the size of LAs. Large LAs result into a small number of LUs, but large number of paging messages, while the opposite is true for small LAs. To circumvent this problem, several approaches have been proposed to reduce the signaling load associated to location management. *Dynamic* (or *local*) LU methods rely on complex algorithms that decide whether to update location information on a per-MS basis [1]. In contrast, *static* (or *global*) LU methods rely on a proper definition of the size and shape of LAs. In these methods, the LA planning problem is formulated as a graph partitioning problem based on user mobility statistics [2]. Alternatively, the paging traffic can be reduced if each LA is divided into several *Paging Areas* (PAs), which are paged sequentially in *selective paging*. The effectiveness of this approach can be improved if the network has additional information about the location in which the MS is located. This approach is known as *intelligent paging* [3]. Several papers have evaluated the performance of these paging methods over idealized network models (e.g., [4][5][6]). However, to the authors' knowledge, no results have been published about the benefit of these methods in actual networks. This paper describes the trial of an intelligent paging algorithm based on last-interaction information in a live GERAN. Section II outlines the basic paging procedure in GERAN, section III describes the intelligent paging algorithm and section IV discusses trial results.

J. Wang is with China Mobile Group Beijing Co., Network Dpt., China (email: wangjinlong@bj.chinamobile.com).

H. Zhang is with Nokia Networks, China, (email: EXT-Hongqing.Zhang@nokia.com).

M. Toril is with the University of Málaga, Communications Engineering Dpt., Málaga, Spain (email: mtoril@ic.uma.es).

V. Wille is with Nokia Networks, Performance Services, United Kingdom (email: volker.wille@nokia.com).

II. Paging in GERAN

The paging algorithm is Mobile Switching Center (MSC) functionality. The MSC also includes the definition of LAs and how these are searched after an incoming call. To aid paging decisions, the Visitor Location Register (VLR) in the MSC stores location data for users currently in the service area of the MSC. When an incoming call or Short Message Service (SMS) is received, the MSC determines which PA should be paged first. In the basic approach (hereafter referred to as *normal paging*), paging is performed over the entire LA simultaneously. For this purpose, the MSC sends a message with the user identity to all BSs in the LA. This message is transmitted in the Paging CHannel (PCH) of the BSs. User identity can be a *Temporary Mobile Subscriber Identity* (TMSI), which in conjunction with the LA identity refers unambiguously to the user, or the full *International Mobile Subscriber Identity* (IMSI) [7]. The length of TMSI is 4 octets, whereas IMSI consists of 9 octets. Thus, more than twice as many MSs may be paged with TMSI, which makes it the preferred option. To protect against message loss, any paging request can be sent several times. The re-paging process is not standardized but vendor-dependent. Nonetheless, most implementations enable operators to start with a number of attempts using TMSI in LA, then with IMSI in LA, and finally with IMSI in the whole network. To avoid unnecessary repetitions, an interval between consecutive attempts is applied.

In theory, the minimum size of an LA is a cell, while the maximum size is an entire MSC. In practice, most operators configure one LA per BSC to cope with paging requests even in high call/SMS traffic situations. In addition, re-paging is often disabled for SMS. While the former results in more LUs, the latter reduces SMS success rate. It is therefore clear that any method that reduces paging traffic allows for the definition of larger LAs, which results in fewer LUs. Thus, the need for LU-related signaling capacity is reduced and the latency of database queries is minimised. Likewise, re-paging could be enabled for all services, which could improve success rate of service establishment significantly.

III. The Hypo-LA Intelligent Paging Algorithm

The proposed algorithm, referred to as *Hypo-LA paging* (HLA), is a selective paging algorithm based on information about the last interaction of the subscriber with the network. The aim of the algorithm is to eliminate the linear dependence of paging traffic on the LA size. For this purpose, the LA is divided into several PAs, which are paged sequentially. The probability of a successful first paging is higher if it is made in the PA where the MS had the last interaction with the network. The latter information can be obtained from the most-recent cell information, stored in the VLR [3]. This information consists of the LA Code (LAC) and Cell Identifier (CI) of the last interaction, which is updated after each transaction (e.g., call set-up, handover, SMS delivery, LU).

The main drawback of selective paging is an increase of the paging delay if the first attempt is unsuccessful as the user has moved to another PA. To avoid loss of traffic due to excessive paging delay, the latter must be kept below certain limits. This is more easily achieved if the number of paging steps is minimised. Thus, in practice, selective paging is often carried out in two steps: the first attempt with TMSI in one PA and the second one with IMSI in the entire LA, with global paging disabled. In addition, it is recommended that a large share of paging requests succeed in their first attempt. This can only be achieved if the initial PA is large enough. Experiments presented later show that a PA comprising one BSC is large enough to provide adequate results. Hence, the final algorithm can be summarised as follows:

1) MSC finds out the BSC of the last interaction based on LAC and CI and starts paging with TMSI only in this BSC. HLA paging is repeated based on LA-specific parameters that define the number of re-paging attempts and the re-paging interval,
2) If HLA paging with TMSI fails, MSC continues a normal paging with IMSI in the entire LA. This LA-wide paging is repeated based on the same parameters mentioned above.

IV. Field Trial

A. Trial Methodology

The above-described method has been trialed in a live GERAN network to quantify the benefits that it can provide to operators. The aim of the trial was to assess the ability of the HLA feature to reduce paging traffic without impacting paging success. The trial scenario consists of 2 BSCs (denoted as BSC A and B), which initially form a single LA. Both BSCs are geographically next to each other, providing contiguous coverage in an urban/dense-urban environment. BSC A has 102 BSs and 277 transceivers, while BSC B has 123 BSs and 413 transceivers.

The paging process used in the network under normal conditions and when using HLA is shown in Table I. In the table, it is observed that, in normal paging, the first attempt is carried out using TMSI for the entire LA. This paging is repeated once before paging with IMSI in the entire LA. Re-paging intervals are set to 3 seconds in either case. The paging sequence for HLA only differs to the normal one in the fact that the first paging attempt is carried out in a single BSC. In the most extreme case, it takes about 9 seconds to notify the target subscriber about an incoming call/message when having to re-page three times, and 12 seconds to notify the source subscriber about a paging request failure.

TABLE I
Parameters in paging methods.

	Normal paging	Hypo-LA paging
Paging type	TMSI	TMSI
Paging area	LA	BSC
Re-paging attempts	1	1
Re-paging interval [s]	3	3
Paging type	IMSI	IMSI
Paging area	LA	LA
Re-paging attempts	1	1
Re-paging interval [s]	3	3

The HLA feature first sends paging messages only to the BSC where the subscriber was last ”seen” by the system. Thus, the overall paging load should reduce as messages only have to be sent to part of the LA. The gain of the method will strongly depend on the reliability of the location information in the VLR, which is linked to the mobility of subscriber between the two BSCs. As LUs are not triggered when subscribers change BSC, subscribers that have moved between BSCs can only be reached by LA-wide paging.

During the trial period, which lasted several days, the HLA feature was enabled at different times of the day. Paging performance data was collected during the entire trial period. At the end of the trial, two comparable 2-hour periods were selected to determine the impact of the feature on paging performance. Comparable, in this context, means that identical time periods on the same day of the week were selected during which paging demand in the selected BSCs were as similar as possible.

B. Trial Results

The effect of HLA on paging traffic can be seen in Table II. It is observed that the paging traffic in the 2-hour period is about 159300 paging messages. Without HLA, this traffic load is observed in both BSCs. The difference in paging traffic between the two BSCs with normal paging is due to measurement inaccuracies related to time synchronization. With HLA, paging traffic reduces significantly. In BSC B, paging traffic is almost halved. For BSC A, the reduction is even larger. The difference in paging traffic between the BSCs with HLA is due to different subscriber load (as BSC B covers a larger geographical area than BSC A). These results show clearly that HLA is able to reduce paging traffic in the LA.

To assess the value of the feature, it is also important to evaluate its impact on paging success. Table III presents the ratio of paging requests that succeed after 1-4 attempts and those unsuccessful for both methods in the whole area. In the table, it is observed that the success rate in the first paging attempt for HLA is well below the respective value of normal paging (i.e., 93% for normal paging, 88% for HLA). The unsuccessful paging attempts performed in the wrong BSC must be repeated in the entire LA. This is the reason for the increased ratio of requests that succeed in the third attempt (i.e., 3.3% increase in absolute terms). After the fourth paging message, the success rate is almost identical to the case of

TABLE II
IMPACT OF METHOD ON PAGING TRAFFIC.

	Normal paging		Hypo-LA paging	
	BSC A	BSC B	BSC A	BSC B
Nbr. of paging messages	159459	159254	61060	84894

TABLE III
IMPACT OF METHOD ON PAGING SUCCESS RATES.

	Normal paging	Hypo-LA paging
	BSC A + BSC B	BSC A + BSC B
1st page	92.81%	87.94%
2nd page	2.22%	3.05%
3rd page	0.43%	3.75%
4th page	0.30%	0.48%
Unsuccessful	4.24%	4.78%

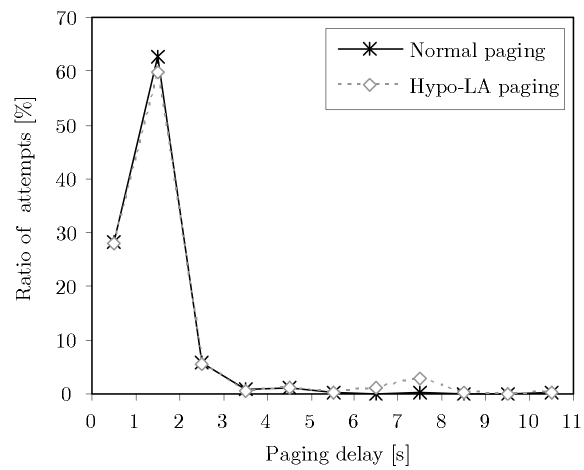

Fig. 1. Impact of method on paging delay.

normal paging. In absolute terms, an increase of 0.54% in the ratio of unsuccessful paging requests is observed. A change of this magnitude is well within the statistical variability inherent in the available data. Hence, it can be concluded that, despite the fact that the success rate of the first page has reduced, the overall paging success rate remains unaltered.

End-users will notice the decrease in the success of first paging as an increase in the paging delay. If the paging process takes longer, then the calling party will have to wait longer before the called party is able to answer. Statistics, not shown here, indicate that, with normal paging, the average paging response time is 1.39 seconds. When HLA is enabled, this indicator increases to 1.62 seconds. This small increase in absolute terms is confirmed by the analysis of the paging response time distribution. Figure 1 shows the histogram of paging response time rounded to the next integer value for both methods. In the figure, it can only be seen that, when HLA is enabled, the ratio of paging requests with delay 1-2 seconds decreases by 3.2% in absolute terms, while the ratio of requests with delay 6-8 seconds increases by the same amount. Likewise, the ratio of paging requests with a delay longer than 8 seconds remains unaltered. These results are consistent with Table III.

Having outlined the ability of HLA to reduce paging load significantly while reducing paging success and user-experience only marginally, it is useful to describe how network operators can benefit from this feature. There are two possible use cases. Operators can keep the existing LA design and deploy the method to cope with abnormally high paging load during events such as Chinese New Year or Christmas. Thus, HLA is used to handle extreme paging traffic loads. Alternatively, operators can use HLA to enlarge LAs without increasing paging traffic. The benefit of utilizing enlarged LAs is that signaling channels, which were used for LUs, can be converted to traffic channels. Thus, capacity is added to the network without the need for any capital expenditure, which is crucial for mature technologies such as GERAN. In this sense, it is worth noting that the HLA feature only relies on the MSC, and can thus be used with any Base Station Subsystem vendor's equipment.

V. CONCLUSIONS

In future cellular networks, it is expected that higher traffic demand and smaller cell size will increase the constraints on the management of location areas. This paper has presented trial results of an intelligent paging algorithm based on last-interaction data in a limited area of a live GERAN. Results have shown that, by defining a BSC as the minimum paging area, a two-fold reduction in paging traffic is achieved when compared to simultaneous paging in the old location area comprising two BSCs. More importantly, it has been proved that, by selecting a whole BSC for the first attempt, paging delay can be kept within reasonable limits. This low delay should be maintained, regardless of the size of the location area. This feature can help GERAN operators to create larger location areas or ensure that the network is operating well even under conditions of unusually high paging traffic loads.

ACKNOWLEDGMENT

The authors would like to thank Xiaomei Wu for the support in organising and conducting the trial.

REFERENCES

[1] A.Bar-Noy, I.Kessler, and M.Sidi, "Mobile users: To update or not to update?" *Wireless Networks*, vol. 1, no. 2, pp. 175–185, Jul 1995.

[2] P. S. Bhattacharjee, D. Saha, and A. Mukherjee, "An approach for location area planning in a personal communication services network (PCSN)," *IEEE Transactions on Wireless Communications*, vol. 3, no. 4, pp. 1176–1187, Jul 2004.

[3] G. Lyberopoulos, J. Markoulidakis, D. Polymeros, D. Tsirkas, and E. Sykas, "Intelligent paging strategies for third generation mobile telecommunication systems," *IEEE Transactions on Vehicular Technology*, vol. 44, no. 3, pp. 543–554, Aug 1995.

[4] S. Mishra and O. Tonguz, "Analysis of intelligent paging in personal communication systems," *Electronic Letters*, vol. 34, no. 1, pp. 12–13, Jan 1998.

[5] O. Tonguz, S. Mishra, and J. Hanwook, "A simple analytical framework for location management in personal communication systems," *IEEE Transactions on Vehicular Technology*, vol. 47, no. 2, pp. 428–439, May 1998.

[6] O. Tonguz, S. Mishra, and R. Josyula, "Intelligent paging in wireless networks: random mobility models and grouping algorithms for locating subscribers," in *Proc. 49th IEEE Vehicular Technology Conference*, vol. 2, Jul 1999, pp. 1177–1181.

[7] M. Mouly and M.-B. Pautet, *The GSM System for Mobile Communications*. Telecom Publishing, 1992.