Utilization of Cellular System Reports for Real-Time Network Traffic Monitoring

Panayotis Georgopoulos and Christos Tsakiris
National Technical University of Athens
Heroon Polytechniou 9, 15773 Athens, Greece
E-Mail: ctsakir@yahoo.gr

ABSTRACT

The rapidly rising number and the sophisticated needs of the cellular mobile network users have led to a change in our conception of these networks: While in the past the sources provided were regarded abundant, now they seem at times inadequate to suit the traffic that new services and new subscribers are about to require. Disasters such as earthquakes and terrorism attacks or normal events like New Year’s Eve and the Olympics are only some of the cases that these networks fail exactly when they are most needed. As a result, new structures that offer significant monitoring and/or control over these networks are obviously necessary. A direct renewal of the whole network infrastructure may be theoretically appealing; real life, on the other hand, demands economically sound techniques that avoid large-scale alterations. In this paper we are presenting a powerful yet cost-effective software method that monitors in real-time the traffic that flows through the channels of any cellular mobile network system. The results obtained by means of this software can be processed in many ways, primarily by prediction techniques that calculate the state of the traffic flow in the immediate future from the past traffic measurement values, thus offering a tool of great importance to the network administrators.

I. INTRODUCTION

Traffic monitoring systems, e.g. the systems that measure the actual utilization of a certain set of resources related to traffic flow, to be used by cellular mobile network systems are of great importance nowadays. Several hardware and software implementations of these systems are on sale and many more are under development, all enveloping varied concepts and techniques. The reason is, partly, the obvious need to supervise the use and effectiveness of the network resources, in practice, of a very complex, rapidly growing, constantly changing and expensive network system [1]. Diagnosis of congestion [2] regarding some specific cells that require allocation of additional frequencies is also crucial, if such allocation is possible. Besides, the cells where the frequencies allocated are underused can be detected. In the case of a redistribution of frequencies or of a dynamic frequency allocation scheme, the sources will be transferred from where they are not utilized to where they are needed [3]. Another important use is the detection of the point where, within a safety margin, the network resources are adequate, resulting to optimization of the use of the invested capital. Moreover, traffic monitoring systems enable the diagnosis of network errors, that in most cases provoke abnormal traffic indications, and possible tracking of the cause of these errors. Additionally, these systems evaluate the effect that alterations, modifications and extensions to the hardware and the software part of the cellular mobile network system have, in practice, to the real capacity and use of the provided resources. Furthermore, traffic monitoring systems are of use regarding the derivation of various statistical reports, for example, the time of the day or the days of the year when congestion appears, or the places where traffic is greater than others. These reports can be useful by many ways (evaluation of development plans, marketing, e.t.c.). Lastly, these systems allow the evaluation of the number of channels of every type that are provided by a cellular mobile network system. This evaluation can lead, for example, to the conclusion that a certain channel type is frequently congested, whereas another one is usually free. This is critical to the development of future cellular systems. Among the various interfaces of the cellular mobile network system, the air-interface, e.g. the interface between the mobile terminal (in most cases the mobile phone) and the antenna system called Base Transceiver Station (BTS) [4], is by far the most important as far as traffic is concerned. Because of its very nature, it is subject to uncontrollable factors that exceed every human capability for prediction. Moreover, it is subject to interference and noise. Its technological capabilities are strongly bound, too, by elements such as the high cost of the equipment needed, public health and the lack of suitable and permitted locations for this equipment. Thus, it is subject to congestion. In the case of the other interfaces (terrestrial interfaces that do not involve transmission through air), even a suspicion for congestion can give rise to great modifications to the interface in order to face events of congestion. This is not possible for the air-interface, because of the reasons stated before. As a result, the focus of the present paper is set on the air-interface of cellular networks. Also, among the various quantities that were possible to measure, the key factor of the air-interface channel utilization was specifically chosen, as every other valid quantity (for
example, blocking ratio or successful call ratio) can be deduced from it straightforwardly [5]. The paper is organized as follows. Section 2 defines the problem in detail and provides a brief technical background regarding the air-interface. The logical channels of wireless networks and the basic procedures of this interface are examined, in particular. The particularities of the interface and the specific need for a traffic monitoring tool that takes them into account will be clarified. In section 3, the basic concept of the proposed monitoring system is presented in detail. Here, the entities called clearing codes are introduced, together with a description of how these entities are processed in order to generate traffic flow calculations. In section 4, the implementation of the method will be examined, accompanied by display and evaluation of the results produced. Several charts are presented and commented specific conclusions of the behavior of the proposed monitoring system are deducted.

II. PROBLEM DEFINITION

Generally, in wireless networks, traffic and signaling have to pass through the same interface, namely air-interface. Due to the nature of wireless access and the need to make efficient use of the system’s resources, a single mobile terminal is never accompanied with a permanent link to the base station which answers for its communications tasks; a series of logical channels and operational procedures have been defined instead. The aforesaid channels and procedures enable the network to provide the promised services.

The first layer of abstraction in dealing with this interface are channels consisting of certain number of bytes that are used for each packet burst. These channels are offered by the multiplexing scheme used by the interface (usually FDMA/TDMA or CDMA [6]) and are called physical channels. The second layer of abstraction are the logical channels that are constructed by physical ones and represent a defined operation of the interface.

An excellent example of the situation is depicted in GSM [6]. The GSM architecture makes use of a FDM and TDM as a method of multiple access. A group of usually predefined frequencies are assigned to each cell and the access to every single frequency is divided in time. In other words, for each frequency there is a number of time slots (TS), namely 8, which are repeated within a cycle called TDMA frame. All these time slots, exactly similar, are in essence the physical channels in the air-interface. The TDMA frames are enumerated by means of a number called Frame Number (FN). According to the numbers of TS and FN that characterize each time slot, the network operator may transmit different types of information and protocols. Therefore, a number of logical channels is constructed.

Two main categories of logical channels are distinguished: the Traffic Channels (TCHs) and Control Channels (CCH) [7], [8]. TCHs are used to convey digitally coded speech or user data in both directions. Full rate TCH is the most common with a 22.8 kbps total rate and 9.6 , 4.8 and 2.4 kbps for user data.

CCHs are used to convey signaling and system control information, that is, the data transmitted through them is not accessible from the user’s application. Their major tasks include establishment, preservation and release of the TCHs, mobility management and access control to the air-interface’s logical channels. Specifically, three kinds have been defined: Broadcast Control Channel (BCCH) Common Control Channel (CCH) and Dedicated Control Channel (DCCH).

The BCCH is a point-to-multipoint unidirectional control channel from the fixed subsystem to the mobile terminal that is intended to broadcast a variety of information to mobile terminals, including information necessary for the mobile terminal to register to the system, such as frequency correction and synchronization sequences. A CCCH is a point-to-multipoint (bi-directional control channel) that is primarily intended to carry information necessary for access management function (e.g., allocation of dedicated control channels). It is involved in the procedures of paging and initial access for both directions. The CCCH can include the following: (1) paging channel (PCH), (2) random access channel (RACH) and (3) access grant channel (AGCH). A DCCH is a point-to-point, directional control channel. Two types of DCCHs are used: (1) stand alone dedicated control channel (SDCCH) and (2) associated control channel (ACCH). Stand alone SDCCH is a DCCH whose allocation is not linked to the allocation of a TCH. The SDCCH is used for system signaling during idle periods and call setup before allocating a TCH. For example, mobile registration, authentication, and location update takes place through this channel. The ACCHs convey supportive information, namely measurement reports and handover signaling. Similar channels are defined for GPRS and other network implementations [9].

The logical channels of any cellular mobile network system are assigned to mobile terminals in a predefined order so as to perform a necessary operation (e.g. mobility management task) or to accomplish the requested service. The types and the alternation order of the logical channels depend exclusively on the operation that takes place. The basic operations normally include connection request, paging, identification of the mobile terminal, authentication, ciphering, call clearing and attach and detach from the system (usually called IMSI attach and detach) [10]. The connection request is performed every time the fixed subnetwork addresses the mobile terminal or vice-versa. The paging procedure is used to offer a single mobile terminal with given identity during the connection request procedure. By means of identification, the mobile terminal provides the system database with further information of its identity in case of an internal system error or when needed while the authentication reveals whether the mobile terminal truly is what is claimed to be with the help of secret numbers, hopefully known only to the fixed subnetwork for protection against the eavesdroppers and the call clearing releases all the resources which were dedicated to the mobile terminal.
During the use of the service and restores the mobile terminal to the idle state. The IMSI attach and detach procedures register and unregister the mobile terminal to the system. If the mobile terminal is attached, it will be paged in the local area where it is present. If the mobile terminal is detached, the system will not waste its resources in paging for an incoming call. Based on the fundamental procedures, top-level operations can be constructed such as location update, handover process, mobile-to-land line call in both directions, mobile terminal-to-mobile terminal call, sending and receiving an SMS message and other services [11]. All these necessary operations add large amount of traffic particularly in the air-interface, thereby in the knowledge of the type and the order in which the logical channels are captured we can devise a monitoring system.

III. PROPOSED MONITORING SYSTEM

When a usage of a service provided by the cellular network, for example a call or an expedition of an SMS message, has terminated, whether successful or not, the network is capable of providing a report on software level, thereby referred as “clearing code” [12]. This report is derived by the billing report (e.g. CDR - Charge Data Record for GSM networks) produced by the network switching center and contains an amount of information that is useful in determining the usage of the network resources of the air-interface. This subset of the clearing code includes the reason why the usage of the service ended (event which shall be called from now on simply as “termination”), usually represented by a hexadecimal number that belongs to one of the two ample categories: normal termination and abnormal one. The first does not necessarily mean that the call, for example, was successful. It does however mean that the network is not liable for an unsuccessful call, if the case is that. The user is assumed to be responsible if he is situated in an area of poor radio coverage or if the air-interface is congested. The time when the usage of the service started and ended, by two numbers called timestamps, is also included, together with the specific cells that served the mobile terminal. Whether the mobile terminal started the use of the service or not, e.g. whether a call was incoming to it or outcoming from it, is stated in this basic set, too. Other pieces of information contained are the subscriber’s identity and various service parameters.

One and only one clearing code is produced for each of the subscribers to the cellular network mobile service. There is no clearing code overlapping whatsoever. Until the termination, no clearing code is produced and no information is received by any system that uses them for the purpose of traffic monitoring. Thus, a caller that continues to discuss on his mobile terminal cannot be taken into account by the proposed monitoring system. These clearing codes are assumed to be collected by a software sniffer that extracts the useful information mentioned above. The amount of information gathered this way is the first source for the proposed monitoring system.

When a clearing code is produced, this means that exactly before the time indicated by the timestamp of the end of the service, a defined sequence of logical air-interface channels of the serving cell(s) was occupied. This chain of occupied channels is determined by the basic operations mentioned earlier. The appropriate procedure, of course, is determined by the number indicating the reason of the termination. The procedure starts at the time indicated by the first timestamp and ends at the time indicated by the second timestamp (when the received clearing code can be produced and certainly after the authentication and ciphering subprocedures). So, what is of importance in this point of view is the effect of the clearing code on the traffic of the air-interface and not the special meaning of each one. As an example, one –the most frequent– clearing code is 0000-normal end of call that comprises the following sequential occupation of channels: PCH (only if the mobile terminal is the called part), RACH, AGCH, SDCCH, FACCH, TCH, each for a computable amount of time [13]. The information database derived this way for each reason of the termination and for each of the calling and called parts is the second source for the proposed monitoring system.

At a specific moment in time, a number of network operations may be active which yield as many clearing codes as the operations, produced at the termination of these operations. These codes are mapped to sequences of logical channel seizures. The core function of the proposed monitoring system receives the clearing codes (source 1) and produces the corresponding logical channel seizures according to the timestamps and the database (source 2). Further analysis calculates average utilization values over various periods of time (Fig. 1).

These operations are performed by a software entity called MU (Monitoring Unit) and result in the derivation of the utilization of all the logical channels in a given area. The MU must comply with a number of requirements:

- It receives the clearing codes as input in real time so as the results will be as valid as possible. It features an internal clock that must stay synchronized with the master clock of the whole network. The configuration of the logical channels in the air-interface are known to the

---

Fig. 1: Generation of clearing codes at the end of the corresponding sequences and calculation of utilizations
MU. Finally, it preserves the number, type and duration of the logical channels’ seizure in a data structure. This structure must be constructed and updated in real-time and it must possess the following properties:

It is required to be dynamically allocated in order to be altered in real-time and to occupy as less space in memory as possible. Due to the size of the structure, its nodes must include proper fields so as to group much common data in one or a few nodes. Specifically, it is advisable that an occupied channel may be presented by one node with information about the duration and the type the channel. The notion of time succession regarding the seizures must be implemented. Simultaneous channel seizures (irrespective of type) taken place in the same period should be presented by the same node.

A noticeable property of the MU’s function is that keeping record of the state of the air-interface in the past allows us to apply the technique of delaying the generation of the results of the present for a short while so as to obtain a more accurate estimation. This happens because the MU is supplied with more clearing codes which refer to a given time. Therefore, the method of looking back is used. The previous property leads to a compromise between the validity of the results and the feature of real-time. It is obvious that it is not feasible to accomplish both. The best we can do is to find the right value for the parameter of backlooking so as to combine almost valid utilization with a slight delay of the demonstration of the results in the present time.

IV. SIMULATIONS AND EVALUATION OF RESULTS

The previously described MU was realized as a simulation program. The basic principles discussed earlier were followed with two exceptions: Firstly, the input (source 1) does not emanate from the normal operational procedure of the network (e.g. no socket communication between the simulator and the actual network intending to supply us with clearing codes is implemented), but the same clearing codes are collected to an input file instead. Secondly, the simulation does not function with an internal clock. The sense of the passed time rises totally from the clearing codes themselves. Particularly, a noticeable assumption made is that the clearing codes become known to the program only at the moment of the end timestamp of that clearing code. The simulation was designed so as to describe the utilization of the air-interface of the GSM networks specifically. Other assumptions are that the channels of the initial access are not retransmitted, for example due to noise or packet collisions. The program includes full control of all configuration parameters.

As an example, the simulation was performed with various values of the backlooking parameter. The input of the clearing codes referred to a real event of heavy traffic flow within a cell including services like telephony, SMS and data and lasting approximately 8 minutes.

Fig. 2 depicts the variation of utilization in the TCH for 4 values of backlooking. The area that is relevant to smaller values must be deemed to be in the foreground, that is they veil the areas with greater values. It is clear that the use of little values of backlooking results in poor performance from the view of the validity of results, but we gain in demonstration time. The last area (230 sec) reveals the real situation over the TCH. This was proved after comparing the results with the actual utilizations. Fig. 3 leads to the same conclusions, even though the utilizations and the durations of the seizures are limited because the channel here is SDCCH.

Fig. 2: TCH utilization in relation to time

Fig. 3: SDCCH utilization in relation to time

Fig. 4: Mean diversity of the normal and actual measures of utilization in relation to delay time
A very interesting chart is located in Fig. 4. It depicts how much the utilization measures provided by the simulation differ from reality as the backlooking parameter increases. A vertical line which suggests the mean duration of the clearing codes is also displayed. It is the approximate point where the curves proceed to a flex point and after that flex point they drop roughly. Consequently, it is not baseless to believe that the mean duration is a fine first value in order to produce reliable results. Moreover, we can notice a difference in the form of the lines. The RACH and the PCH/AGCH curves drop quite roughly, because these channels need great values for backlooking in order to be detected. On the contrary, the TCH curve has a very smooth form as it refers to the last occupied and the most lasting channel so it is easily detectable.

V. CONCLUSIONS AND FUTURE WORK

The proposed monitoring system presents a number of important and evident advantages. Firstly, it involves no essential alteration regarding the cellular mobile communication network. The utilization of present structures and operations is maximal. This fact is equivalent to the principal advantage of low cost. Second, the proposed monitoring system is applicable to any generation of mobile network systems, whether the system runs under GSM, GPRS, UMTS or any other present or future system. The requirement for clearing code production and gathering is reasonable, simple to satisfy and generally fulfilled. Special attention must be drawn to the definition of the quantity of utilization for the latter systems, as they function by packet switching rather than circuit switching. Moreover, it permits acceptably real-time applications, always in the frame that past data processing sets. Enough accuracy is provided for a delay of some minutes, which is nominal for every cellular network traffic monitoring system. Furthermore, the results generated refer to specific cells and specific logical channels. Exact information for the particularities of the congestion faced by the network is provided, making the confrontation of each type of congestion easier. Last but not least, the amount of information needed is minimal and consists only of few bytes of data. There is no possibility of overloading a potentially already congested network. Even the needed CPU and memory sources are kept relatively low.

The simulator or even the actual software program which will be installed in a future monitoring system will in the immediate future be improved in the following points: The validity problem of the results can be alleviated by means of prediction techniques which calculate a correction factor for the recent utilizations using older values. What is more, various statistical models can be used for the calculation of the duration of any channel which seizure cannot be determined accurately such RACH in GSM. Much work is still to be done, but the path is a fruitful one.

ACKNOWLEDGEMENT

This work has been performed in the framework of the project IST CAUTION, which is partly funded by the European Community. The Authors would like to acknowledge the contributions of their colleagues from National Technical University of Athens and especially Sofoklis A. Kyriazakos, VTT Information Technology, Cosmote Mobile Telecommunications S. A., Telia Mobile and Motorola S.p.A.

REFERENCES