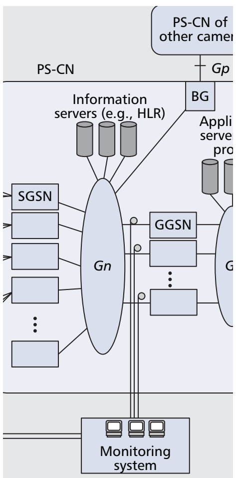
TRAFFIC MONITORING AND ANALYSIS FOR THE OPTIMIZATION OF A 3G NETWORK

FABIO RICCIATO, FORSCHUNGSZENTRUM TELEKOMMUNIKATION WIEN



The availability of high-quality traces can greatly empower the measurement-based optimization cycle, with human experts in the loop, thus driving an already operational 3G network toward improved performance. The authors discuss the contribution that TMA can provide to the optimization of an operational 3G network.

ABSTRACT

Recent years have recorded a surge of research activities on IP traffic monitoring, enabled by the availability of monitoring hardware and large-scale storage at accessible costs. More recently, passive monitoring has been applied to operational 3G networks. The passive observation of network traffic, coupled with advanced traffic-analysis methods, can be a powerful and cost-effective means to infer the network status and localize points of performance degradation without requiring complete access to all network elements. Furthermore, the availability of high-quality traces can be exploited to predict the load of the network under hypothetical conditions, variations of the actual network configuration at the capturing time. Both approaches can be useful for some engineering and reoptimization tasks that are commonly encountered in the lifetime of an operational 3G network. In abstract terms, the availability of high-quality traces can greatly empower the measurement-based optimization cycle, with human experts in the loop, thus driving an already operational 3G network toward improved performances. In this article we discuss the contribution that traffic monitoring and analysis (TMA) can provide to the optimization of an operational 3G network.

INTRODUCTION

In this work we discuss the potential role of traffic monitoring and analysis (TMA) for the engineering and optimization of a 3G cellular network. Research initiatives based on the passive monitoring of Internet traffic have surged in recent years. Enabling factors are the availability of high-performance acquisition hardware (e.g., DAG cards, wiretaps) and large-capacity storage systems (e.g., RAID) at accessible cost. Several methodologies for the analysis of the collected data are being developed by importing concepts and tools from different areas (e.g., signal-processing, data-mining). Some of these techniques can find very concrete applications and provide powerful support for the operation and engineering of real networks, with recognizable benefits in terms of revenue protection and/or cost saving.

Since 2004 we have been involved in a research project aimed at exploiting traffic monitoring and analysis for the engineering of a GPRS/UMTS network [1]. The initial idea was to capture packet-level traces from the live network and use them to distill synthetic user models fitted to the observed data. The models were meant to feed network simulations and/or mathematical analysis in order to derive dimensioning guidelines for the real network. The conceptual trajectory is represented by the simulation path shown in Fig. 1. As a side value, it was expected to extract fine-grain indicators about the current performances and quality of the underlying network (e.g., packet delay, completion time of signaling procedures, etc.). During the investigation we soon recognized that the availability of high-quality traces (to be exactly defined below) yields a much higher potential for improving the engineering practice of the real network, far beyond the mere opportunity to fit abstract models to observed patterns. More generally, TMA can play a major role in several technical areas within the running of a real 3G network: operation and maintenance, troubleshooting, planning and optimization, design and engineering, security monitoring and fraud detection, and so forth. Its reach can be easily extended to less technical areas like for example marketing, service engineering, billing, and tariff design. However, in the spirit of the current Special Issue we focus here only on the aspects related to the network engineering and optimization.

In this contribution we present a few exemplary applications of TMA to the operation and engineering of 3G network. The goal is to provide an overview of the recent and ongoing work, and to promote further research on TMA in 3G. The referenced works[1] come from a few research groups active in this area and primarily from the METAWIN project [1].

The rest of this article is organized as follows. We define the scope of TMA, and then we discuss the evolution from the classical approach for network monitoring toward large-scale TMA. We introduce the requirements for an advanced traffic monitoring system for 3G networks. We provide examples of the potential role of TMA in the optimization of an operational 3G network. Finally, we draw a summary of the conclusions.

 1536-1284/06/$20.00 © 2006 IEEE

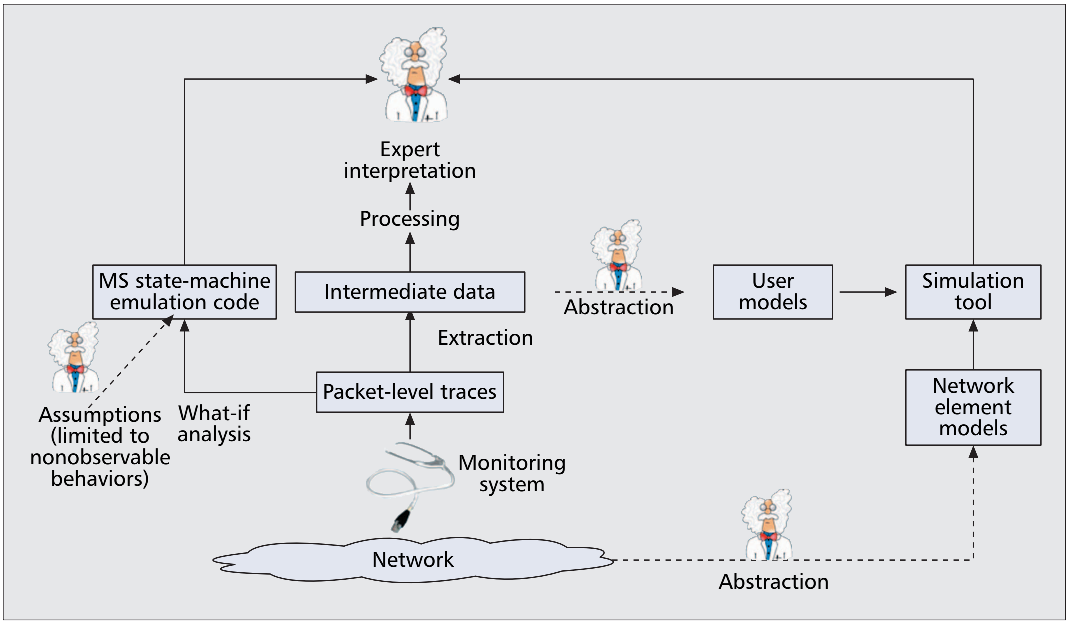


Figure 1. *Conceptual paths to network measurements exploitation.*

With the term traffic monitoring and analysis (TMA), we refer to the ability to passively monitor the traffic crossing the network at a very fine granularity, typically at the packet level, and to process it in order to gain insight into the status and performance of the network and into the behavior of the user population.

Definition of Traffic Monitoring and Analysis

With the term traffic monitoring and analysis (TMA), we refer to the ability to passively monitor the traffic crossing the network at a very fine granularity, typically at the packet level, and to process it in order to gain insight into the status and performance of the network and into the behavior of the user population. Traffic monitoring is typically achieved by inserting passive wiretaps on selected network links so as to "sniff" transit traffic and store locally a copy of each frame, integral or partial (e.g., header only), labeled with externally-derived additional information (e.g., timestamps, L2 identifiers). The resulting data, called "packet-level traces" or simply *traces*, represent a highly accurate representation of "what happened" in the network. The traces can be post-processed in various ways depending on the particular application. Often the first processing step is the extraction of specific embedded data. For example, from packet-level traces it is possible to reconstruct *flow-level traces* — by whatever definition of flow: L4 connections, traffic from/to address prefix, etc. — by listing the set of observed flows along with their attributes (start/end time, volume, endpoints identifiers, etc.). For some applications it is useful to extract *packet-level subtraces* constituted by all packets matching some specific criteria (e.g., TCP SYN towards port 80). Counting specific occurrences of certain events in fixed time-bins will deliver *discrete time-series* (e.g., number of packets transmitted in each direction, or number of distinct hosts seen in each bin). Note that the same methods can be applied to the user-plane as well as control-plane traffic at any protocol layer. For instance in the context of 3G network (Fig. 2) a wealth of information can be extracted from the monitoring of control information at the 3GPP layers below IP (Fig. 3).

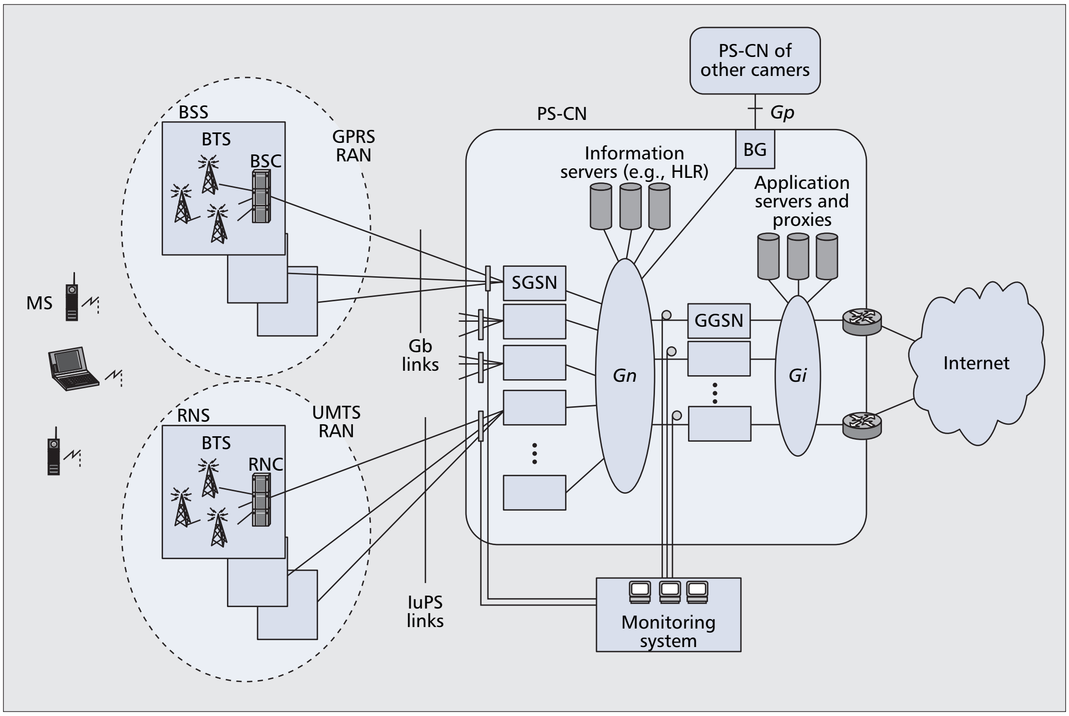
The above few examples illustrate the broad range of *intermediate* data that can be extracted from packet-level traces to serve different applications. This is just the first step in the post-processing chain that will eventually lead to a human action (e.g., reconfiguration of some network element), as sketched in Fig. 1. The following stage takes as input the intermediate data and applies processing methods that are highly specific to the particular application; examples include:

- Basic statistical analysis to extract synthetic indicators that can be immediately interpreted by the network staff (e.g., simple averages and percentiles)
- Direct visualization and multidimensional exploration by network experts
- Application of advanced ad hoc algorithms and techniques (e.g., data mining, signal processing)
- Feeding the intermediate data into another tool (e.g., trace-driven simulations or "what-if" analysis, as discussed below)

Network Monitoring: The Present and the Vision

The classical approach to network monitoring for the purposes of network management relies largely on routine collection of data delivered by the network equipments themselves: built-in counters, logs, SNMP MIBs, and so on. This approach has some limitations. First, the quality of the available data is not always adequate: the time granularity and the aggregation level are coarse, the data semantic is limited, and in some special cases their reliability can be questioned, (e.g., in case of overload or malfunctioning). Second, the process of extracting, gathering, and correlating such data involves considerable costs given the broad heterogeneity of equipment types, vendors, software releas-

The choice of a meaningful and convenient key performance/quality indicator for the PS can not be reduced to a variation of those successfully adopted in the CS. Instead, it requires the application of sound "dual expertise" in the fields of both mobile cellular and packet networking.



■ **Figure 2.** *Structure of a GPRS/UMTS network (packet-switched domain only).*

es, data formats, and even data semantic. Moreover, every change in the network (e.g., replacing of equipment or software upgrades) is likely to require changes in the monitoring infrastructure as well. Above all is a general fundamental problem with such an approach, namely, the lack of decoupling between the monitored system and the monitoring tool, which produces obvious ambiguity problems and makes the latter unreliable in case of equipment malfunctioning.

To complement the routine large-scale data collection from network elements, sporadic fine-grain measurements are performed with small-size network protocol analyzers. These measurement interventions are generally limited in time and space (one or few interfaces) and are often used for troubleshooting actions, after a problem has been detected by external means (e.g., customer complaints).

In summary, the current status of the network monitoring practice is often a combination of large-scale routine data collection from network elements plus sporadic and local fine-grain measurement actions. The future vision of an advanced network monitoring foresees monitoring tools that:

- Are capable of performing routine collection of fine-grain measurements in the large scale
- Are completely decoupled from the production network
- Can cleverly process the recently collected data and deliver reports and alarms proactively

This is exactly the vision of a large-scale TMA infrastructure based on passive wiretapping at key network links, with null or very limited interaction with the production equipment. Such a global system would be intrinsically multipurpose. Observing the traffic allows for the derivation of network-related as well as user-related data. While the former are used by technical departments (operation and maintenance, planning, optimization, etc.) the latter can be useful for marketing purposes.

Large-scale monitoring systems are available on the market and are regularly in use for 2G systems (e.g., GSM). With the deployment of 3G, a novel packet-switched (PS) domain has been added to the legacy circuit-switched (CS) domain, and some legacy CS monitoring systems were extended to cope with the PS domain. However, the current generation of commercial systems falls short of exploiting the full potential of passive monitoring a packet network. To some extent, this is perhaps due to the fact that they were conceived and developed as mere extensions of the legacy system for the CS domain. The PS section was implicitly considered just as a set of additional protocols to be parsed. Instead, the differences between the two domains are deeper: the PS protocols have completely different dynamics, some of which can be exploited to the operator advantage (e.g., TCP close-loop dynamics can be exploited in order to infer delays and packet loss). Furthermore, the PS section supports a completely different usage environment (user populations, terminal types, applications, services, etc.) that is infinitely more heterogeneous and complex than the CS telephony. Accordingly, the choice of a meaningful and convenient key performance/quality indicator for the PS can not be reduced to a variation of those successfully adopted in the CS. Instead, it requires the application of sound "dual expertise" in the fields of both mobile cellular and packet networking. The latter might not be necessarily present in legacy development groups from the CS world.

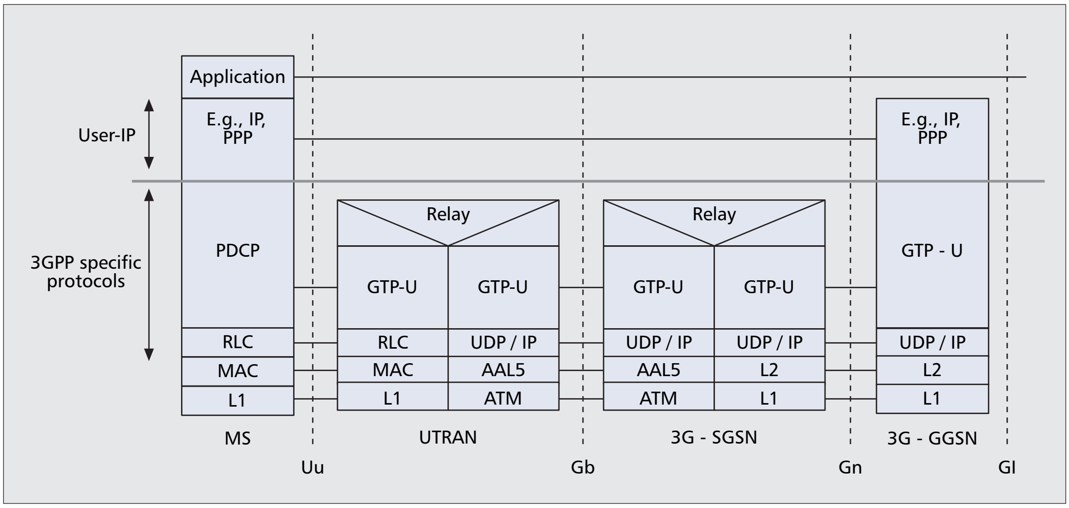


Figure 3. *Partial scheme of the 3G protocol architecture (UMTS data plane).*

The system prototype was developed from scratch on the Linux platform. We used Endace DAG acquisition cards and high-end standard PCs. The system is currently operational in the GPRS/UMTS network of a large mobile operator in Austria, EU.

A Monitoring Framework for 3G Mobile Networks

The prerequisite to fully exploit the potential of TMA in the context of an operational 3G cellular network is the availability of an advanced monitoring infrastructure. In our project [1] we developed a working prototype of a large-scale monitoring system covering selected links on all logical interfaces of the Core Network. In this section we list the most important features of the monitoring system enabling the analysis tasks and applications discussed in the remainder of the article.

Complete Capture — The traffic volumes carried by the internal links of 3G mobile network are such that all frames can be captured with standard hardware equipments, with no need to resort to packet sampling. This is a major simplification compared to TMA in backbone networks with multi-Gigabit traffic rates.

Joint User and Control-Plane Capture — The monitoring system must be able to capture and parse the complete frame, including the 3GPP headers below the user-IP layer as well as the signaling packets at each layer.

Anonymization — For privacy reasons, all subscriber-related fields (e.g., IMSI, MSISDN) are hashed with a noninvertible function. The resulting string univocally distinguishes the mobile station (MS), but cannot be referred to the user identity. For simplicity we will maintain the term "IMSI" to refer to the hashed string. This approach preserves packet-to-MS associations while preserving the anonymity of the subscriber. Also, to preserve content privacy we do not store the user payload above the TCP/IP layer.

Stateful Tracking of Associations — For many applications it is highly desirable to label each packet with certain information associated with the MS it was generated by or directed to. The most important are the MS identifier (hashed IMSI) and the current location (e.g., cell ID). For other applications, it might be useful also to know the MS type (e.g., handset, laptop card, PDA, etc.) and the equipment capabilities. The former can be directly retrieved by the TAC code included in the IMEI, while the latter are usually advertised by the MS during the Attach request. Similarly, it is often useful to associate individual packets to their PDP-context and hence to their attributes (e.g., the assigned IP address). The relevance of such associations is highlighted in the next sections. A generic frame crossing the network does not include all such information in its fields, but a passive monitoring system can dynamically reconstruct such associations by smartly tracking the message exchange between the MS and the network. More specifically, it is required to inspect signaling procedures and certain fields of the lower-layer control information (e.g., TLLI, T-IMSI) and to maintain for each entity (e.g., MS, PDP-context) a dynamic record of associations (we skip the technical details here). The point to be taken is that, in general, any attribute that is exchanged between the MS and a generic network element can be captured and later associated to future packets. The associations between packets, PDP-contexts, and (hashed) IMSI can be extracted on any interface between the MS and the GGSN (e.g., Gn). The localization of the terminal can be achieved in GPRS/EDGE by sniffing the Gb interface (for a detailed description of IMSI-to-cell tracking on Gb, see [2, section IV.C]). For UMTS/HSDPA instead, sniffing on IuPS would allow the localization of the MS only at the granularity of routing areas (RAs), as intra-RA cell changes are not reported to the SGSN. Exact cell-level localization for UMTS/HSDPA would require monitoring the Iub interface between the MS and the RNC.

The system prototype was developed from scratch on the Linux platform. We used Endace DAG acquisition cards and high-end standard PCs. The system is currently operational in the GPRS/UMTS network of a large mobile operator in Austria, EU.

Sample Applications of TMA in 3G

In this section we provide examples of how *high-quality traces* can be exploited in support of network (re)engineering and (re)optimization. By the term "high-quality" we mean packet-level traces that are:

The key idea of observing TCP dynamics in order to infer abnormal network status can be extended to detect areas of performance degradation in the Radio Access Network. The basic idea is again to merge information extracted at the TCP layer with spatial information derived from the 3GPP layers.

- **Complete**: All frames seen on the link are included, on both the user and control planes, and with no loss of information from the lower 3GPP layers (except anonymization of subscriber-specific identifiers for privacy).
- **Labeled**: Each frame is labeled with the most important associations discussed above: generating MS, PDP-context, GPS-synchronized timestamps, and so on.

The extraction of high-quality traces requires an advanced monitoring system to be in place with all the features discussed above.

DETECTING POINTS OF CONGESTION

Most of the current 3G traffic is TCP [3]. The TCP control-loop introduces correlation between the traffic dynamics and the status of the end-to-end path, hence, *observability*: a trouble affecting a TCP flow at some point can be detected by observing the TCP dynamics at a different point along the path (e.g., by the analysis of retransmissions, time-stamps of DATA-ACKs pairs, statistical analysis of observed throughput, etc.).The problem of inferring network conditions out of TCP passive monitoring has been an important area of research and nowadays is mature enough to be applied operationally to real network. Large-scale measurements of TCP-inferred round-trip times, throughput, and retransmission rate from a few monitored links have been reported in some past works [4–6].

These measurements enable the detection of capacity restrictions in the Gn network by monitoring only few Gn links near the GGSNs, an approach considered in two recent papers [7, 8]. By using information available at the 3GPP layers, in this case, the IP address below GTP (Fig. 3), it is possible to discriminate the traffic components associated with each SGSN/GGSN and analyze each of them separately. A capacity restriction on some Gn link will have an impact on the TCP behavior during the peak hour (e.g., higher retransmission rate, compressed marginal rate distribution) that can be detected from traces captured on another link, without the need to access other network elements or to maintain detailed information about the network deployment and configuration (e.g., provisioned bandwidth). The latter feature increases the robustness of the scheme, as it can reveal capacity restrictions due to configuration errors and equipment malfunctioning, and at the same time reduce the maintenance efforts of the tool. An exemplary case found in the real network is shown in Fig. 4, where a bottleneck in the Gn network was detected from the analysis of the marginal rate distribution observed at a different monitoring point (for more details see [7, 8]).

The key idea of observing TCP dynamics in order to infer abnormal network status can be extended to detect areas of performance degradation in the Radio Access Network. The basic idea is again to merge information extracted at the TCP layer with spatial information derived from the 3GPP layers. With reference to the GPRS/EDGE section, monitoring on the Gb links near the SGSNs is sufficient to retrieve the packet-to-cell association for each data packet in both directions. Then TCP performance indicators can be measured for each cell, and such data are used to pinpoint the need for capacity increase in those areas yielding recurrent signs of poor performance. Such a novel approach is still in the exploratory phase. So far it has been considered by two pioneering works: in [2] the authors use the maximum achievable *throughput* as the relevant performance indicator, while in [9] we used frequency of *retransmissions* and *round-trip-times*. The latter approach avoids the problems and limitations associated to the estimation of the available channel capacity from the TCP throughput (e.g., it is not applicable to short-lived connections).

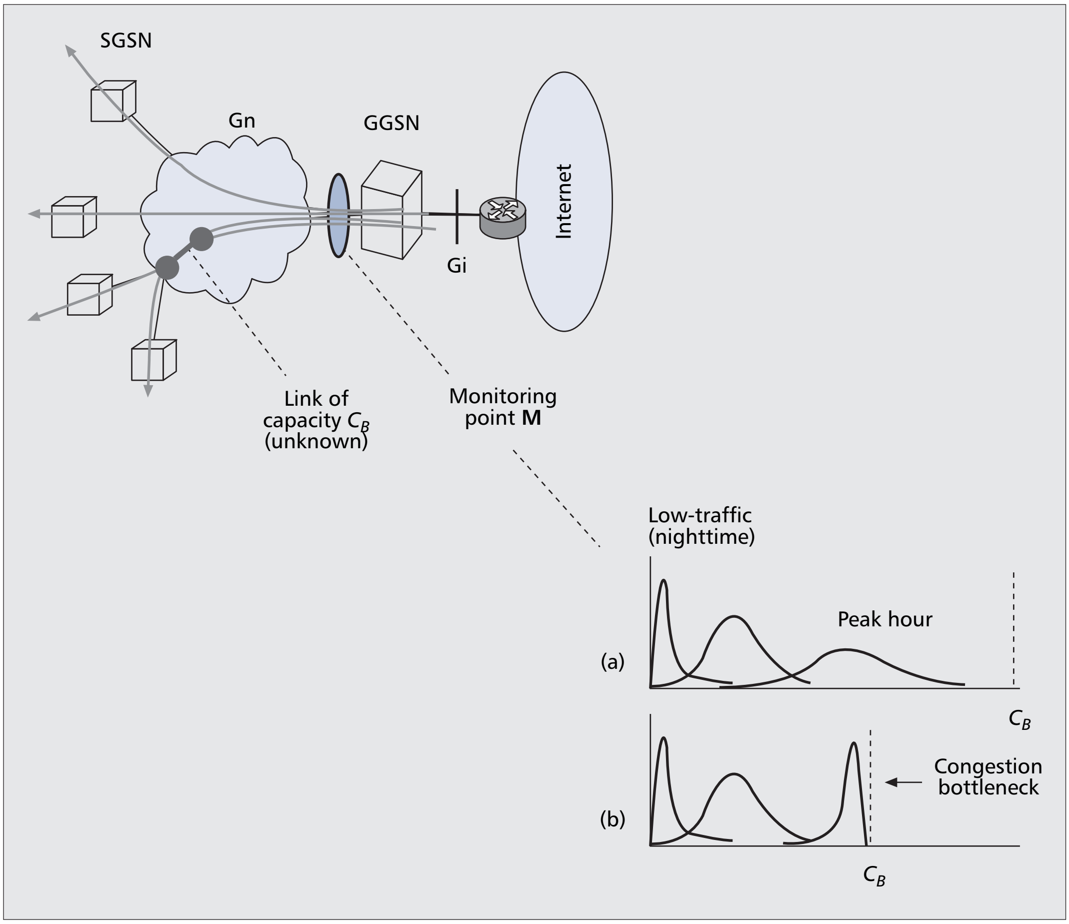
To be effective, these approaches must discriminate those cases where the registered poor performances are due to terminal-specific conditions (e.g., self-congestion due to greedy applications like P2P, terminal malfunctioning, bad radio coverage) rather than accounted on cell configuration. The packet-to-MS association annotated in the traces can be helpful here: in the latter case, poor value of the TCP indicators will be shared by different MSs in the same cell. Note that the problem of discriminating network-related from terminal-related causes of poor performance is also found in the previous Core Network study [7].

With regard to UMTS/HSDPA, monitoring on IuPS would allow a coarser spatial discrimination (i.e., at the level of RAs and RNCs). The cell ID is reported to the SGSN only at the start of the PDP context. Complete cell-level discrimination would require passive monitoring on the Iub interface near the RNC, which is, in general, more costly than monitoring IuPS links. It is not yet clear whether statistical inference techniques can be used to detect the presence of a congested cell in an RA from the TCP indicators measured on IuPS. The lack of spatial discrimination (packet-to-cell labeling) might be compensated by the availability of accurate time granularity, exact MS identification, and partial cell information (at the PDP-context start). For instance, it can be expected that one or more congested cells will generate a *cluster of MSs* with contemporary bad-performance indicators, a pattern that can be perhaps recognized even without explicit cell information, while the cell-information available at PDP-context start might be enough to localize the cell, given the relatively low mobility of most UMTS users (e.g., laptops). This is indeed an intriguing problem for future research, as suggested in [9].

The approaches discussed in this section are still in their pioneering stage. They are promising about the possibility of observing the *actual* quality of the data plane at a sufficient spatial granularity to trigger revision of the locally deployed radio capacity in an area yielding performance recurrently worse than average. The vision is to build a measurement-based reprovisioning loop (with a human expert in the loop) that is *accurate*, as it is based on the indicators of actual transfer quality, and relatively inexpensive, as it requires the tapping of few selected links and local processing.

OPTIMIZATION IN THE CORE NETWORK

An optimization problem found in the engineering of the GPRS Core Network addresses the wiring of Gb links, that is, the optimal associa-



■ **Figure 4.** *Example of bottleneck discovery in the Gn network based on traffic analysis.*

We have implemented such a code and have run it over Gb and IuPS traces for the whole network. Our experience indicates that a carefully optimized code can straightforwardly track the state of all the MSs in the whole network in real time on high-end standard hardware.

tion of BSC to SGSNs. Given a certain setting of the Radio Access Network, and specifically a given grouping of cells into RAs, the problem is then to optimize the assignments of the BSC to SGSNs (i.e., the Gb wiring). There are multiple concurrent optimization goals to be pursued: minimize the monetary cost of the link distance (typically, it increases with the distance); balance the load among the set SGSN (it is typically measured in terms of the peak number of attached users); and minimize the frequency of inter-SGSN RA updates (each of such procedure is signaling intensive and involves four network elements, i.e., two SGSNs, HLR, and GGSN). As a partial subproblem, one might be interested in optimizing only the wiring between a set of SGSN collocated in the same physical site. In this case, only the last two minimization objectives apply. This problem was considered in [10] and formalized as a mixed-integer linear programming (MILP) problem. The input data for the optimization are two discrete time-series:

- The number of attached MSs in each routing area i at time t_k
- The number of MS passing from routing area i to j in the interval (t_{k-1}, t_{k-1})

Both time-series can be extracted directly from high-quality traces taken on Gb/IuPS, as detailed in [10]. The basic idea is to reproduce the behavior of each MS with a simplified state-machine, including only the states relevant to the specific application and use the messages found in the traces to trigger state transitions. In our case, only two states are sufficient: "Attached" (A) and "Detached" (D), with the former including the current RA identifier as an internal variable. For example, a successfully completed Attach Request in routing area i will trigger a transition $D \rightarrow A(i)$, RA updates mark transitions $A(i) \rightarrow A(j)$. The transition $A \rightarrow D$ can be triggered by an explicit Detach procedure, but also by the absence of any packet from the MS for a certain timeout after which the MS is considered "Implicitly Detached" from the network. We have implemented such a code and have run it over Gb and IuPS traces for the whole network. Our experience indicates that a carefully optimized code can straightforwardly track the state of all the MSs in the whole network in real time on high-end standard hardware.

TMA for General Parameter Optimization: What-If Analysis

As described in the preceding paragraph, we have adopted the method of "reading" the traces by means of simplified state-machines in order to reconstruct the internal behavior of each MS (i.e., its internal state). This approach can be extended to reconstruct the *hypothetical* behavior that would have taken place under different network conditions (e.g., for different parameter settings).

Consider the following simple example. The

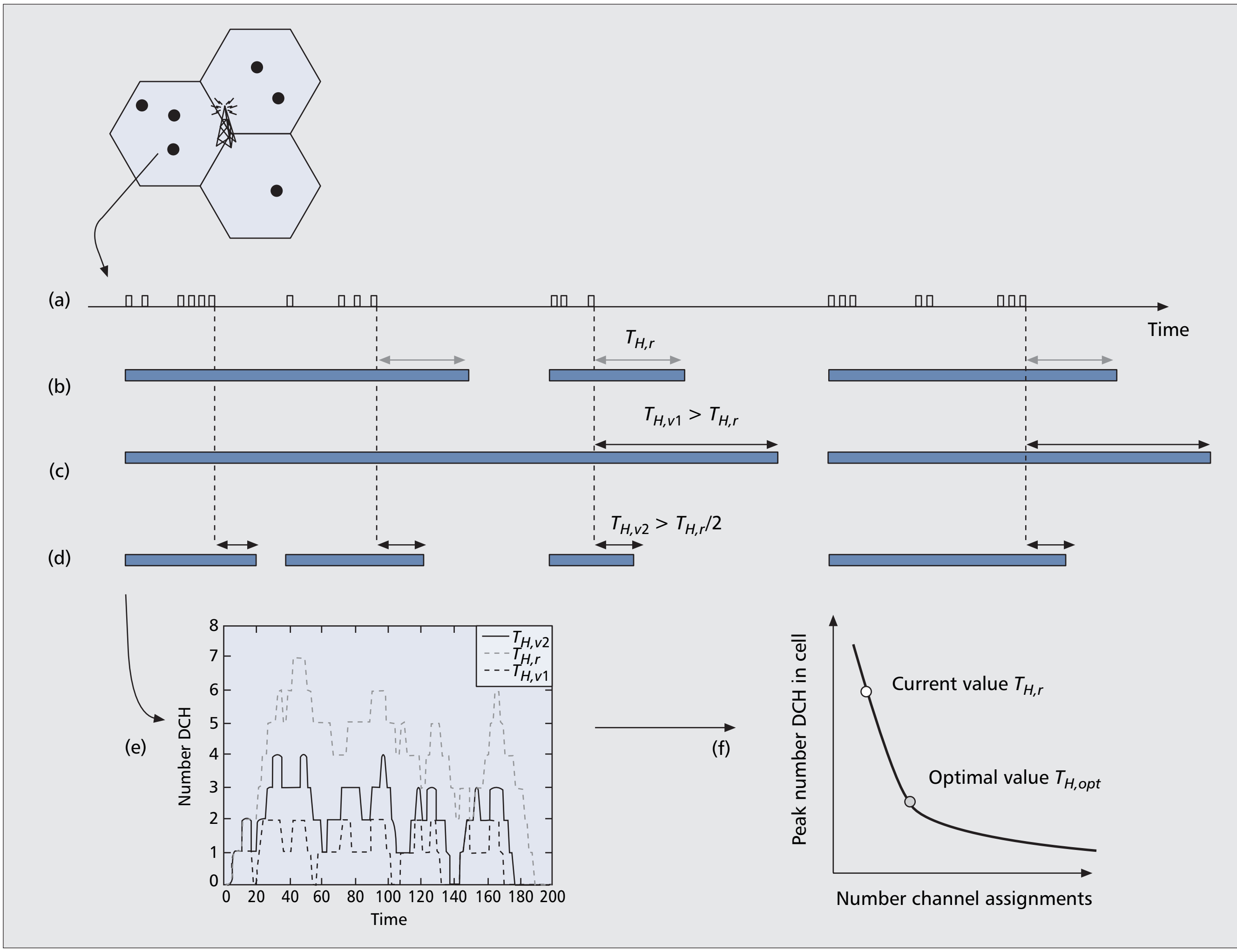


Figure 5. *General scheme of measurement-based parameter optimization for* T_H *timeout: a) observed packet pattern for a single user; b) estimated DCH assignment process with the current parameter value, and c, d) with different timeout settings; e) estimated cell-level DCH occupancy process for different parameter settings; f) summary trade-off curve and new optimal value.*

DCH channel in UMTS is assigned and released dynamically to each MS and its bandwidth is adjusted dynamically, depending on the instantaneous usage. The algorithms for dynamic resource allocation are implemented in the RNC and are often vendor proprietary, with several parameters tunable by the operator. In the simplest case, the DCH is assigned when the first packet is transmitted or received, and released after that no packets are seen for a timeout T_H (the DCH holding timeout). The optimal value of T_H must trade-off between two competing goals: long values cause resource wastage, while short values involve frequent release/reassign cycles. The latter worsen the user experience, since each reassignment procedure introduces an additional delay on the arriving packet, and increases the signaling load on the radio link. When the DCH is in place, its assigned bandwidth can be dynamically adjusted (increased/decreased by multiples of 64 kb/s) based on thresholds of the traffic rate measured on a certain time window.

The optimal value of T_H as well as of the other involved parameters (e.g., rate thresholds, window length) can be obtained by mathematical analysis or via simulations. In both cases the input is a certain model of user behavior. If high-level traces from the real network are available, the model can be fitted to the data. However, the process of distilling a model out of empirical data is complex and cannot avoid introducing severe simplifications. This approach can be successful in those cases where:

- The user population is highly homogenous
- The user population's behavior involves few dimensions

This is typically the case for voice calls in traditional telephony, with the additional advantage of a pretty good mathematical tractability of the resulting models. But in multiservice networks such as 3G, this is no longer the case: the behavior of the user population is marked by high heterogeneity, large disparity (also known as the "elephants and mice" phenomenon), and large dimensionality. Therefore, the hope that the resulting model is usable at all for practical engineering vanishes. Coming back to our example, we recognize that the availability of high-quality traces offers the possibility to estimate what would have been the DCH occupancy pattern of each MS for different values of T_H by adopting the simplified-state machine approach. We can build up a simplified state-machine for each MS, with states corresponding to the different values of DCH bandwidth plus the "nonassigned" state, and state transition rules according to the bandwidth assignment algorithms (e.g., release the channel after T_H from the last packet). The recorded traces were captured for a single value of T_H, say, $T_{H,r}$ (real). By feeding such traces into state-machines with a different parameter setting, say, $T_{H,v}$ (virtual), we can reconstruct the DCH occupancy profile for each MS in the different network conditions ($T_{H,v}$ instead of $T_{H,r}$). This approach, sketched in Fig. 5, is similar to trace-driven simulations. The reconstruction is not exact, as it implicitly assumes that the user behavior is independent from the network setting (in our case, the value of the DCH timer), so that the packet pattern observed for $T_{H,r}$ would be maintained for $T_{H,v}$ as well. In other words, it neglects the impact of the parameter setting onto the traffic pattern. This is a source of error, but certainly smaller than the model-based simulation approach, wherein the lack of *observability* of the would-be user behavior for $T_{H,v}$ is cumulated with abstractions and simplifications of the actually observed behavior for $T_{H,r}$.

In some cases the above trace-driven approach can be extended beyond parameter-tuning. There are a number of engineering questions that can be fruitfully supported by a smart "replay" of high-quality traces. As a purely illustrative example, assume that the operator is considering introducing a new GPRS/EDGE service for all users of APN X, involving the sending of periodic messages by all MSs with an active PDP-context in X towards a central server every period T. Using trace-driven analysis, it is possible to predict rather accurately the additional message load on the server as well as in each cell for different values of T. Similarly, it is possible to predict the load on the network following large pathological events, for example, a large-scale infection by scanning worms affecting a certain fraction of terminals.

Conclusions

Traffic monitoring and analysis (TMA) can play an important role in the operation and performance optimization of an already operational 3G network. It is a cost-effective approach to network monitoring compared to the classical practice based on direct access to the nodal equipment. Wiretapping of selected key links can be sufficient to evaluate the performance of nonmonitored points, by exploiting the well-known dynamics of the TCP/IP protocols. Nowadays the hardware required to perform continuous fine-grain traffic monitoring on a large-scale typically comes at accessible costs, at least for the current volumes found in typical 3G networks.

Using a few examples, we have shown that high-quality traces extracted by a smart monitoring system can support the optimization of network parameters and the decision process about new features. We do not claim that any optimization task can be resolved from traces. The main limitation is that the collection of large-scale complete traces is cost-effective only in the Core Network, where the aggregation level reduces the number of physical links to be monitored. On the other hand, not all network phenomena are observable from the Core Network (e.g., intra-RA cell transitions in UMTS are not "seen" in IuPS). This suggests that TMA should be considered as a powerful and cost-effective complement, rather than a replacement, for the current engineering practice.

TMA can play an important role in the operation and performance optimization of an already operational 3G network. It is a cost-effective approach to network monitoring compared to the classical network monitoring practice based on direct access to the equipment.

References

[1] METAWIN and DARWIN, http://userver.ftw.at/~ricciato/darwin/index.html

[2] T. Borsos *et al.*, "A Measurement Based Solution for Service Quality Assurance in Operational GPRS Networks," *Proc. IEEE INFOCOM*, Barcelona, Spain, Apr. 2006.

[3] P. Svoboda *et al.*, "Composition of GPRS/UMTS Traffic: Snapshots from a Live Network," *Proc. 4th Int'l. Wksp. Internet Perf., Simulation, Monitoring and Measurement*, Salzburg, Austria, Feb. 27–28, 2006.

[4] P. Benko, G. Malicsko, and A. Veres, "A Large-Scale, Passive Analysis of End-to-End TCP Performance over GPRS," *Proc. IEEE INFOCOM*, Hong Kong, Mar. 7–11, 2004.

[5] F. Vacirca, F. Ricciato, and R. Pilz, "Large-Scale RTT Measurements from an Operational UMTS/GPRS Network," *Proc. 1st Int'l. Conf. Wireless Internet*, Budapest, Hungary, July 10–15, 2005.

[6] Y. Lee, "Measured TCP Performance in CDMA 1x EV-DO Network," *Proc. Passive and Active Measurements Conf.*, Australia, Mar. 30–31, 2006.

[7] F. Ricciato, F. Vacirca, and M. Karner, "Bottleneck Detection in UMTS via TCP Passive Monitoring: A Real Case," *Proc. ACM CoNEXT '05*, Toulouse, France, Oct. 24–27, 2005.

[8] F. Ricciato and W. Fleischer, "Bottleneck Detection via Aggregate Rate Analysis: A Real Case in a 3G Network," short paper, *IEEE/IFIP NOMS '06*, 2006.

[9] F. Ricciato *et al.*, "Passive Tomography of a 3G Network: Challenges and Opportunities," poster at *IEEE INFOCOM*, Barcelona, Spain, Apr. 2006.

[10] F. Ricciato, R. Pilz, and E. Hasenleithner, "Measurement-Based Optimization of a 3G Core Network: A Case Study," *6th Int'l. Conf. Next Generation Teletraffic and Wired/Wireless Advanced Networking*, St. Petersburg, Russia, May 29, 2006.

Biography

FABIO RICCIATO (ricciato@ftw.at) received a Laurea degree in electronics engineering in 1999 and a Dottorato di Ricerca (Ph.D.) in information and communications engineering in 2003, both from University La Sapienza, Rome, Italy. He has participated in several national and European research projects within the area of networking. He is now a senior researcher and project manager at Forschungszentrum Telekommunikation Wien, where he leads the DARWIN project on traffic monitoring and analysis in 3G mobile networks.