A Customer Service Assurance Platform for Mobile Broadband Networks

Alessio Botta and Antonio Pescapè, University of Napoli Federico II (Italy)
Claudio Guerrini, Accanto Systems (Italy)
Marin Mangri, Telefonica Germany (Germany)

ABSTRACT

In this article, we discuss trends, issues, requirements and solutions for customer service assurance (CSA) platforms for mobile broadband networks. We propose a distributed probe-based architecture called intelligent CSA (iCSA), and demonstrate how it is a key component of an advanced OSS. iCSA provides support to OSSs, addressing a number of important issues: increased bit rate, joint analysis of control and user plane, multidimensional analysis, root cause analysis, and so forth. To provide real evidence of the benefits of our proposals on a real mobile broadband network, we also illustrate experimental results on two hot topics: *mobility and session management* and *root cause analysis of TCP connections*.

INTRODUCTION

The traffic over mobile networks has been strongly increasing due to the growing number of users, terminals (i.e., smartphones and tablets), and applications (i.e., video streaming and social networks). At the same time, mobile operators are facing several challenges in their value chain (e.g., increased infrastructure management costs, reduced average revenue per customer), while sustaining the migration toward evolved packet systems architecture (Table 1, SR1). Due to a plurality of access and core network technologies being used to deliver a complex set of services, setting up and running mobile broadband networks is becoming increasingly complex. In addition, increased competition and customer churn are driving service providers to be even more customer-centric and innovative with their services. In this evolving scenario, both industry and academia are paying more and more attention to customer experience management (CEM) and customer service assurance (CSA). CEM refers to the collection of processes an operator uses for tracking, overseeing, and organizing every interaction between a customer and the organization throughout the customer life cycle (from service support to new sales, from trouble resolution to billing inquiry, etc.); CSA refers to the part of CEM that deals with service quality, a measure of how individual users experience the services they purchase [1]. As a result, CSA platforms supporting the operational support system (OSS) are indicated as a mandatory ring in the management chain for mobile broadband networks and are consolidating in a clear framework [1, 2].

While mobile broadband networks are completing their transformation in fully packet-based architectures, traffic monitoring systems (based on passive probes) are continuously evolving. In particular, they are experiencing a continuous shift towards being a key tool in the area of CSA platforms able to track and manage mobile subscribers' experience, when properly fed and configured [3]. OSSs market analysis [4] positions probe-based traffic monitoring systems into the ecosystem of *service assurance*, alongside other OSS applications for *fault management*, *performance management*, and *service quality management*. The service assurance market generated $2.3 billion revenue in 2009 and is forecasted to grow up to $3.4 billion in 2014, resulting in a compound annual growth rate (CAGR) of 8.3 percent [4]. Probe-based systems are the largest subsegment in terms of revenue, and it is estimated to increase from $843 million in 2009 to $1.18 billion in 2014, a CAGR of 7 percent [4]. The market of mobile telecommunication services and, consequently, mobile network operators are facing the following challenges, which will tend to increase in the upcoming years:

- A dramatic increase in cost and complexity for managing mobile networks and services
- The need for heavy investments in infrastructures to meet the growing demand of data communications, while radio access capacity is not scaling accordingly [5]
- A reduction in average revenue per customer (average revenue per user)
- A shortage of human resources with the appropriate skills to manage the growing complexity

These changes will necessitate a number of macro-requirements for OSSs, which will focus on:

- Overall customer experience, in order to minimize customer churn.
- Policies to control access to resources based

0163-6804/11/$25.00 © 2011 IEEE

SR1	3GPP TS 23401, "General Packet Radio Service (GPRS) Enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) access."
SR2	3GPP TS 23.207, "End-to-End Quality of Service (QoS) concept and architecture."
SR3	TM Forum, "TR 148 Technical Report: Managing the Quality of Customer Experience."
SR4	TM Forum, "TR149 Technical Report: Holistic e2e Customer Experience Framework & Sample Workbook."
SR5	3GPP TS 29.060, "General Packet Radio Service (GPRS); GPRS Tunneling Protocol (GTP) across the Gn and Gp interface."
SR6	3GPP TS 25.415: "UTRAN Iu interface user plane protocols."
SR7	3GPP TS 24.301, "Non-Access-Stratum (NAS) protocol for Evolved Packet System (EPS)."
SR8	3GPP TS 24.008, "Mobile radio interface Layer 3 specification; Core network protocols."
SR9	3GPP TS 23.003, "Numbering, addressing and identification."
SR10	3GPP TS 32406, "Performance Management (PM); Performance measurements; Core Network (CN) Packet Switched (PS) domain."
SR11	3GPP TS 32426, "Performance measurements Evolved Packet Core (EPC) network."
SR12	3GPP TS 23.002, "Network architecture."

Table 1. *Set of relevant standards.*

on all the key business variables, such as: customer segments, devices, services and network load [6]. In this area, the Third Generatioin Partnership Project (3GPP) has defined and is continuously improving a policy architecture (Table 1, SR1) on top of the well-known quality of service (QoS) architecture (Table 1, SR2).

- Operational efficiency [2].

In this complex scenario, probe-based platforms have to cope with a number of issues, in order to provide adequate support to OSSs and to implement a CSA strategy. We present a platform, called intelligent CSA (iCSA), which aims at addressing the key issues through innovative solutions:

Increase of bit rate: The bit rate of links probed close to key network nodes is continuously growing. Typical gateway general packet radio service (GPRS) support node (GGSN) capacity is around several gigabits per second, and it is expected to increase suddenly during the upcoming years, especially with the transition to *evolved packet systems* (Table 1, SR1). iCSA supports specialized packet capturing and preprocessing hardware and software designed to exploit modern multicore CPUs.

Split of user and control plane metrics: Modern telecom networks adopt the split of control and user planes vs. all relevant business dimensions (customer groups, device or service types, key network parameters). iCSA performs the aggregation of user plane and control plane metrics in different parts of the network, through different components of its architecture.

Complex network architectures: Practical deployment and the need to correlate information collected from different probing points mandate the implementation of a complex and coordinated distributed solution. A typical use case is the analysis of the S1 protocols in the Evolved Universal Mobile Telecommunications System (UMTS) Terrestrial Radio Access Network (E-UTRAN): analysis of the user plane (e.g., TCP dynamics) can be done close to the serving gateway (Serving-GW), while the control plane (e.g., device type) is typically probed at the mobility management entity (MME) (Table 1, SR1). iCSA implements a highly distributed architecture, in which the different components are managed by a centralized entity.

(Near) real-time availability of key business metrics: One of the major challenges of a CSA platform is to provide measurements as quickly as possible to lead to fast error detection and correction. When applied to a probe-based system, this means that monitored protocol packets have to be continuously analyzed in order to provide summaries and measures (e.g., session and mobility management procedures failed in the last five minutes). As a probe-based system becomes a key part of an OSS ecosystem, continuous and near-real-time availability of data becomes a key requirement. iCSA provides a large quantity of information in near real time, even in high-speed networks, distributing the computational load among different hardware and software components of its architecture.

Root cause analysis: The last major challenge is related to the quick identification of root causes by using proper metrics and analysis tools, such as "Is the service problem affecting this segment of users in the network or not?," "Is the network affecting the performance of this TCP connection?" Advanced data manipulation and presentation capabilities, together with innovative techniques for user plane analysis, allow iCSA to provide fast and accurate answers to these questions.

There are three main approaches or perspectives to *service assurance* that have emerged over time (Table 1, SR3 and SR4, and [1]): *resource-centric*, *service-centric*, and *customer-centric*. Each of these approaches has strengths and weaknesses, and no single method by itself can provide a fail-safe and effective way to CSA. Modern networks require a combination of all three assurance models to fully monitor, report, and troubleshoot problems. This combined approach, as described earlier, is adopted by the iCSA platform. Different measurement methods may be used for implementing these approaches: *element-based*, in which performance measurements are directly reported by network elements; *terminal-based*, in which software agents are placed on user equipment; and *probe-based*, in which data is passively collected, capturing the traffic flowing through the network. The latter method allows visibility on the entire multivendor network, without adversely affecting the components of the network or installing intrusive software on the user equipment. iCSA is a probe-based system.

A Platform for CSA

In order to cope with the issues presented earlier, we propose a probe-based CSA platform called intelligent customer service assurance (iCSA), which provides deep analysis and cross-relational capabilities across the network, services, devices and subscribers. As depicted in Fig. 1, the two main components of the platform are the *iCSA central server* (single or multiple) and *iCSA probes*. iCSA can be enriched with external data sources (e.g., information on devices, services, and customers) and integrated into external management systems.

As shown in Fig. 1, the iCSA central server consists of the following key subsystems: the *iCSA server platform* and *iCSA applications*. End users can access the iCSA applications by means of web clients. The iCSA server platform is a set of distributed components that receive extended data records (xDRs) from all the iCSA probes. By the term xDR, we mean a summary containing the most important information regarding each single transaction. Such a transaction could be related to both the control and user planes of a simple call, an intelligent network request, a session or mobility management transaction, a TCP connection, and so on. The iCSA server platform implements the following capabilities (more details are reported later), exploited by the iCSA and/or other OSS applications:

- Retrieval of both xDRs and raw frames from probes (captured by the data collector and by the data collector server proxy, respectively)
- Binding of xDRs pertaining to the same transaction and enrichment of their content by means of external information (performed by the binding/enrichment function).
- Computation of different measurements, at different levels, such as elementary counters and key performance indicators (performed by the data management function)
- Optimal storage of xDRs (performed by the iCSA central database)
- Interface toward external data sources and external OSS applications

iCSA applications access and manipulate data available across the iCSA platform for different purposes:

Troubleshooting: Analyze specific protocol sessions by retrieving xDRs pertaining to this session. This requires searching within a large distributed database of xDRs: for a mobile operator, several hundred xDRs per active user may be stored every day.

Multidimensional analysis: Analyze, for planning and optimization purposes, relevant counters related to the amount and quality of network procedures and services, through different multidimensional views, such as network, customer groups, device types, and areas. Counters are also combined into key performance indicators (KPIs), inheriting multidimensional views.

Proactive monitoring of the network: Monitor continuously the health of the network and related services, and trigger alarms in case of issues in key network elements, services or customer groups such as corporate accounts and VIPs.

Figure 2 summarizes the architecture of the iCSA Probe and highlights its main components:

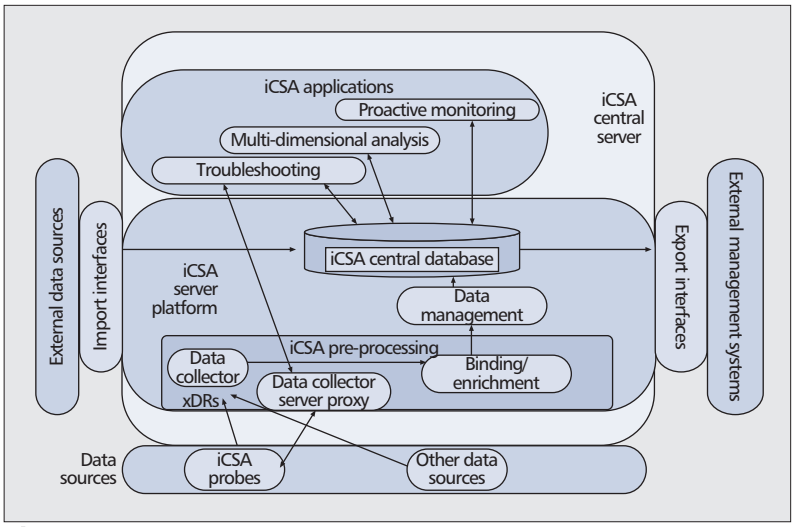

Figure 1. *High-level view of the iCSA architecture.*

Packet capturing and processing: This module interfaces with the links connecting the nodes of the network being monitored. It is a dedicated acquisition card, with custom firmware and drivers: it provides in-hardware timestamped output packets, with an accuracy of microseconds, and it is able to analyze tens of gigabits per second of traffic. Moreover, it may implement in-hardware packet filtering for the probe to analyze only the portion of relevant traffic, thus offloading the CPUs.

xDR Generator: This component decodes and analyzes timestamped packets, and generates statistics at the protocol layer (messages and events counting) and xDRs. For each new transaction, the xDR generator builds a new record and keeps it in memory. When the transaction reaches a significant phase (e.g., start and end of a call or of a TCP connection, timeout expiration, etc.), the xDR generator closes the record and stores the data in the local storage system. The xDRs are then transferred to the iCSA central server. In order to properly generate the xDRs, the state machine of each protocol is implemented, which allows messages pertaining to the same transaction to be bound (e.g., message related to the same Packet Data Protocol [PDP] context) and thus to relate subsequent xDRs, which represent the state evolution of a certain session.

Storage: This component implements an indexed storage of low-level protocol statistics and alarms, raw frames, and xDRs. The xDRs are continuously transferred to the iCSA central server, while a long-term storage of raw frames and alarms is provided by the probe.

Application Interface: This component acts as a proxy for requests coming from applications (mostly in the area of troubleshooting) that require access to statistics/alarms and frames.

The hardware of the iCSA probes is based on high-end server technology that exploits multicore processors. The number of cores ranges from 4 to 12, depending on the traffic and the complexity of the requested analysis, while RAM capacity varies in the range of 4–16 Gbytes. Storage availability can be configured in the range of

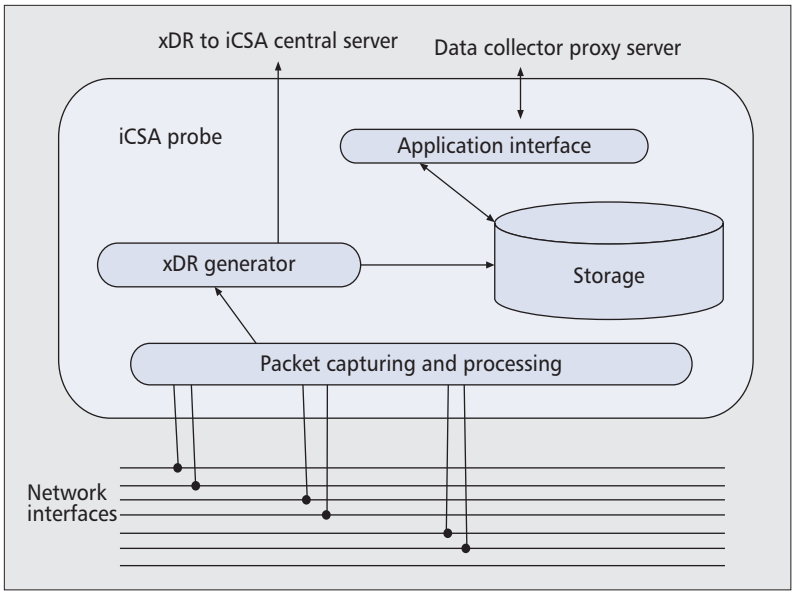

Figure 2. *The iCSA probe.*

1–28 Tbytes. The software of the probe is designed so that it tracks the continuous evolution of server technology, especially in the area of multicore processors (discussed below). Acquisition cards are specific to the transport technology in use in the network under analysis (PDH, SDH, Ethernet, etc.). For mobile broadband networks, Ethernet is the dominant technology; while Gigabit Ethernet is the most common case, 10 Gigabit Ethernet is gaining momentum. The iCSA probes can monitor protocols at any access and at any core network interfaces of 2G-3G mobile networks and evolved packet systems, and in the IP Multimedia Subsystem (IMS).

Mode of Operation

The iCSA monitoring chain starts within the probes. xDRs are continuously transferred to the server, while raw protocol messages are stored in the local probe hard disks and then transferred to the server only on demand (e.g., when users are performing detailed troubleshooting analysis and require their visibility). This avoids the exchange of large amounts of data between the probes and the server during online monitoring. In the iCSA central server, xDRs are subject to an initial preprocessing (Fig. 3) mainly to:

- Bind xDRs pertaining to the same transaction, but coming from different probes. These xDRs are identified by applying specific matching rules across parameters of xDRs (e.g., a specific match among subscriber identifiers).
- Enrich the content by means of external static or semi-static information (metadata), such as service models, segments of customers, and types of devices.

Then xDRs follow two processing chains inside the data management component of the iCSA server platform (Fig. 3):

- They are analyzed in order to verify whether they contain information on abnormal network or service conditions (e.g., a transaction failed because of a server failure), which can trigger an alarm, and then loaded into a dedicated database to be used for troubleshooting applications (right part of Fig. 3)
- They are processed by the X-Ray engine (left part of Fig. 3). This component generates multidimensional measurements called elementary counters (ECs), on both the control and user planes. These counters are combined into key performance indicators (KPIs), which can also trigger alarms in case specific thresholds violations or profiles are identified.

The application for multidimensional analysis (described before) is implemented through the following mechanisms:

Splitting over xDR dimension: ECs and thus KPIs are projected over a specific element of an xDR representing a key point of analysis, exploiting this element in its full cardinality. A relevant example could be the probing point. For instance, by comparing evolution over time of TCP round-trip time (RTT) for specific classes of services (e.g., HTTP download) in different parts of a mobile broadband network, it is possible to understand the contributions of different parts of the network to the RTT, and thus points of congestions or backhaul issues.

EC and KPI grouping: When groups are defined, ECs and KPIs are calculated both for the totals (as before) and for the specific segments corresponding to these groups, such as, device types, service categories, customer groups, etc. A group is defined by a *regular expression* over any number of fields in the xDRs (e.g., the field devoted to International Mobile Equipment Identity [IMEI]). Once a group is defined, the xDRs, ECs, and KPIs are grouped according to it and measures are calculated for each group (e.g., all mobile devices produced by a certain manufacturer). When groups are not defined, the measures are aggregated on all the xDRs. It is worth noting that a single field in the xDRs can generate multiple group types, considering different parts of the same field as different fields; vice versa, multiple fields in the input events can be used for a single group type (e.g., performing a logical AND among these fields).

One of the key functionalities implemented by the iCSA in the monitoring process just described is to maintain the link among the different layers: frames, xDRs, ECs, and KPIs (also when split and grouped as described before). In this way it is possible to drill down from measurements referring to a specific group into the specific xDRs that determine the measurement for this view, and down to the correspondent frames. The chain allows building multi-dimensional KPIs as well as intersecting different multi-dimensional views, thus implementing the CSA concept.

The iCSA platform allows to cope with the issues presented earlier. In particular, the hardware and software of the iCSA probes are designed to cope with the *increase of bit rates* in mobile broadband networks. The software is designed in a way so that it can automatically adapt to the numbers of cores available in the CPUs. Load balancing among different cores is adaptive and processing is spread based on different steps of analysis, type of protocols, range of IP addresses or other criteria. The load distri-

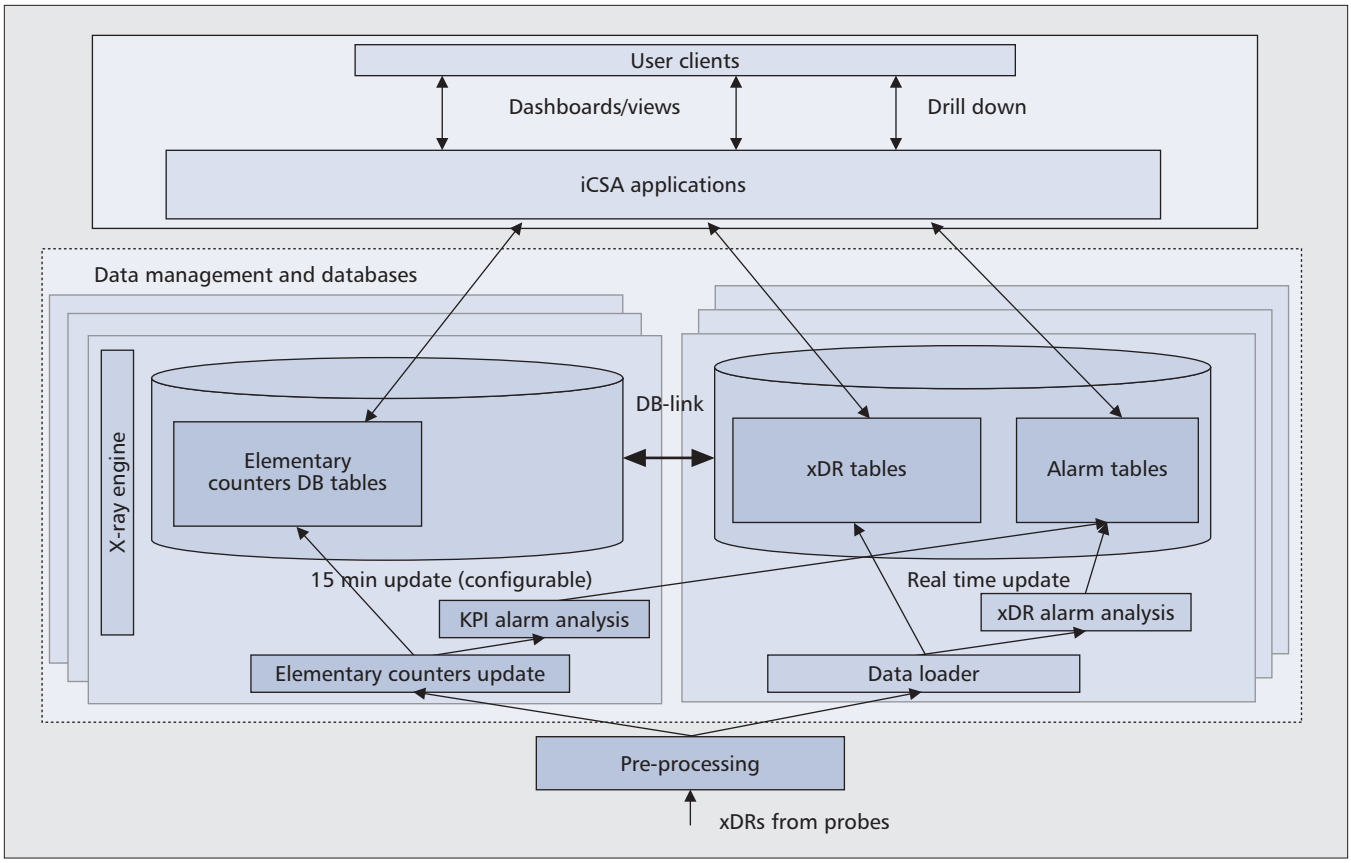

Figure 3. *Detailed view of the iCSA server.*

bution function is partially implemented directly in the acquisition card, partitioning collected traffic over different core queues. In order to have flexibility in the acquisition of user plane traffic, semi-static and dynamic filtering are applied. This capability controls the user plane being analyzed based on specific conditions defined on the control plane: for example, only the user plane pertaining to certain Access Point Name (APN) or to specific groups of customers are acquired and analyzed. For what concerns the *split of user control planes*, binding of user and control plane starts in the iCSA probes, where user plane measurements (e.g., exchanged/lost packets, throughput) are associated to a specific PDP context by applying a specific binding key that can be the IP address given to the user and/or tunnel identifiers of GPRS Tunneling Protocol User Plane (GTP-U) protocol being used at the Iu-Ps and Gn interfaces (Table 1, SR5 and SR6).

This binding continues in the iCSA central server, where xDRs coming from different probes are bound and xDRs related to the user plane are enriched with control plane parameters (customer group, device types, geographic area, etc.). Regarding the necessity to cope with *complex network architectures*, iCSA coordinates the work of different probes in a real-time fashion. A good example is the de-ciphering of control messages over the Gb (the same applies to the Iub interface). In order to give to all the probes in the network the possibility to decode these messages, it is necessary to dispatch deciphering keys both from the MAP-Gr interface or from the Gn interface in case of inter-SGSN mobility (Table 1, SR12). Similarly, deciphering of Non-Access Stratum messages at the S1 interface (Table 1, SR7) requires retrieving keys from the S6a interface and from the S10 interface in case of inter-MME mobility.

In order to provide *near-real-time availability of key metrics*, the load for the computation of metrics at different abstraction levels (from frames to ECs and KPIs) is distributed among iCSA probes and central servers, thanks to the architecture of the platform and to the monitoring chain herein described. As a result, protocol frames are continuously analyzed and consolidated in proper summaries and measures are updated in near real-time (delay is configurable according to the volume of traffic, typically it is on the order of 5 min).

Finally, the multidimensional view provided by iCSA allows a holistic root-cause analysis to enforce quality assurance in complex network and service scenarios, such as mobile broadband networks. Once a specific KPI highlights an issue (e.g., a significant increase of failures of a certain network procedure), the multi-dimensional views allow understanding dimensions and specific instances that are responsible for most of the reported problems. A detailed workflow of this approach is reported next. Even though aggregated counters seem not to highlight a problem, this could not be the case for a specific segment of users, when intersected with specific dimensions of their network or service experience. iCSA multiviews support assurance on a per segment basis. Beside this, a key aspect of

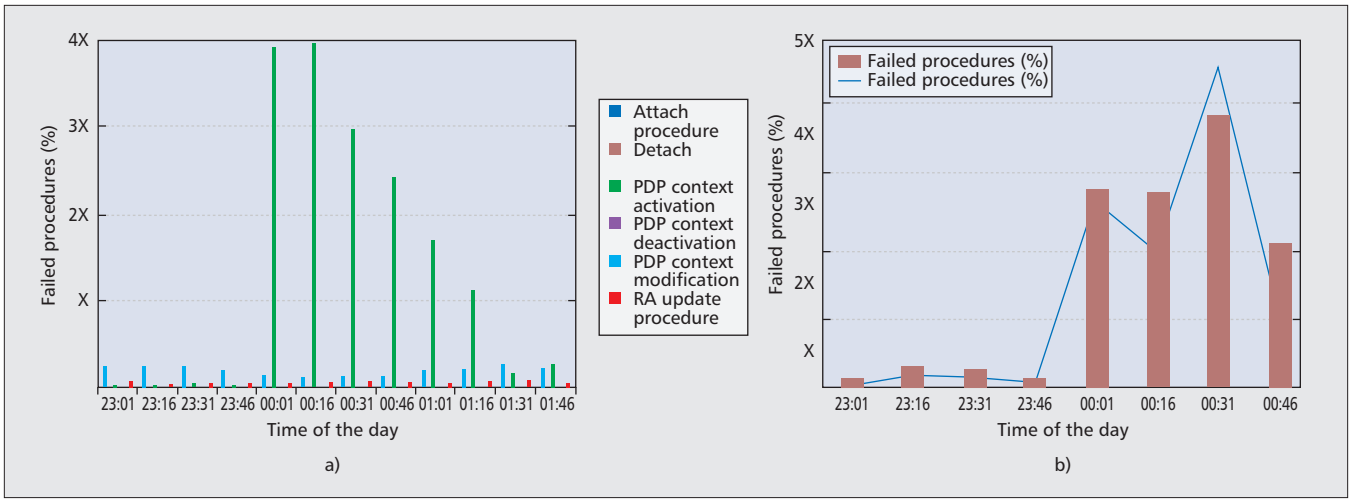

Figure 4. *Time behavior of mobility management and session management procedures: a) at Iu-PS; b) at Gn.*

the root cause analysis is the quality of basic measurements collected at the probe level and inserted in the xDRs. Later we report an example of a novel advanced approach, validated using the proposed iCSA to identify the root causes of TCP performance issues.

iCSA Platform at Work

As evidence of the goals achievable with iCSA, we present experimental results from two measurement campaigns in a real mobile broadband network. In the first example, we explain how to disclose possible network problems by properly analyzing mobility and session management metrics. In particular, we monitor a set of views of different kinds of metrics, while forcing a part of the network (only used for tests) to work in a shortage of resources. In the second example, we describe how to understand the performance experienced by users and the possible causes of limitation through the analysis of metrics extracted from the IP and TCP headers. In this case, we analyze the data traffic flowing through the operational part of the network, and results are derived only from TCP/IP headers, without any content analysis, and having anonymized all user addresses.

Mobility and Session Management: A Multidimensional Analysis

This section shows how mobility and session management metrics may be analyzed and what are in practice the views that are jointly analyzed for understanding where possible problems originate. In order to achieve this goal, we consider the trend of key measurements focusing on mobility and session management procedures (Table 1, SR8). In this experiment, traffic and configuration of the network have been purposely tuned to emphasize the aspects hereafter described. We obtained visibility on session and mobility management, thanks to the distributed nature of iCSA, probing the Gb and Iu-PS interfaces (similar results could be obtained capturing at the S1 interface of a Long Term Evolution [LTE] network). In order to track the most important events related to the interaction between users and network, xDRs have been generated monitoring the following procedures:

- GPRS Attach/Detach (explicit or implicit by inactivity timeout)
- PDP Context Activation/Modification/Deactivation
- Routing Area Update

A multidimensional representation was achieved with the following splitting mechanisms and related views:

- Device view based on proper parsing of IMEI
- Customer view based on the analysis of anonymized user identifiers, typically used to assess the quality of key corporate accounts
- Location view based on routing areas and service areas (Table 1, SR9)
- Service view based on APN

Figure 4a illustrates the evolution over time of session and mobility procedures that failed. Measurements are made at Iu-PS. As shown in the figure, the percentage of PDP context activation failures is rapidly increasing in a short period of time (from 00:01 to 00:16) and suddenly steps back to normal values (from 00:16 to 01:31). The iCSA platform makes it possible to go much deeper and see:

1) The failure is not polarized by any specific location, customer groups, and device types.

2) When splitting over different APNs considered in the traffic being analyzed, one of them (the one used for the tests) exhibits a spike in the PDP context failure.

3) When intersecting this specific APN with the information regarding the session management causes (e.g., key session management information included in protocol transaction that identifies the reason of a failure), only one appears as causing most of the failures in the observed period: *activation rejected by GGSN*.

4) A view on the measures made at the Gn interface (between SGSN and GGSN) for the Gn PDP activation procedure highlights a similar trend seen on the Iu-Ps (Fig. 4b).

This experiment shows that the multidimensional view allows to identify the worst performing elements per different dimensions or the

possible cooperation of different aspects to determine potential problems (e.g., device and APN). This enables the possibility of tracking potential issues — resulting in a loss of revenues and low customer experience — which cannot be addressed by analyzing aggregated counters. This is one of the key added values of iCSA. In this light, for instance, the identification of users highly impacted by a failure, while accessing the service, could drive quickly and easily to the most suitable operation needed to recover the issue: reconfiguration of terminal parameters to access a certain APN, tuning of the capacity on a per APN basis, and so on.

Multilayer Root Cause Analysis of TCP Connections

TCP connections experience a number of performance issues when using wireless networks. The novel multilayer root cause analysis (MRCA) of TCP connections implemented in the iCSA framework allows us to infer the performance of the TCP connections and to determine the causes that limit their throughput (also called root causes) The causes of throughput limitation can be grouped into three main categories:

- Application or user behavior
- TCP stack configuration
- Network performance

The MRCA is aimed at identifying the network-related limitations (the most important for the network operator), in some case automatically; in others, by carefully observing the results. This approach improves and integrates different techniques proposed in the literature (see e.g., [8, 9]), and it is based on a number of xDRs calculated for three different points of view: *aggregate*, *connection*, and *host*. In the following, we report the results of the MRCA on the mobile broadband network cited before.

Aggregate: Figure 5a shows a summary of the results. The retransmissions are on the order of 2 percent, which is acceptable for a cellular network [10]. The impact of these retransmissions on the performance of the users is further analyzed below. Additionally, what is really surprising is the high number of connections having packets with the TCP reset flag set. Grouping the connections by xDR fields (retransmissions, reordering, number of packets, etc.), we identify three main causes:

- Mobile stations trying to open TCP connections toward closed ports (mostly due to malware and unwanted traffic)
- Mobile stations experiencing bad network conditions, resulting in reordering and retransmissions of packets (the RST packets were sent after receiving unexpected packets)
- Mobile stations using a non-standard TCP implementation (the TCP RST packets were sent after receiving the TCP FIN packet)

In general, we could verify that those events were not impacting the throughput of the users.

Connection: This analysis reveals a number of TCP connections whose performance is not dominated by the application. Such connections are identified thanks to the methodology proposed in [9]. In particular, we use this methodology to divide the parts of the connections for which the application does not stress the network enough from those for which the application always sends large-sized packets at a high rate. However, due to the time-varying conditions of the cellular channel, this methodology does not allow to identify the root cause, we can only exclude the causes related to the application. The reasons are reported in the *host* analysis.

Host: Thanks to the *connection* analysis, we could identify the set of connections whose performance is not limited by the application. In the following, we show the results of the host analysis of two mobile stations whose connections are in this set. Figures 5b and 5c illustrate the time behavior of the throughput, retransmission ratio (also called retransmission score: it is the percentage of packets that are retransmitted), RTT, and number of parallel connections of these mobile stations.

The first user (Fig. 5b) downloads about 20 Mbytes from a web server on a single connection, and at the same time, it opens a few other connections, transferring a small amount of bytes. Comparing the time behavior of throughput and retransmission score, we see how this last parameter influences the throughput: for most of the samples, when the throughput decreases, the retransmissions increase. If we only look at the retransmission score, however, we can see that it is generally quite low, with a few spikes. This explains why the connection analysis did not identify the root cause of the performance: the retransmission score averaged over the entire period is not high enough to be identified as the limiting cause. Comparing the other two plots, we can observe the effect of the buffering (in the mobile network and in the TCP stack): when more than two connections are active, the RTT increases.

The results for the second user are reported in Fig. 5c. For this mobile station we observe a different behavior:

- The throughput is higher, reaching values up to 2.5 Mb/s.
- The number of retransmissions is also higher, and they do not correlate with the throughput.
- The RTT is very high and reaches a few seconds.

Basically, this user is able to reach the maximum throughput allowed to him by the network, but the buffering is playing a big role, and the RTT increases, which may cause problems to some applications.

Finally, we can state that the throughput of the connections of these users is limited by the network, and that the multilayer analysis allowed us to easily identify this limitation.

The root cause analysis of TCP connections permitted to spot some issues that affect the experience of the users. While for some of these issues (e.g., a high number of retransmissions in the network) it is possible to set up an alarm that automatically alerts the operator when the aggregated counter reaches a certain threshold, some other steps of the analysis (e.g., the host analysis) still require the observation of the results by an expert. We are currently working toward the identification of other xDRs (e.g., correlation coefficient of the throughput and

TCP connections experience a number of performance issues when using wireless networks. The novel Multi-Layer Root Cause Analysis of TCP connections implemented in the iCSA framework allows to infer the performance of the TCP connections and to determine the causes that limit their throughput.

Connections analyzed	159445
Packets analyzed	5379640
Connections with retransmitted packets in either direction	36608 (23%)
Connections with 1 Byte retransmitted in either direction	17389 (11%)
Connections with reset packets in either direction	47851 (30%)
Reordered packets	4327
Out of sequence packets	13085
Packets retransmitted	100706 (2%)
Packets retransmitted for timeout in uplink	36465
Packets retransmitted for fast retransmit in uplink	3937

(a)

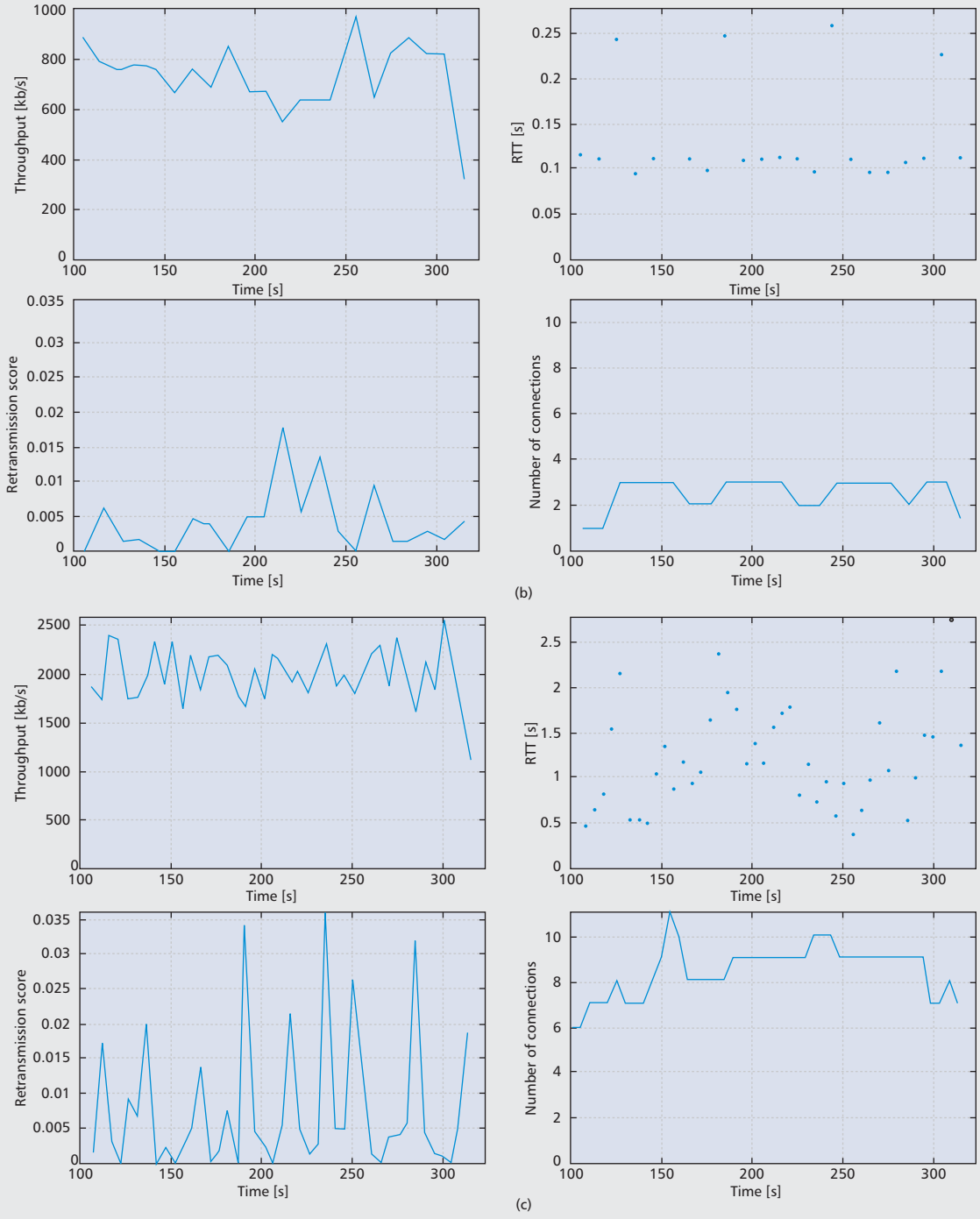

Figure 5. *Results of the Multi-Layer Root Cause Analysis of TCP connections.*

retransmission score time series) that allow the automatic detection of these and other issues.

Discussion and Conclusion

In this article, we review the main trends, issues and requirements of CSA architectures, and propose a novel CSA platform called iCSA. The importance of the topics discussed in this article and relevance of the novel features of iCSA are also highlighted, on one hand by the debate involving NGMN, 3GPP, and TM Forum on operations in next-generation networks and how they are reflected in standardization bodies [2], and on the other hand by the proliferation of advanced service assurance platforms. The proposed solution relies on several innovative contributions, ranging from the adopted distributed solution to the probe architecture, from the new multi-dimensional approach to the advanced root cause analyses.

As for the multidimensional approach, performance measures reported by network elements are typically not separated for relevant dimensions (e.g., customer segments, device type, services etc.), but simply provide an aggregation of different events on a per node-area. iCSA xDRs track user interactions with the network at different protocol layers, thus allowing a definition of a customer/service-centric data model, and to aggregate data on such relevant dimensions. As for the root cause analyses, if a network counter reports an issue, the most relevant information provided by a network element is the reject cause from control-plane related events. The multidimensional view provided by iCSA identifies dimensions and specific instances that mostly aggregate the problem in the control and user plane. Finally, the iCSA architecture allows a detailed analysis by recovering xDRs to troubleshoot a certain problem, and correspondent packets for a deep protocol analysis. iCSA data model allows to define workflows that bind proactive monitoring, multidimensional analysis and troubleshooting.

We plan to extend and improve iCSA according to several directions. First, we plan to fully automate the root cause analysis presented in this article. Second, we envisage progress of the iCSA probes in line with the trends of network technologies, line speed and hardware architectures. Third, we plan the further integration with other systems (e.g., policy and charging rules function [PCRF] systems). Fourth, we plan to improve iCSA functionalities to include a link between root cause analysis and higher level analysis (e.g., quality of experience [QoE]).

References

[1] Stratecast Perspectives and Insight for Executives (SPIE) SPIE 2011 #07 – Feb. 18, 2011, "Adaptive Customer Service Assurance – Measuring and Managing the Level of Customer Service Quality"

[2] NGMN TOP OPE Recommendations, NGMN Alliance, Version 1.0, 21 Sept. 2010.

[3] A. Kind *et al.*, "Advanced Network Monitoring Brings Life to the Awareness Plane," *IEEE Commun. Mag.*, vol. 46, no. 10, Oct. 2008, pp. 140–46.

[4] Analysis Mason "Service Assurance Systems: Worldwide Forecast 2010–2014," July 2010.

[5] Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2010–2015, Cisco white paper, Feb. 2011.

[6] S. Dixit, Yile Guo, and Z. Antoniou, "Resource Management and Quality of Service in Third Generation Wireless Networks," *IEEE Commun. Mag.*, vol. 39, no. 2, Feb. 2001, pp. 125–33.

[7] P. Agrawal *et al.*, "IP Multimedia Subsystems in 3GPP and 3GPP2: Overview and Scalability Issues," *IEEE Commun. Mag.*, vol. 46, no. 1, Jan. 2008, pp.138–45.

[8] F. Ricciato, "Traffic Monitoring and Analysis for the Optimization of A 3G Network," *IEEE Wireless Commun.*, vol. 13, no. 6, Dec. 2006, pp. 42–49.

[9] M. Siekkinen *et al.*, "A Root Cause Analysis Toolkit for TCP," *Computer Networks J.*, vol. 52, issue 9, 26 June 2008, pp. 1846–58.

[10] W. Lum Tan, F. Lam, and W. Cheong Lau, "An Empirical Study on the Capacity and Performance of 3G Networks," *IEEE Trans. Mobile Computing*, 2008, pp. 737–50.

Biographies

Antonio Pescapè [SM] (pescape@unina.it) is an assistant professor at the Department of Computer Engineering and Systems of the University of Napoli Federico II. He received his M.S. Laurea degree in computer engineering and his Ph.D. in computer engineering and systems, both from the University of Napoli Federico II. His research interests are in the networking field with focus on models and algorithms for Internet traffic, network monitoring, measurements and management of heterogeneous IP networks, and network security. He has coauthored over 90 journal (*JSAC*, *IEEE Communications Magazine*, *IEEE Wireless Communications*, *IEEE Network*, etc.) and conference (SIGCOMM, IMC, ICC, GLOBECOM, etc.) publications, and he is co-author of several patents pending. He has served and serves on more than 60 technical program committees of IEEE and ACM conferences. He has served as an Editorial Board Member of *IEEE Communications Survey & Tutorials* and has been Guest Editor for *Computer Networks* (Special Issue on Traffic Classification and Its Applications to Modern Networks). He was a Program Chair of IEEE AINA 2009. In 2009 he was awarded with the IET Communications Premium Award, and in 2010 with the best local paper award at IEEE ISCC 2010.

Alessio Botta [M] received his M.S. Laurea degree in telecommunications engineering from the University of Napoli Federico II and his Ph.D. in computer engineering and systems from the same university. From February to August 2009 he visited the Networking and Security Department at Eurecom, Sophia Antipolis, France, working on the monitoring of 3G networks. Currently he holds a postdoctoral position at the Department of Computer Engineering and Systems of the University of Napoli Federico II. His research interests fall in the area of networking, with specific regards to network monitoring and measurements.

Claudio Guerrini received his M.S. Laurea degree in telecommunications engineering in 1998 from the University of Bologna (with honors) and an M.B.A. from Almagraduate school-University of Bologna in 2008. He joined CSELT (now Telecom Italia LAB) in 1999, working in the area of 3G radio access networks. In 2002 he became a company delegate in 3GPP WG4, following standardization from Release 99 to Release 8 (LTE), being rapporteur of the specifications for Radio Resource Management (3GPP TS 25123/25133). In 2006 he joined Sunrise Telecom PPG (now Accanto Systems) as a product manager for monitoring systems. He filed several patents in the area of mobile networks.

Marin Mangri is a Diplomat Engineer in electronics and telecommunications from Technical University Gheorghe Asachi-Iasi (Romania), where he graduated in June 1989. Between 1989 and 1997 he was involved in many fields of electronics, telecommunication, and computers programming being a technical product manager in an audio-TV factory and a designer of mobile radio networks. From October 1997 to December 2000 he held a BSS-Expert position at Orange-Romania, a GSM operator. In 2001 he joined Telefonica Germany (O2) as an NSS specialist and starting in 2005 till present, he acts as lead tracing and protocol optimization specialist in the same company. Currently he is also a Ph.D. student in the Faculty of Electronics and Telecommunications of the University "Politehnica" Timisoara (Romania) working in the area of tracing optimization of real-time protocols in IMS.

The proposed solution relies on several innovative contributions, ranging from the adopted distributed solution to the probe architecture, from the new multi-dimensional approach to the advanced root cause analyses.