Adaptive Terminal Reporting for Scalable Service Quality Monitoring in Large Networks

Jan Groenendijk, Yangcheng Huang and Liam Fallon
Network Management Lab, LM Ericsson
Athlone, Co. Westmeath, Ireland
Email: Jan.Groenendijk, Yangcheng.Huang, Liam.Fallon (at)ericsson.com

***Abstract*—Terminal reporting provides the most accurate way of evaluating how a customer is experiencing a service but building a scalable service quality monitoring solution is not easy because service usage patterns and network conditions are unpredictable. The maximum resource usage of existing terminal reporting solutions is statically constrained, impairing their effectiveness in problem diagnosis when presented with highly variable session rates. This paper revisits the fundamental principles of terminal reporting and presents an adaptive terminal reporting scheme for scalable service quality monitoring in large networks. The main novelty of this approach is that it automatically coordinates terminal reporting activities across all services being monitored and dynamically adapts the reporting characteristics to network and service conditions.**

I. INTRODUCTION

Today's complex and constantly-evolving networks put considerable strain on service assurance. At the same time, accurate and actionable insight into customer service experience has never been more essential for operators competing in a market increasingly driven by end user expectations [1]. The most reliable way to determine the true quality of a service is to monitor that service at the point where that service is delivered; that is at the terminal that is running a service.

Terminal reporting measures service quality metrics at the terminal and reports them back to an entity in the network that records and collates them centrally. By measuring at the terminal, the operator can measure fluctuations in delivered service quality, allowing implementation of measures to reduce churn, which is a large problem for operators.

A number of initiatives for terminal reporting are or have been standardized by IETF [2], [3], 3GPP [4], [5], [6], BroadBand Forum [7], [8], [9] and ISO/IEC [10]. However, large-scale introduction of these solutions is not taking place, one of the reasons being the bandwidth load of terminal measurements on the transport capacity of the network.

While each individual solution includes mechanisms allowing an operator to control measurement reporting load, operators with multiple end-user services deployed have mixed terminal reporting solutions. This means that operators must manage a number of mechanisms manually to control reporting bandwidth load, leading to increased operational costs.

Another drawback of current solutions is that their maximum resource usage is statically constrained, impairing their effectiveness in problem diagnosis when presented with highly variable session rates.

While an extensive body of prior work on terminal reporting mechanisms for individual services exists, very little examination of unified quality reporting solutions for multiple services for network-wide holistic problem diagnosis has taken place. The need for such solutions has become increasingly imperative and urgent where operators running mobile broadband applications demand quick, effective, and efficient problem diagnosis. There is a pressing need for techniques that can dynamically adapt reporting characteristics to network and service conditions.

This paper presents an adaptive terminal reporting scheme for scalable service quality monitoring in large networks. The proposed solution includes two stages, a *zoom-out* reporting stage and a *zoom-in* analysis stage.

In the *zoom-out* stage a reduced level of terminal reporting is used, maintaining low reporting traffic overhead. The system stays at *zoom-out* monitoring stage unless monitored terminal service sessions indicate a degradation of service quality in at least some terminal service sessions. When this happens, the system enters the *zoom-in* stage.

In the *zoom-in* stage, the system selectively collects reception reports from terminals that satisfy certain criteria to analyze and pinpoint causes of the quality problems observed in the zoom-out stage.

The main novelty of the proposed scheme is that it automatically coordinates terminal reporting activities across *all* services being monitored and dynamically adapts the reporting characteristics to network and service conditions [1].

The rest of the paper is organized as follows. Section II gives a brief summary of related work in the area of service quality reporting from terminals. Section III describes the proposed *zooming* scheme in details. Section IV shows how the algorithms shall be deployed in a terminal reporting system and outlines an illustrative use case for the approach. Section V gives conclusions and describes further work currently being undertaken.

II. BACKGROUND AND RELATED WORK

This section presents a brief non-exhaustive summary of existing work in the field of service quality terminal reporting.

---

[1] It is worth noting that security aspects of the proposed scheme are out of the scope of this paper. The authors rely on the appropriate standardization bodies to take care of authentication, authorization and integrity of terminal reporting.

A. IETF

RTCP (Real-time Transport Control Protocol) [2] periodically transmits control packets to participants (including multiple content senders and receivers) in a RTP based streaming multimedia session, enabling group size estimation and the distribution and calculation of session-specific information such as packet loss and round trip time to other hosts.

As defined in RFC 3550, RTCP traffic is limited to a small and known fraction of the RTP session bandwidth: small, so that the primary function of the transport protocol which is to carry data is not impaired and known, so that control traffic can be included in the bandwidth specification given to a resource reservation protocol and that each participant can independently calculate its share. It is recommended that the fraction of the session bandwidth added for RTCP be fixed at 5%. If it is assumed that terminal reporting takes 75% of bandwidth, the interval at which receiver reports are sent is:

$$\text{reporting interval} = \#\text{receivers} \times \frac{\text{average RTCP packet size}}{0.75 \ \times \ \text{RTCP bandwidth}}$$

RFC 5760 [3] proposes a hierarchical terminal report aggregation model, aiming at increasing the maximum number of users limit further and enabling user Quality of Service (QoS) measurement.

Note that RTCP and its extensions are not limited to multicast sessions. The session can be unicast or multicast. The protocols may also be applicable to services other than RTP streaming.

B. 3GPP

3GPP QoE reporting mechanisms [4] [5] specify that QoE metrics be reported after a service session via a reception reporting procedure. The specification for Multimedia Telephony [6] describes a mechanism to randomize terminal reporting for a population of terminals.

C. Broadband Forum

TR-069 [7] specifies a communication protocol between an Auto-Configuration Server (ACS) and Customer Premise Equipment (CPE). The ACS server in the service provider's network has the ability to control and monitor CPE with the TR-069 protocol. The IPTV [8] and VoIP [9] equipment data models are being updated to support QoE reporting.

D. ISO/IEC

The ISO/IEC Moving Picture Experts Group (MPEG) working group is currently standardizing MPEG media over HTTP streaming. Terminal reporting is included in solutions such as DASH (Dynamic Adaptive Streaming over HTTP) [10] with the semantics of quality metrics being defined at each observation point. Similar to 3GPP approaches, the delivery protocol for quality metrics should be HTTP.

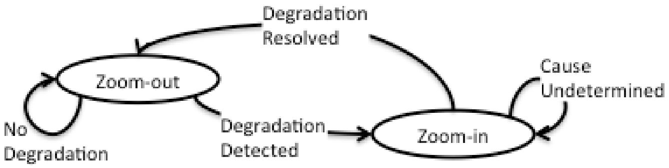

Fig. 1. Zoom-in and Zoom-out reporting

III. ADAPTIVE TERMINAL REPORTING SCHEME

Terminal reporting gives an extremely accurate view of the end user service quality experience but it is expensive in its use of terminal and bandwidth resources. This scheme uses two reporting stages to resolve this apparent conflict, a *zoom-out* reporting stage and a *zoom-in* analysis stage. During normal operation, zoom-out reporting synopsizes the service quality being experienced by users in an extremely efficient manner, using terminal reporting on a small representative set of terminals. When service degradation is detected the system turns on zoom-in reporting, using more detailed terminal reporting for a focused set of services and locations where degradation has occurred for detailed problem detection and resolution. Fig. 1 shows a state transition diagram which illustrates transitions between zoom-out and zoom-in reporting.

A. The Zoom-out Stage: Reporting

To reduce the reporting traffic volume, it is crucial to determine a minimum set of services that should be monitored in order to get a representative picture of overall service quality. To achieve this, the proposed scheme analyzes reporting coverage of each service.

The quality metrics reported in terminal reports for a particular service reflect the status of the paths between terminals and the corresponding servers for that service. Therefore, the *reporting coverage* of each service can be defined as *the set of physical and logical network entities whose status can be reflected by terminal reports of that service*.

There is inherent redundancy in the reporting coverage for all services in a network because different services running on the same terminals and network will reflect the same network conditions. Zoom-out reporting uses an algorithm to eliminate that redundancy so that monitoring need only be activated for that service in each network segment which gives the most representative view of user service quality in that network segment. Conventional statistical techniques such as those used by the 3GPP [6] can then be applied to sessions of a service selected for monitoring to further reduce the resource requirements of terminal reporting.

The algorithm is now described. The aim of the algorithm is to identify a *minimum* set of services whose reports satisfy the monitoring requirements and are a representative sample of the status of all elements in the monitored network. That is, the union of their report coverage *contains* all of the monitored entities.

The algorithm has two sets of input: a set U with n elements containing the network resources to be monitored

```
MinCoverServiceSet(U , S)
{
 remove services from S with very low coverage
 FOR each time window w
   REPEAT
      select service S[i] from S as a candidate
             service, satisfying, within w, S[i] covers the
             maximum number of uncovered elements of U
      mark those elements in U covered by S[i] as covered
  UNTIL all elements in U are marked as covered
}
```

Listing 1. Minimum Coverage Set Algorithm

and a collection S with m sets, each contains the reporting coverage for one of m services.

The content of set U is populated using operator inputs and configuration data sources such as topology. The operator specifies the nodes and resources to be monitored such as radio access nodes, backbone routers, or transport links.

$$S = S_1, S_2, ..., S_m$$

The reporting coverage in collection S is a collection of resource coverage sets, each containing the reporting coverage for one of m services. The elements of a coverage set S_i of S may be any network resource. The coverage set for a service is populated by analysing collected terminal reports and correlating them with other data sources such as topology, call records, or packet inspection traces from probes. Consumption patterns such as temporal distribution and volume of service access may also be included in the coverage set of the service. Examples of coverage sets are as follows:

$$\begin{aligned} S_{mobileTV} &= \{Cell(c_1, c_2, c_3), SGSN(sgsn_1), GGSN(ggsn_1, ggsn_2)\} \\ S_{WB} &= \{Cell(c_4, c_5), SGSN(sgsn_3), GGSN(ggsn_6, ggsn_7), \\ &\quad Time(09:00-11:00), Usertype(1001)\} \end{aligned}$$

The reporting coverage set of the Mobile TV service $S_{mobileTV}$ covers Cells c_1, c_2 and c_3, SGSN [2] node $sgsn_1$, and GGSN [3] nodes $ggsn_1$ and $ggsn_2$. The Web Browsing reporting coverage set S_{WB} covers Cells c_4 and c_5, SGSN node $sgsn_3$, and GGSN nodes $ggsn_6$ and $ggsn_7$ during the time period 09:00 to 11:00 for users of type 1001.

Determining the minimum coverage service set, i.e. the minimum set of services whose reports represent the status of the monitored network, is a minimum set cover problem. Such problems are known to be NP-Hard, for which there is no fast algorithm to solve. The problem to be solved here is essentially more complex since elements in the sets are heterogeneous and multi-dimensional, but with relaxed requirements in terms of number of subsets. A greedy approximation algorithm shown in Listing 1 is proposed as the best mode implementation and suffices for practical purposes.

Services with very low coverage of nodes or low usage are removed, based on pre-defined thresholds. The algorithm loops over a list of predefined time windows and identifies the minimum cover service set for each time window. The final step in the algorithm selects the service with the lowest service access rate as the service to use for monitoring.

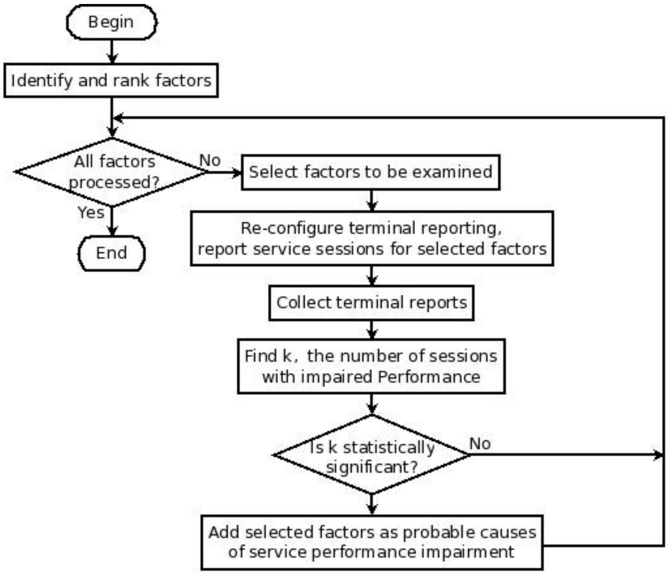

Fig. 2. The Zoom-in Algorithm

TABLE I
TYPICAL SERVICE DEGRADATION FACTORS

Factor Category	**Factor**
Service	IPTV VoIP Web Browsing
Terminal	Terminal Type Manufacturer Software Version
Network	Cell Equipment Logical Entity

The proposed algorithm generates the minimum cover set for each time window. The identified sets over different time periods may either be merged into one service set which is used to configure terminal reporting in the network as a once off operation. Alternatively, the minimum cover set may be used to reconfigure terminal reporting on the network at the beginning of each time window.

B. The Zoom-in Stage: Analysis

In the *zoom-in* stage, the system collects reception reports from all terminals that satisfy certain criteria to determine causes of quality problems that have been observed in the *zoom-out* stage. The zoom-in algorithm is shown in Fig. 2.

The zoom-in stage uses the concept of *factors* to determine on what services and network segments detailed terminal reporting should be enacted. A factor is an aspect of a service that might identify a common cause of service degradation. Table I lists some typical factor categories and factors.

---

[2] Serving GPRS Support Node

[3] Gateway GPRS Support Node

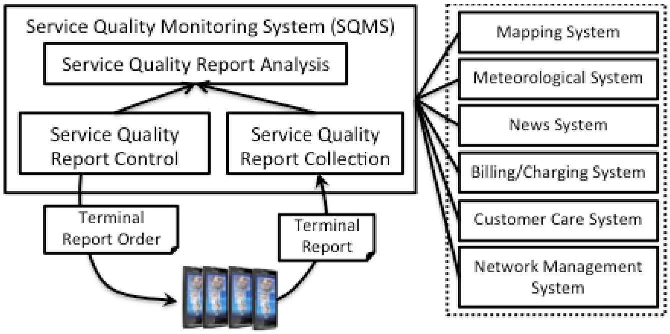

Fig. 3. Service Quality Monitoring System

Factor analysis is carried out on sessions monitored in the zoom-out stage whose terminal reports indicate degraded quality. The terminal reports are correlated with other data sources such as service, terminal, or network information using IDs in terminal reports like the IMSI (International Mobile Subscriber Identity). Once a list of factors has been compiled, the factors are ranked based on the number of degraded sessions associated with each factor. There is a degree of uncertainty in the ranking because the ranking is based on a sample of terminal reports taken from the zoom-out stage.

The algorithm then processes each factor in order. One or more factors are selected and the network is instructed to turn on full terminal reporting for the services and network segments identified by those factors.

The resulting terminal reports are analysed statistically assuming a normal distribution, so a sufficiently large sample of reports (at least 30, but ideally many more) is gathered. Once the report sample is gathered for the factors, the number of service sessions with impaired performance in the sample is calculated. If the service sessions with impaired performance are caused by the factors in question, the number of reception reports reflecting impaired performance shall be statistically significant among all reception reports collected for those factors, given a certain confidence level.

The algorithm outputs a list of factors that have been identified as probable causes of service impairment. That list can be used to raise a service quality alarm, to generate a report, or to activate automated root cause analysis.

IV. DEPLOYMENT AND USE CASE

In previous work [11], we described a Service Quality Management System (SQMS), the server side component in a terminal reporting system. A SQMS has three main functions, as illustrated in Fig. 3. Adaptive terminal reporting is deployed in the Service Quality Report Analysis component of the SQMS, which has a feature that implements the state transition diagram in Fig. 1 as well as the zoom-out and zoom-in algorithms described in Section III.

A typical use case that illustrates the use of adaptive terminal reporting is as follows.

A large operator monitors a mobile network using the proposed scheme. Zoom-out monitoring shows that terminals of certain cells in certain areas are experiencing decreased service quality, indicating possible problems with service delivery. Zoom-in analysis is activated automatically, increasing the number of terminal reports from the problematic cells and areas and ranking the possible causes.

The selected factors are:

- Origin, model and serial number of the device (based on IMEI [4])
- Software version of the device (based on IMEISV [5])
- Types of services currently running inside the monitored network
- Mobility characteristics of the device (walking, running, cycling, driving etc)

Based on the reports collected from zoom-in monitoring, the factor analysis results show that certain types of terminals have a higher than expected level of signalling traffic with the network, increasing the load significantly. As the popularity of those phones increases in particular areas, the quality of service delivery decreases. Laboratory tests confirm a software fault in the terminals in question. A software upgrade is then deployed to fix the fault.

V. CONCLUSIONS AND FURTHER WORK

The adaptive terminal reporting approach described in this paper provides substantial advantages over existing approaches. Rather than using a statically configured budget of bandwidth as other approaches do, it uses bandwidth in a frugal manner during normal operation. When problems are detected, it targets only those areas, enabling the richness of full terminal reporting. This allows operators to optimize their terminal reporting bandwidth budget and still have improved insights when abnormal situations occur, thus enhancing the quality of service delivery to service consumers.

We are evaluating the effectiveness and scalability of adaptive terminal reporting as an approach for service quality monitoring. Two specific topics are of particular interest: we wish to quantify minimum bandwidth usage to give a representative synopsis of end user service quality in zoom-out mode, and we wish to determine how tightly zoom-in reporting can be parameterized to accurately diagnose problems and to minimize potential negative impact on end-user sessions.

ACKNOWLEDGMENT

This work was partially supported by Science Foundation Ireland under the FAME SRC Grant No 08/SRC/I1403.

The authors would also like to thank Johan Granath and Piotr Stasiewski at Ericsson for their support with this work.

REFERENCES

[1] Ericsson. Keeping the Customer Service Experience Promise. Ericsson. [Online]. Available: http://www.ericsson.com/res/docs/whitepapers/wp_service_assurance.pdf

[2] H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson, "RTP: A Transport Protocol for Real-Time Applications," *RFC 3550*, July 2003.

[4] International Mobile Equipment Identity

[5] IMEI Software Version

[3] J. Ott, J. Chesterfield, and E. Schooler, "RTP Control Protocol (RTCP) Extensions for Single-Source Multicast Sessions with Unicast Feedback," *RFC 5760*, February 2010.

[4] "Transparent end-to-end Packet-switched Streaming Service (PSS); Protocols and codecs," *3GPP TS 26.234*, December 2010.

[5] "Multimedia Broadcast/Multicast Service (MBMS); Protocols and Codecs," *3GPP TS 26.346*, October 2010.

[6] "IP Multimedia Subsystem (IMS); Multimedia Telephony; Media Handling and Interaction," *3GPP TS 26.114*, December 2010.

[7] J. Blackford, H. Kirksey, and W. Lupton, "CPE WAN Management Protocol," *TR-069 Amendment 3*, November 2010.

[8] M. Quacchia, W. Lupton, G. Straub, G. Bathrick, and H. Kirksey, "Data Model for a TR-069 Enabled STB," *TR-135 Issue 1*, 2007.

[9] J. Bernstein and B. Stark, "DSLHome Provisioning Parameters for VoIP CPE," *TR-104*, 2005.

[10] ISO/IEC JTC 1/SC 29, *Information Technology, MPEG systems technologies Part 6: Dynamic Adaptive Streaming over HTTP (DASH)*, ISO/IEC Std., 2011.

[11] L. Fallon and Y. Huang, "GSQR: A Generic Service Quality Reporting Protocol for Terminals," *Proc. 4th IFIP/IEEE International Workshop on Distributed Autonomous Network Management Systems (DANMS 2011), to appear*, May 2011.