SPECIAL TOPIC

SLA Implementation Technologies and Their Applications

Abstract:
With more and more network services developed, users now have higher requirements on the Quality of Service (QoS). Service Level Agreement (SLA) is thereby proposed to manage telecom services with guaranteed QoS and to address the QoS issues between Service Providers (SP) and users. SLA representation template, violation process and metrics evaluation are three key techniques to implement SLA. A typical SLA management system includes SLA data management, SLA problem management, and SLA management. SLA research is still in its primary stage and SLA management is yet to be standardized. With general and special management approaches to be defined in the future, the unified industry standard will finally take shape.

Zhang Ruoying
Xu Zhifa
Lu Chuncong
(Telecommunication Planning Research Institute of MII, Beijing 100037, China)

As Quality of Service (QoS)[1] becomes increasingly important, Service Level Agreement (SLA)[2] is proposed to manage telecom services with guaranteed QoS in order to straighten up the service quality issues between Service Providers (SP) and end users. An SLA is a formal contract (or part of an agreement) between an SP and a user that specifies, usually in measurable terms, the service quality level, priority, and obligations. It is a standard evaluating telecom service.

The objective of SLA is to build up a healthy network operation environment where users can enjoy services guaranteed by regulations instead of verbal "promises", therefore, their rights are well protected. With the SLA, sophisticated telecom carriers are able to attract VIP users who bring stable revenues while start-up carriers may sharpen their competitive edges.

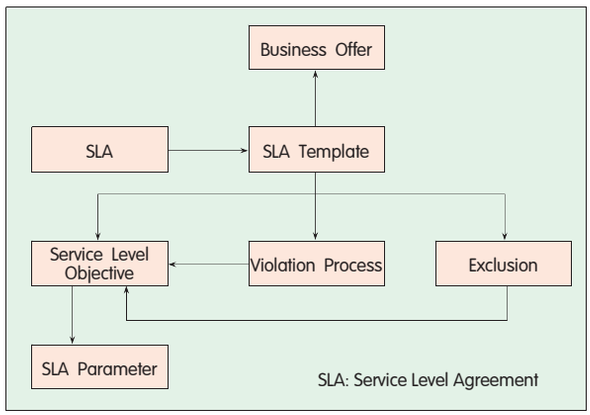
An SLA involves SLA parameters' definition and calculation, SLA representation method, and SLA management methods. Presently, numerous research works are done on SLA. TMF701 of TeleManagement Forum (TMF) discusses two management methods—SLA parameter frame work and SLA life cycle; GB917[3] defines service availability parameter and performance report contents not referring the representation; IETF drafts[4,5] put forward the necessity and importance of SLA representation. However, the key implementation technology and system application of SLA (e.g., SLA parameter selection/measurement, SLA representation method/format, and SLA violation process) are yet to be defined.

1 SLA Management Requirements

In order to exactly define the requirements of SLA management, there should be in-depth comprehension and analysis. The SLA management requirements are involved in four aspects: service fulfillment, service assurance, user interface management, and other management needs. The first three comes from the definition of enhanced Telecom Operations Map (eTOM).

In the service fulfillment stage, SLA management focuses on SLA negotiation and subscription. The SLA should clearly define the following: measurable service performance metrics and parameters that are understandable to the user; user and SP obligations; service performance measurement method, measurement cycle, and report cycle; SP operations triggered by SLA violation; service-related report types (including report content, format, destination address, conditions, and transfer mode); SLA services' definitions and expiration dates. Concerning of any services, a user should be able to select the parameters to be guaranteed and the value range of parameters.

The service assurance stage is when the services have been configured and the SP is providing its user with service guarantee. The SLA in this stage should monitor the service quality level and offer information report to the user. The SP must follow the SLA to monitor and measure the actual service performances in a range acceptable to the user or an entrusted third party. All SLA parameter-related user-oriented service information must be sent in time to the



▲Figure 1. SLA template composition.

user according to SLA terms. The SP should set soft threshold for each parameter and give warning whenever the threshold is exceeded. According to SLA terms, the user should also be told the possibility of SLA violation caused by service degradation.

For user interface management, the SLA needs pay attention to interfaces between the user and the SP, as well as how the SP responds to user services and SLA queries. The SP should respond quickly to the queries about the level of user service quality. Meanwhile, the user should be able to report problems/faults, request for processing, query service status through phone, fax and email means, and get answers/responses through various ways.

Other management requirements include:

• SLA should define and uniquely identify each service module;

• The performance report should use the service identifiers defined in SLA;

• The exceptional service or performance processing and user obligations should also be defined, such as the priority selection mode for reporting problem to the SP and provision of the contact mode.

2 Relationships between SLA and QoS/Network Performance

The network performance is divided into service performance and network performance on account of different objects, attributes, functional scopes, and measurement scopes. The two performance levels are described in performance factors, which in turn is the combination of different parameters.

The SLA is a formal and negotiatory agreement between SPs and users to define services, priority, and responsibility, which focuses on the service level. Figure 1 shows the basic 3 aspects of the SLA template: service level objective, violation process and force majeure statement. The service level objective is described with defined SLA parameters. The SLA parameters include QoS, service level priority, parameter weights, and other high–level parameters obtained by computing existing SLA parameters to evaluate the overall operation quality, such as the parameter of service availability.

ITU–T E.860[6] defines QoS as the consistent degree of service performance and negotiated terms in SLA. The QoS level is obtained by comparing the objective QoS to the measured QoS which is used to value the overall service performance. QoS is the metrics of a specific service for the NGI services, negotiated and defined in SLA. It is related to service quality level and network status and should be guaranteed. It is all or part of the performance factors that define service performance and network performance. QoS parameters can be used or ignored according to the actual circumstance. Therefore, SLA, QoS, and IP network performance collectively evaluate the network quality and service quality in a scientific and measurable way.

3 SLA Implementation Technologies

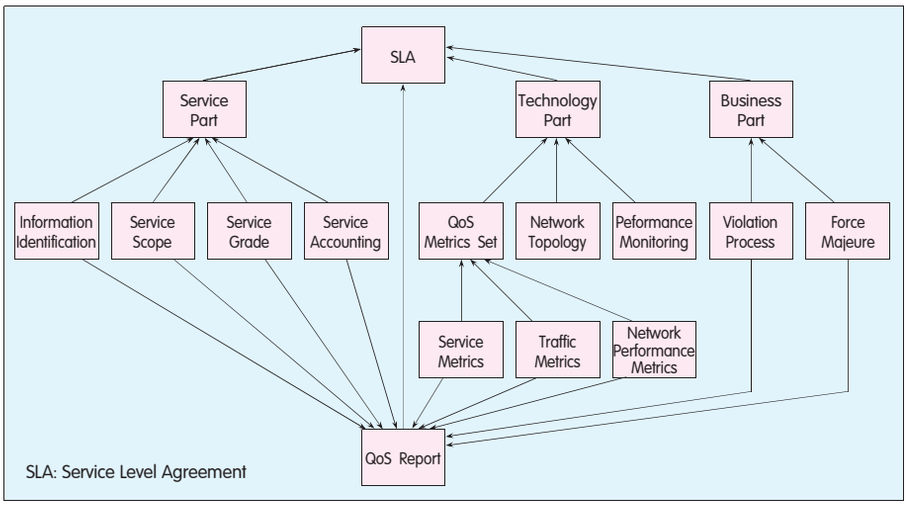
3.1 SLA Representation Template

The SLA representation template is a method used by both SPs and users to define the SLA content of a certain service from the aspects of service level, service quality, priority, and obligations. All aspects are described in this template.

The difficulty in SLA negotiation lies in the shortage of a referential SLA representation template. Without a common template, the SP has to work out a new SLA whenever he negotiates SLA issues with the user. It adds the workload and slows down the exploiture course. Worse, still, as different SPs describe the SLA in their own terminologies, the SLA metrics of the same service may mean distinct differently to the users. It's understandable that the lack of a common representation template will affect the user's satisfaction and no commonly acceptable contracts could be signed.

The SLA representation template uses a template to present SLA contents. It simplifies SLA negotiation process and standardizes the service flow, thus better guarantees service for SPs and users. The SLA representation template defined in this article is a common one that depends on neutral service or implementation technology and thus is quite useful in the early stage of SLA negotiation[7]. We give a sample SLA representation template, shown in Figure 2, to standardize the SLA representation method and contents, which adopts the object–oriented design. Its construction process is the modeling process of SLA representation template, and has 4 parts: service part, technology part, business part, and QoS report part.

• The service part breaks into information identification, service scope, service grade, and service accounting. The information identification describes the basic information of the service, SPs and users. The service scope is the network scope of the service (between service access points or between network access points) defined by the SP when he specifies the service level to the user. According to the scope of service,



▲ Figure 2. SLA representation template.

the proper service parameters are selected, and related network devices, lines, and users are determined. The service grade, as selected by the user, corresponds to relevant service quality and charges. The service accounting is to propose user account bills for the service by certain billing rates and service levels.

• The technology part consists of the QoS metrics, network topology, and performance monitoring. The QoS metrics are the summary of all quality–related metrics, including service metrics, traffic metrics, and network performance metrics. The network topology offers visual network impression for easy understanding of the general network. The performance monitoring is to monitor the service quality data and adjust the service level accordingly.

• The business part includes the violation process and force majeure. The violation process describes the violation conditions and steps to be taken in case of violation. The force majeure statement gives additional description of the "act of God"— exceptional violations that can exempt SPs from penalties.

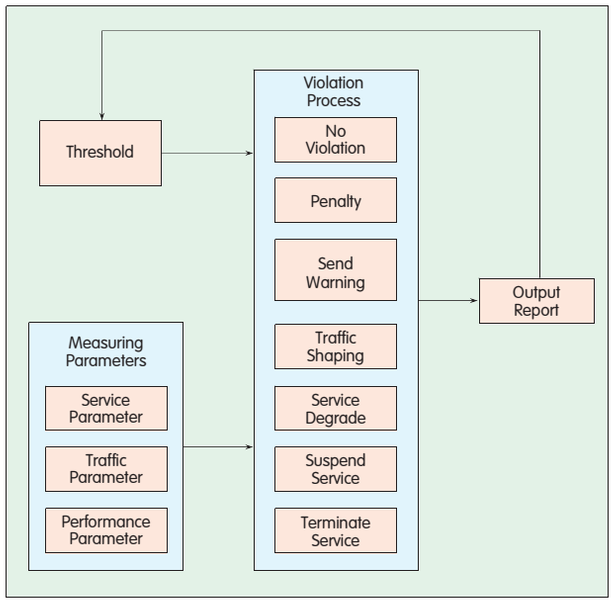
• The QoS report is a classified service quality report offered to the users and SPs. It extracts data from the service part, technology part, and business part and gives the service quality data and statistics as specified in the SLA for further evaluation.

3.2 SLA Violation Process

The SLA violation process sets the content and means of the penalty paid by SP to the user in case the SP fails to provide the expected service level. This violation process is supposed to ensure the fairness, legality and effectiveness of SLA negotiation and is very important in the negotiation process. It also helps to build fair rewards and punishment mechanism and protects user's benefits as well.

The violation process should describe the violation conditions and steps taken in case of violation. When the agreed traffic mode or service quality fails, the SLA is violated.

There are two kinds of SLA violation. One, SP finds out that the network status is not able to meet the SLA terms, in case of network traffic congestion or network disconnection; An SP or a user finds that the service quality is degraded below the level as set in the SLA. The other, a user finds out that the service quality is degraded. Figure 3 shows the violation process flow. Three measuring parameters are used here: the service parameter, traffic parameter, and performance parameter. Whether or not a violation is occurred depends on the parameter values against their thresholds. Steps should then be taken accordingly. Different penalty algorithms should apply to different violations. The violation severity determines whether the warning information should be sent. The parameter's deterioration degree determines the necessity of traffic shaping, and service suspension or termination. The process will ultimately be output in a report (to the threshold module, for instance) as the basis for adjusting the thresholds of some parameters.

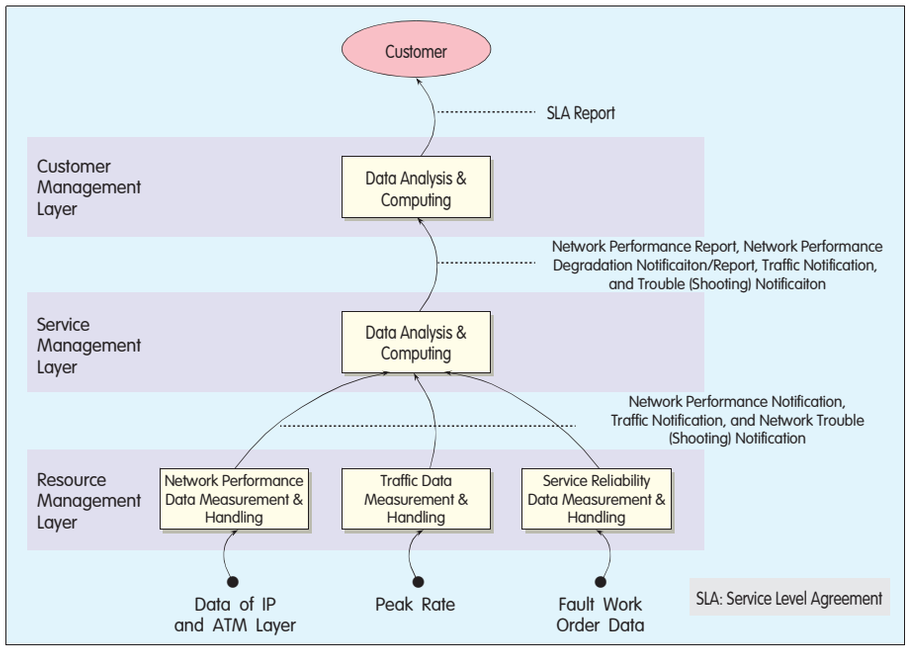


▲ Figure 3. Violation process flow.

3.3 SLA Metrics Evaluation System

As the SLA management requires managing all related network data, service data, and user data, it's necessary to collect mass original data from different layers and processes, compute, and manage them. Figure 4 shows the SLA metrics evaluation system. In case of a single SP, the system is composed of three layers: resource management, service management, and customer management. Data of these layers are collected, analyzed, computed, and transferred among the layers.

• The resource management layer does data measurement, filtering, and computing. End–to–end data, including network performance, traffic and service reliability are generated in this layer to affect the SLA. It sends to the service management layer the network performance notification, traffic



▲ Figure 4. SLA metrics evaluation system.

notification, network fault notification, and network troubleshooting notification.

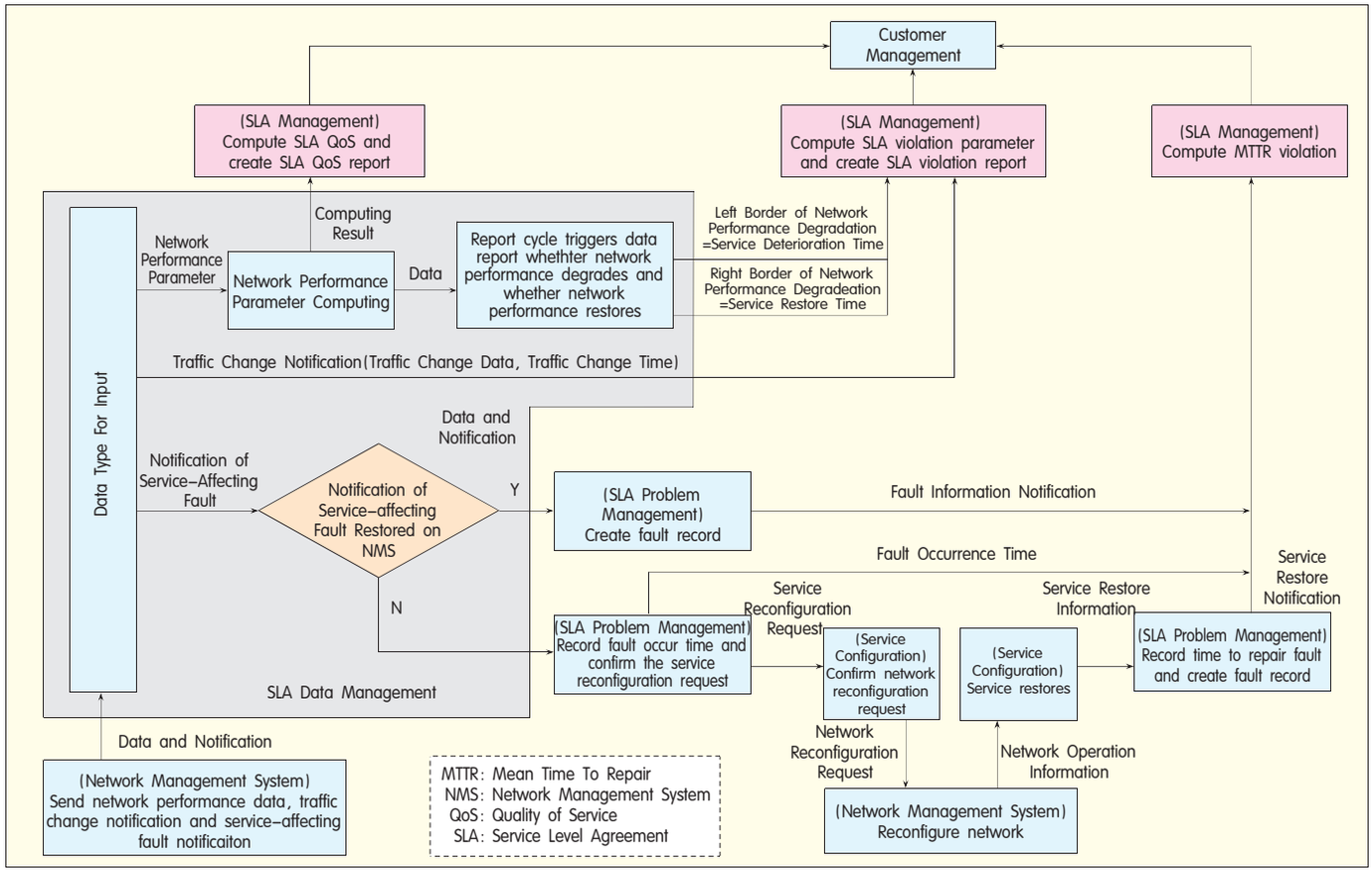
• The service management layer does data analysis and computing. It sends to the customer management layer the network performance report, network performance degradation notification and network performance degradation report, traffic notification, fault notification, and troubleshooting notification.

• The customer management layer does data analysis and computing. It outputs various SLA quality evaluation reports to the SP and user.

4 SLA System Application

With the SLA system application, users can clearly state their requirements and check the SLA implementation. The application helps the SP understand user requirements and use states of the network, thus the SP is able to plan its service quality management development and improve its services and competitiveness[8]. The universal SLA management system functional architecture is composed of three parts: SLA management, SLA problem management, and SLA data management, as shown in Figure 5.

The SLA management module monitors, manages, and outputs reports



▲ Figure 5. Functional architecture of SLA management system.

for the service quality. It compares the quality information against the defined SLA parameters to generate a report whethes there is an SLA violation. This process is concerned with the metrics of a service instance that are related to the SLA, including network performance parameters (delay, jitter), and service performance parameters (service availability, mean time to repair). If the service provided by the SP fails to meet the SLA requirement, billing adjusting may be resulted in.

Functions of SLA management module are:

(1) SLA Evaluation

Manage service quality, ensuring that the service quality meets the agreement signed between the SP and the user. Check the quality–related service data coming from other processes and give alarm to the functional module if the data is not satisfactory.

(2) SLA Violation Management

Ensure to notify the user and related module of the service degradation and violation, and ensure that steps are taken to solve the degradation and violation problems. Analyze the SLA violation information, handle the violation issue and notify the user of the service quality and violation process information when there is an SLA violation.

(3) QoS Report

Report service quality, generate and describe the reports on service level, and customize the quality report and report query based on user demands.

The SLA problem management module reacts promptly to service–affecting faults, and invokes service configuration module or triggers problem solving process.

Functions of the SLA problem management module are:

(1) Problem Diagnosis

Confirm an existing problem that is reported from the resource management layer and notify the user of the problem. Request the resource management layer to check the user complaint to make sure whether the problem exists and feeds back the problem.

(2) Problem Solving

Propose a solution of the problem based on the fault information and performance information, and trigger related modules to carry out the solution scheme.

(3) Problem Closure and Reporting

Perform necessary tests to ensure that the services will be restored to normal quality level. End the problem solving and send fault clearance report to the user.

(4) Fault Information and User Complaint Query

Query fault cause, the current fault that affects the service, the service configuration, and the performance information that related with the current fault and the user complaint information.

The SLA data management module collects and processes information of the network configuration, performance, fault, and billing, and transfers information to relevant processes. It also traces network traffic changes, monitors network faults, estimates network resource use, and sends performance data to the SLA management functional modules and the SLA problem management functional module.

The software part of SLA management system may adapt the 3–layer (data collection, resource management, and user management) structure. It is recommended to use a distributed measurement and centralized management mode. With the distributed measurement, data is obtained regularly for mapping, computing and analysis. The result of data analysis together with the network operation features and SLA requirements will collectively reflect the network status, the service level, and the SLA implementation status.

5 Conclusions

The article has discussed the research work done by international bodies, and the existing problems regarding SLA. It covers the SLA management requirements, relationship between SLA, QoS, and IP network performance. It has discussed three technologies required to implement SLA management (SLA representation template, violation process, and metrics evaluation) and the SLA system application as well.

SLA research is still in its early stage; SLA management is yet to be standardized. With general and special management approaches to be defined in the future, the commonly used industry standard will finally take shape.

References

[1] ITU–T Rec E.800. Terms and Definitions Related to Quality of Service and Network Performance including Dependability [S]. 1994.
[2] TMF701 v2.0, Performance Reporting Concepts and Definitions [S]. 2001.
[3] TMF GB917 v2.0, SLA Management Handbook [S]. 2004.
[4] Salsono S, Ricciato F, Winter M, et al. Definition and Usage of SLSs in the AQUILA Consortium[R]. IETF draft–tequila–sls–00.txt.2000.
[5] Goderis D, T'Joens Y, Jacquenet C, er al. Service Level Specification Semantics and Parameters[R]. IETF draft–tequila–sls–00.txt. 2000.
[6] ITU–T Rec E.860. Framework for Service Level Agreement [S], 2002.
[7] Zhang R Y, Qiu X S, Meng L M. SLA Representation and Applications in the NGI Service Management [A]. Proceeding of the IEEE International Conference on E–Commerce Technology for Dynamic E–Business (CEC–EAST' 2004) [C]. Beijing(China), 2004. Los Alamitos (CA,USA): IEEE Computer Society, 2004: 242–245.
[8] Evans J, Filsfils C. Deploying Diffserv at the Network Edge for Tight SLA, Part 2 [J]. IEEE Internet Computing, 2004, 8 (2): 61–69.

Manuscript received: 2005–11–16

Biographies

Zhang Ruoying, Ph.D, graduated from Beijing University of Posts & Telecommunications. As an engineer of Telecom Operation Research Dept., Telecom Planning Research Institution of MII, she is responsible for telecom operation planning, management, consultation, and network management and communications software research. Over 10 papers have been published and she is one of the first authors drafting the ITU–T M.GDSRT.

Xu Zhifa is Director and a senior engineer of Telecom Operation Research Dept, Telecom Planning Research Institution of MII, with years of experience in telecom operation planning, management, and consultation.

Lu Chuncong is the Vice President of Telecom Planning Research Institution of MII, and Vice Director of Communications Network Operation & Maintenance Committee, China Association of Communication Enterprises (COMC, CACE).