Service Level Agreement

- SLA is a formal contract that describes the service provided and the cost of the service.
- Contract has two parties:
  - Customer
    - End-user
    - Home user
  - Service Provider

ISP Hierarchy

More on this topic in the Peering lecture.

Service Level Agreement

- The function of the SLA is to clearly define the roles, rights and obligations of contract parties at the service interface.
- The contents of the contract depend on the service type and structure offered to the Customer:
  - Network service (Internet Service Provider, ISP)
    - The structure and control of the service is ISP responsibility
    - The structure and control of the service is Customers responsibility
  - Application service (Application Service Provider, ASP)
    - Web, email etc.
  - The contract should contain only those services/applications over which the ASP/ISP has direct control.
Service Level Agreement

- Contract defines
  - Service(s) offered
  - Cost of the service offered
  - Methods to control the service level
  - Procedures to follow in case of network malfunction

- The definition of a service should contain all areas of the service and quality parameters
  - Network service
  - Application and other value-added services
  - Support services

Network services — private access

- Private access
  - Private access service is based on the control of the Customer's access line (and control of the attached closed service network)
    - Connecting the Customer's sites with a closed network (Virtual Private Network)
  - Access lines can be
    - Fixed line connections
    - xDSL, DOCSIS or Ethernet technologies for home users
    - Any asymmetrical technology for enterprise users
    - Wireless connections
    - 3G, @450, satellite
  - ISPs may offer differentiated access services based on the Customer's access line
    - Capacity
    - Delay
    - Packet loss
    - Guaranteeing of these quality/performance parameters, independent of time or load level, is called Quality of Service

Network services — public access

- Public access
  - Public access service is based on Customer/transit contracts made by the ISP with other ISPs
    - If a customer uses this service he/she/it is offered
      - Global IP-addresses (no NAT, no private IP-addresses)
      - Access point from where traffic is routed onwards.
    - ISP controls the performance and service level of the access.

Network services — transit

- Transit service
  - Enable the networking of ISPs
  - Small ISPs are customers to larger ISPs
    - Large ISPs forward the traffic as their customer traffic
  - Equal size ISPs work together as partners
    - Each are other’s customers
    - Mutual contracts contain mainly restrictions regarding the forwarded traffic.

Value-added services

- All supportive services that support network traffic are referred to as value added (ed) services
  - Virtual network services (L2, L2+, L3 VPN)
  - Operating application services (email, web hosting...)
  - Security services (Firewall, IDS etc.)
  - Managing name services and their integrity
  - Controlling and managing customer’s configuration and/or equipment (Managed LAN service)

Performance and quality metrics: examples

- Examples of measured performance and quality parameters:
  - Network capacity is a bps and its usability is y % over time z.
  - The web page usability is a % measured over time b. The download capacity is c bps and the response time is less than d ms. The complete data is backed up every f days /weeks /months.
  - Customer VPN is offered a capacity of m bps with maximum end-to-end delay of n ms.
Usability

> The service availability (or usability) indicates the time that the services is available (or usable)

> Common terms and definitions with availability/usability are:

   - MTBF - Mean Time Between Failures
   - MTBI - Mean Time Between Interruptions
   - MTTR - Mean Time To Restoration

   > These terms define the availability slightly better than just plain percentage values.
   > Not a vague definition “over time t...”
   > Not just one long break in the service.

<table>
<thead>
<tr>
<th>Usability Level</th>
<th>MTBF (h)</th>
<th>MTBI (h)</th>
<th>MTTR (h)</th>
</tr>
</thead>
<tbody>
<tr>
<td>90%</td>
<td>0.9</td>
<td>0.2</td>
<td>0.1</td>
</tr>
<tr>
<td>95%</td>
<td>0.5</td>
<td>0.1</td>
<td>0.05</td>
</tr>
<tr>
<td>98%</td>
<td>0.09</td>
<td>0.02</td>
<td>0.01</td>
</tr>
<tr>
<td>99%</td>
<td>0.05</td>
<td>0.01</td>
<td>0.005</td>
</tr>
<tr>
<td>99.5%</td>
<td>0.03</td>
<td>0.005</td>
<td>0.003</td>
</tr>
<tr>
<td>99.9%</td>
<td>0.01</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>99.99%</td>
<td>0.001</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td>99.999%</td>
<td>0.0001</td>
<td>0.00001</td>
<td>0.00001</td>
</tr>
</tbody>
</table>

MTBF - ITU-T definition

> Item

   - “Any part, device, subsystem, functional unit, equipment or system that can be individually considered.”

> Failure

   - “The termination of the ability of an item to perform a required function.”

> Time Between Failures

   - “The time duration between two successive failures of a repaired item.”

> Mean Time Between Failures

   - “The expectation of the time between failures.”

MTBI - ITU-T definition

> Interruption

   - “Temporary inability of a service to be provided persisting for more than a given time duration, characterized by a change beyond given limits in at least one parameter essential for the service.”

> Time between interruptions

   - “The time duration between the end of one interruption and the beginning of the next.”

> Mean Time Between Interruptions

   - “The expectation of the time between interruptions.”

MTTR - ITU-T definition

> Mean Time To Restoration (Recovery, Repair)

> Time To Restoration

   - “The time interval during which an item is in a down state due to a failure.”

> MTTR

   - “The expectation of the time to restoration.”

Service Level Agreement

> The level of SLA parameters defines the cost of offering the service (for the ISP)

   - The higher the bandwidth or delay demands, the more expensive the service is

   - This applies also to availability/usability

Example:

   > 99.99% usability over one week period means a service that is unusable for less than a minute (during that same week).
   > Makes it impossible to maintain equipment without redundancy
   > Redundant equipment and access to network
   > Raises the costs for the ISP
   > 99 % usability enables the maintenance of network equipment without violating the contract (1.5 hours per week)

Service Level Agreement

> It is in the best interest of the customer to control that the contract (=SLA) is fulfilled

   - Measure the parameters mentioned in the SLA with the best applicable practice (that should also be documented in the SLA)

   - Service is measured where that service is used

      - Customer premises, workstations, access points
      - From the ISP provided measurement device located in the customer premises
      - From the transit point of the ISP towards the ISP’s value added service point
Measuring service quality

> When defining quality measure, always separate fact from fiction: measure only that what is quantitative
  - The measured value should be defined in the SLA so that it can be measured by both the SP and the Customer
  - Do not attach any qualitative or otherwise inexact values that are based on human interpretation or experience
  - The human effect is dependent of the person and his/her feelings
  - The end-user can’t be controlled by the Service Provider
  - Vague definitions lead to controversy that most probably leads to a sensation of unsatisfying (total) quality of service
  - “Collective effect of service performances which determine the degree of satisfaction of a user of the service”[ITU-T E.800]

Verifying the SLA of a network service

- Typically measured between the ISP’s transit points
- Service is usually guaranteed up to these points but not beyond

Verifying the SLA of an application service

- The ASP has measurement points within its network where usability is measured from
- Usually placed to transit points (at least)

Verifying the SLA of a VPN service

- From the ends of the VPN-tunnels or measurement points within the customer network
- Each site-to-site VPN-connection is measured independently from others

Measuring service level

> Quantitative service level
  - Describes the service performance on different protocol levels
  - Can be defined unambiguously
  - Service logic illustrated as a flow chart
  - Numerical values of performance can be attached
  - Can be measured

> Qualitative service level (based on customer experience)
  - Illustrates the service level / service response experienced by the end-user (Customer)
  - The response depends on factors that cannot be measured
  - Ambiguous definition
  - Can not be measured

SLA-measurements

> Measured performance/quality metrics include
  - Delay, one-way or two-way
  - Delay variation (jitter)
  - Packet loss
  - Reachability
  - Availability
  - Response times (e.g. Web-server first page Download time, TCP-connection time)
  - Throughput, available bandwidth
RFC 2544

- Originally defined to offer a standardized set of tests for measuring network device performance (latency, frame loss etc.)
  - Test framework can be also used to measure a network service’s SLA conformance
    - Commissioning measurement for VPNs or Ethernet services
  - RFC-2544 outlines specific tests to validate throughput, latency, frame loss, and back-to-back (burst) performance
  - Full line-rate traffic generation with small and big frames

Measurement methods

- End-to-end measurement
  - Probes are sent from point A to point B
    - E.g. UDP or ICMP packets
    - Packets are timestamped on departure and arrival
    - Packets are given sequence numbers
    - Network’s response is measured by observing the timestamps (delay, jitter) and sequence numbers (packet loss) on the packets
  - Sometimes end-to-end not possible
    - Measured path may cross several independent domains with conflicting policies, measurement tools, methods etc.
    - Session Border Controllers break the end-to-end connection of a VoIP call
    - RTCP statistics are not end-to-end but from the caller to the SBC

Loss of end-to-end connection

- SLA-measurement using an emulated VoIP call: MOS, RTT, DWD, Ploss statistics
  - SBC performs transcoding (e.g. From G.711 to G.729)
- The call initiated by Host A is terminated in the gateway and then in turn SBC performs transcoding (e.g. From G.711 to G.729)
- SLA-measurement using an emulated VoIP call: MOS, RTT, OWD, Ploss statistics

Spatial composition

- Based on the idea that measurements of the sub-paths can be combined so that the result estimates the properties of the complete path
- Can be used to get an estimate of the properties of a inter-domain path without a separate end-to-end measurement
  - Each domain measures its edge-to-edge sub-path

Measurement considerations

- Bandwidth usage of SLA-measurements
  - Active measurements cause extra traffic and thus disturb the normal traffic in the measured network
  - Intrusiveness
  - Probing traffic should only use a few percent of the measured network path’s available bandwidth
  - RFC 2544 Commissioning measurement uses up all available BW
  - Reporting measurement data to DB takes up BW as well
- Measurement accuracy
  - Typically only expensive devices reach <1ms accuracy (delay measurement)
  - Operators have accurate measurement devices in their core networks
- The service should be measured on the layer it is offered
  - E.g. a L2-service cannot be measured with TCP-throughput test...
SLA measurement tools

- Tools are based on operating system dependent measurement software or separate measurement devices
  - Measurement Agents (software)
  - Measurement Probes (hardware)
- A measurement device located on the Customer premises
  - Run the tests described in the SLA
  - Transfers the test results to the Service Provider’s centralized database
- SP’s database reports the results to the Customer according to the SLA

Example: Brix Measurement platform

- Several types of tests
  - UDP Echo (delay, reachability)
  - VoIP emulation (MOS)
  - Cisco CallManager
  - HTTP, DNS, SMTP, IMAP
  - DHCP, FTP...
- Probes
  - Brix 100 Verifier
  - Brix 1000
  - Software Agent for PCs
- Used by the biggest operators in Finland

Measurement probes

- Accedian EtherNID
  - Can be installed in-line or one-armed
  - Demarcation device
  - Layer 2 and 3 tests
  - In-service RFC 2544 test
  - Splitter functionality
  - 1 Gbps interfaces with changeable SFPs
- Brix 100M Verifier
  - One-armed installation
  - 100 Mbps interface
  - Layer 3-7 tests

SLA-reporting

- Data from Brix
- One-way delay between two measurement points
- SLA-reporting view
- Webserver performance
- Many tools have a "traffic light"-type of indicator to illustrate the state of the SLA

Summary

- SLA is a contract, not a set of rules
  - It should not be seen as a threat
- The function of SLA is to create a structured view of what the Customer needs and what the SP has to offer and the relation between these two
- Works as a referee in disputes
- Thus, it must be clearly measurable and observable by both the Customer and the SP
- SLA measurement methods must be agreed upon so that they can be used to decide if the contract has been fulfilled or not