Service Level Agreements and QoS Delivery in Mission Oriented Networks

By

Bharat Doshi, University of Massachussets
Paul Kim, Applied Physics Lab, Johns Hopkins University
Burt Liebowitz, The MITRE Corporation
Kun I. Park, The MITRE Corporation
Sherry Wang, Applied Physics Lab, Johns Hopkins University

1 Introduction

The SLA is an essential tool for delivering mission critical networking services over a Global Information Grid (GIG) network infrastructure that will be designed and operated by many different entities. The GIG integrates Network Service Domains (NSD) created by many acquisition programs. Each NSD will be operated by a NSP. Some NSPs will operate more than one NSD.

Some User Communities will be satisfied with best effort services across multiple domains. Others will require some level of guaranteed capacity and quality of service (QoS) in order to meet their mission goals. Service Level Agreements (SLAs) play major roles in achieving this objective. In particular, SLAs enable User Communities to specify their traffic and QoS requirements and enable NSPs to plan and promise service levels across the NSDs they control.

To achieve these goals, it is important to establish a set of GIG-wide standards for creating SLAs between User Communities and one or more NSPs. In particular, it is important to define a common set of metrics and contents for creating SLAs. Standardized templates, guidelines for allocation of metrics among NSDs, and standardized Service Offerings by NSPs, will help structure SLAs in ways that allow end-to-end requirements to be met.

Many User Communities will need service from more than one NSP. These User Communities will benefit from an organizational structure and a process to create a single SLA that incorporates the services of multiple NSPs. This document discusses a notional cooperative process for this purpose.

This paper provides recommendations for implementing SLAs, creating SLA processes and roles, and selecting SLA metrics to be used by Global Information Grid (GIG) User Communities and Network Service Providers. SLAs define the performance metrics, obligations and financial relationships between User Communities and Network Service Providers (NSP), and between cooperating NSPs. Performance metrics include throughput, packet delay, and packet loss. Obligations include reporting requirements, customer support requirements, incentives and methods for resolving disputes. Financial relationships define the cost for the services covered in the SLA.

2 Driving Requirements, Tenets, and Assumptions

2.1 Basic SLA Approach

The GIG network infrastructure consists of multiple Network Service Domains (NSDs), each managed by a Network Service Provider (NSP). A notional view of the infrastructure is shown in Figure 4-1. The GIG network infrastructure serves a set of User Communities. A User Community is a group of users that support a specific mission. A User Community develops and manages one or more edge networks

which work together to support the mission. In addition, individual users within a User Community could reach out to edge networks and supporting hosts outside the control of the User Community.

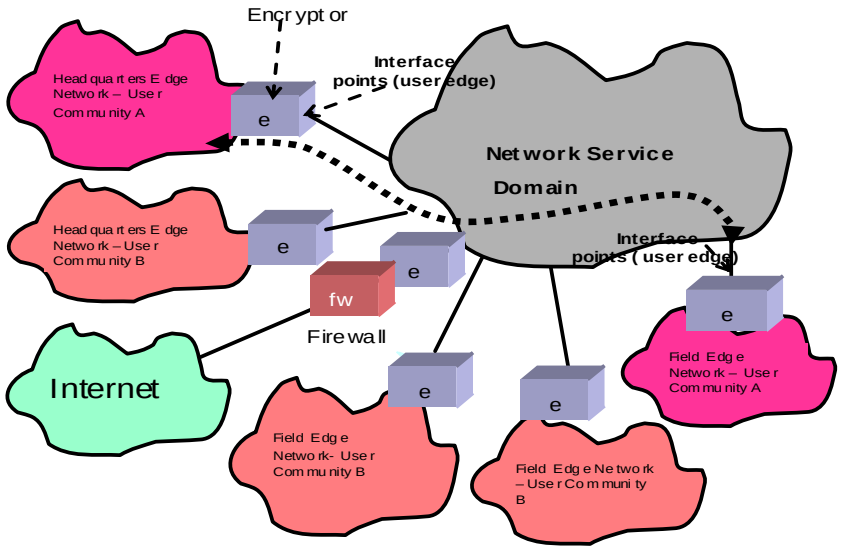

Figure 2-1– User Community/NSP Relationships

A User Community might need transport service from one or more GIG NSPs. This service encompasses connectivity between specific edge points[1], and a level of assured performance and Quality of Service (QoS), typically stated in terms of minimum throughput packet loss, packet delay, and transport availability.

To obtain this support, the User Community shall negotiate a Mission Service Level Agreement (MSLA) specifying the connectivity for multiple services and QoS needed for the mission. The MSLA shall also define reporting relationships and legal and financial obligations between the User Community and the NSPs.

The User Community shall determine its workload and end-to-end performance requirements, including response time, packet delay, and reliability metrics for key transactions, file transfers and real time streams for each service. The development of these requirements shall be under the control of a Mission Planner (MP) associated with the User Community. The Mission Planner shall use the end-to-end performance requirements to derive a set of network QoS metrics to be included in the MSLA. The specific nature of the metrics and their range of values depend upon the applications that support the mission. Example metrics and supporting SLAs will be presented later in this document.

The Mission Planner and a designated representative from the NSP shall negotiate the content of the MSLA. The NSP shall define a set of standard services and terms and conditions as a starting point for the negotiation. The MP and the NSP shall resolve any differences between the User Community requirements and the standard services. The results of this resolution process shall be documented in the MSLA.

Figure 4-1 depicts a situation where a group of independent User Communities require service from one NSD. The case where a User Community needs the services of multiple domains will be discussed in Section 4-2.

[1] An edge point is a location at a User Community network that connects to a location at a Network Service Domain.

2.1.1 User Community

A User Community comprises a set of users who support a specific mission. User Communities vary in size from dozens to millions of individual users. An individual user is a person or automated device that has access to a computer or router which can support IP packet flows.

User terminals are typically connected to local area networks (LAN) that exist within one of the User Community's edge networks. The LAN typically has access to one or more communications lines that are part of a wide area network (WAN). In some cases an edge network might support hundreds of user terminals, in other cases an edge network might contain only one user terminal.

User Communities support a wide range of missions. Some are tactical and involve mobile vehicles. Others are strategic and are based on fixed enclaves. Some use combinations of satellite, wireless and terrestrial links. Different missions support different applications. Some applications are based on TCP which can support reliable communications by automatically retransmitting lost packets. Other applications are based on real time UDP packet streams that require small packet delay and jitter.

The designation of what comprises a User Community shall be left to the various agencies that will connect to the GIG. The connections shall be made at locations agreed upon between the User Community and the GIG NSPs. The connection between a User Community and a Network Service Domain is called the "user edge".

2.1.2 Mission Planner

Each User Community shall create an organization or designate a person to perform the role of Mission Planner. The Mission Planner shall determine the connectivity requirements based on analysis of the user workload generated in support of the mission. This workload consists of but is not limited to file transfers, voice and video calls, email and web transactions.

The Mission Planner shall analyze the frequency and traffic volume associated with the workload. The Mission Planner shall also determine the application response times, delay times, throughput and availability and other transmission metrics necessary to support the mission. These metrics constitute the user's end-to-end requirements for a service.

The Mission Planner shall aggregate the user end-to-end requirements to determine user edge-to-edge[2] traffic flows and categorize these flows by Precedence Level and Quality of Service. The Mission Planner shall determine the communications paths, edge locations and Quality of Service parameters needed to connect the User Community enclaves. The Mission Planner shall use this information to negotiate SLAs with the NSPs whose domains will be used to connect the User Community's enclaves.

2.1.3 Network Service Provider

A Network Service Provider is an entity that develops and maintains an IP-based Network Service Domain. The NSD contains interior and edge routers. The latter are used to connect to other NSDs and to User Community edge networks. The key feature that distinguishes a Network Service Provider from a User Community is that the NSP provides a transit service for other NSPs and User Communities.

Some GIG user organizations will develop and support networks that could be dual purpose. These networks will function as an edge network for a User Community and will also serve as a transit network for NSPs and User Communities. If so, the organization shall be considered a Network Service Provider when serving the other User Communities and NSPs.

An NSP shall, unless otherwise indicated, provide the following features for its client base:

- Points of presence at which the client can interface with a domain's edger router
- A standardized interface at the point of presence
- A set of transport services including:

[2] User Edge to Edge is defined as the connectivity between User Community enclaves, as shown in Figure 4-1.

    - Point to point connectivity
    - Multicast connectivity
    - Partial mesh connectivity
- A set of assured services that, subject to precedence requirements, guaranty performance metrics along the paths used to transmit client traffic flows
- A designated point of contact for User Community Mission Planners
- A set of templates that can be used for negotiating SLAs

The GIG NSPs shall collectively create a set of standardized transport services for the GIG, which are defined and documented as GIG Service Offerings (SO). It is not mandatory that all GIG NSPs support all GIG SOs. Nor is it necessary that NSPs provide only the services defined in the SOs. But if a NSP provides a service that is defined in an SO, the NSP shall provide that service in conformance with the attributes defined in the SO.

2.2 SLA Cooperative Approach for Multiple Service Providers

If the GIG Network Infrastructure was centrally owned, the creation of a MSLA would be relatively straight forward. A User Community would negotiate a single MSLA with a GIG organization designated to engage in such negotiations. This organization would allocate requirements across the various networks controlled by GIG management. However, the GIG is not centrally owned and managed and such an organization does not currently exist. Therefore some User Communities will need transit services from more than one Network Service Domain, as shown in Figure 4-2.

The commercial Internet provides a model for service provider cooperation in supporting user flows. But this model was designed to for a best-effort, connectivity-oriented service. The GIG must support more stringent goals, especially with regard to QoS. Therefore, it is of great importance that the GIG NSPs cooperate to provide a standardized set of Service Offerings.

If the GIG NSPs do not cooperate to provide standardized Service Offerings, User Communities, in many cases, will be forced to negotiate iteratively with multiple NSPs, a potentially expensive and time-consuming process.

The concept of cooperative inter-provider networks has been discussed to some degree in the literature. Reference 3, for example, describes a cooperative approach, where NSPs agree to coordinate their service offerings and establish standards for conveying QoS requirements across multiple domains. However, there is little experience in the commercial IP world from which to create a model for a cooperative approach in the GIG. The collection of entities that comprise the GIG will have to create their own cooperative mechanisms to provide cross-domain QoS. Some of the approaches used by commercial telecommunications providers may be useful in this endeavor.

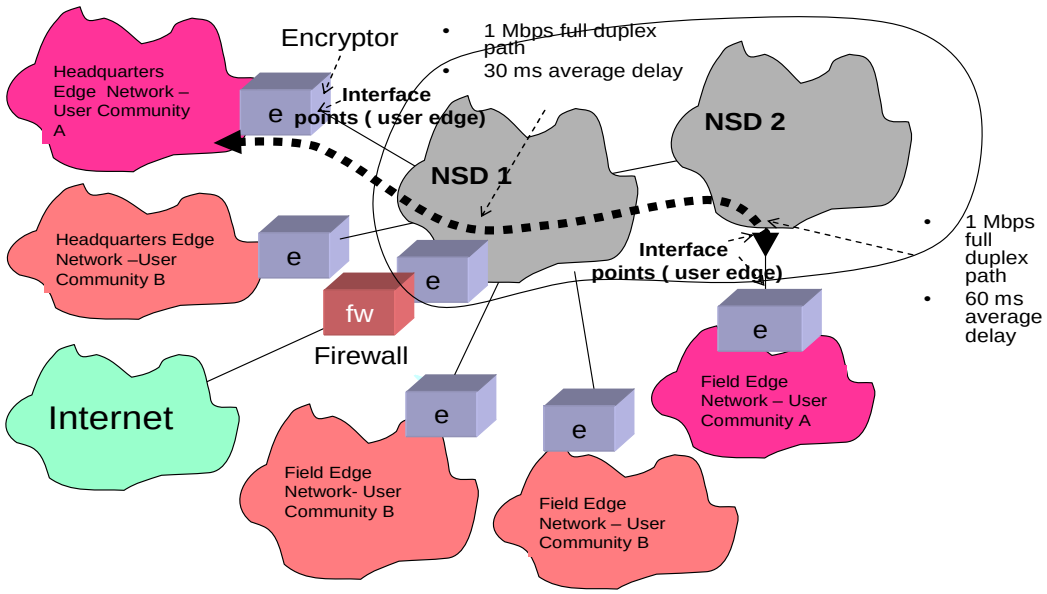

Figure 2-2 Example of Traffic Flow and QoS Metrics in a Multi-provider Network

2.2.1 Description of Cooperative Approach

The Cooperative Approach recommended for the GIG is based on the relationship between User Communities (and their Mission Planners), an entity called the GIG Cooperative Agency (GCA) and the Network Service Providers. The GCA is an interface organization that represents the cooperative NSPs, so that a User Community does not have to negotiate SLAs with multiple NSPs. The Cooperative Approach shall incorporate the following elements:

1. The User Communities (UC) and their Mission Planners
2. The Network Service Providers (NSP)
3. The GIG Cooperative Agency (GCA) which shall include representation from each NSP and possibly a third party. The GCA serves as a collective agent of the NSPs
4. A set of standard Service Offerings (SO) which shall define the types of services provided by the NSPs.
5. A set of standard MSLA templates shall be used as the basis for negotiating mission specific SLAs
6. A process for negotiating specific MSLAs between a Mission Planner and a GCA based on UC requirements
7. A set of standard Network Service Domain SLA (NSD SLA) templates which shall be used by the GCA and individual NSPs as the basis for negotiating specific SLAs. The NSD SLAs shall define the aggregate requirements for a domain and are derived from the MSLAs that require services from the domain.
8. A set of standard Service Domain SLA templates which shall be used as the basis for creating specific Service Domain SLAs. These SLAs shall define data flows and QoS needs between the NSPs for those cases where one NSP is the direct user of another NSP's services

9. A reporting system which shall enable UCs to report service related issues to the GCA and enable the GCA to report issues to NSPs
10. A government-established Adjudication Function (AF) which shall resolve issues related to perceived failures to meet SLA requirements
11. A methodology that shall be used for developing, pricing and advertising Service Offerings

The relationship between these elements is shown in Figure 4-3. The User Community and its Mission Planner shall determine the end-to-end network performance requirements for a specific mission. The MP shall derive user edge-to-edge requirements based on the performance requirements. The MP shall negotiate MSLAs with the GIG Cooperative Agency. The NSPs shall establish and operate GIG transit domains. NSPs shall negotiate NSD SLAs with GCA. The Adjudication Function resolves MSLA and NSD SLA disputes[3].

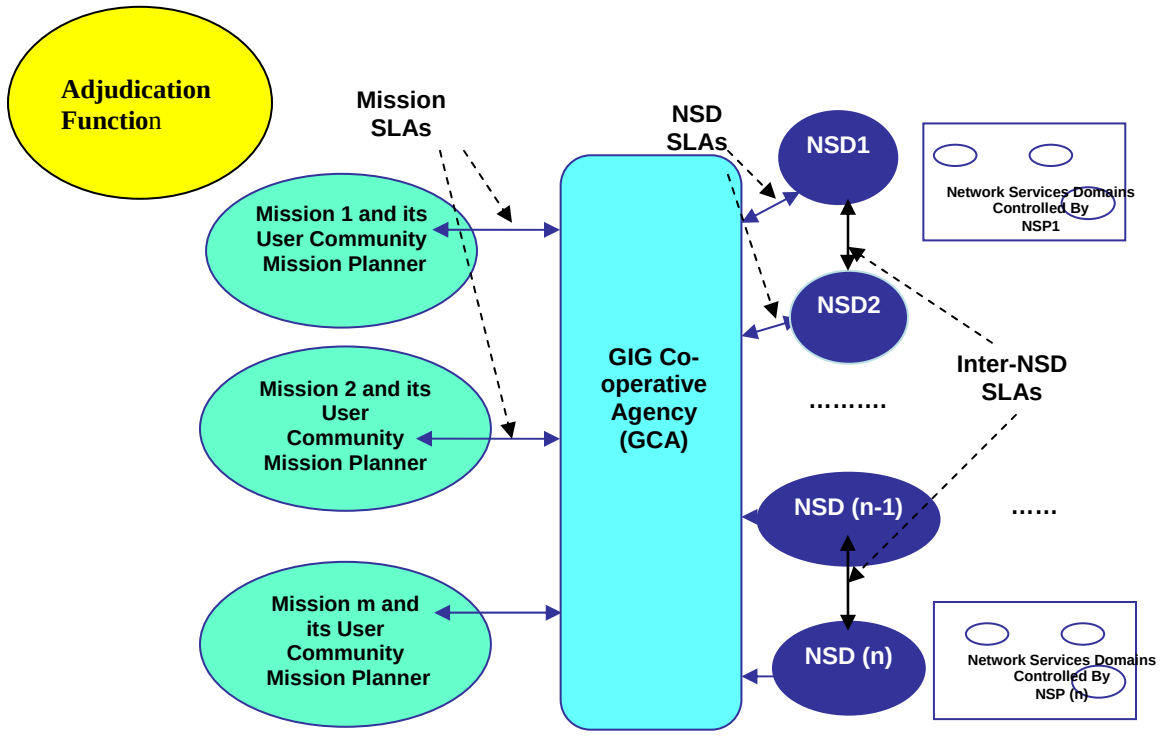

Figure 2-3 – Relationship between User Communities and GCA

2.2.2 The Role of the GIG Cooperative Agency

The government shall establish an entity called the GIG Cooperative Agency. The GCA has multiple functions and roles, including:

- SLA Negotiation The GCA shall negotiate the MSLA with a User Community's MP. The MP determines the end-to-end requirements for a specific UC mission and from these derives the user

[3] For example, a UC might observe that it is not receiving the bandwidth negotiated in a MSLA; a NSP might observe that a UC has consistently transmitted a volume of traffic greater than agreed to in the MSLA.

edge-to-edge requirements. The MP shall present the user edge-to-edge requirements to the GCA. The GCA shall determine if the user edge-to-edge requirements can be met by the Service Offerings and connectivity available from the Cooperative NSPs. If the requirements can be met at a satisfactory condition, the GCA shall create the MSLA, which shall be reviewed and agreed to by the MP. The GCA shall then negotiate or modify existing Network Service Domain SLAs with the supporting NSPs. If the UC requirements cannot be satisfied by the standard offerings, the MP and GCA shall attempt to negotiate a mutually agreeable MSLA. A successful negotiation might require that the UC back off from its original demands or that one or more NSPs provide a non-standard service.

- Support the Establishment and Maintenance of Standards The GCA shall work with the NSPs to develop, publish, disseminate and update technical standards and SOs.
- Technical Consulting The GCA shall, at the request of User Communities, provide technical assistance to MPs regarding the specification of user edge-to-edge requirements. This advice shall be part of a systems engineering service in which the GCA could also analyze UC workload, develop user end-to-end requirements and from these help derive user edge-to-edge requirements.
- Support Negotiation of Inter-provider SLAs The GCA shall, at the request of the NSPs, assist in the development of Service Domain SLAs.
- Creation of SLA Templates The GCA, working with representatives from the NSPs, shall create templates for MSLAs, NSD SLAs and inter-provider SLAs.
- Disseminate Service Information The GCA shall hold periodic meetings with NSP staff to discuss and initiate new service offerings. The GCA will also hold periodic meetings with UC representatives to present new Service Offerings and to solicit inputs regarding the need for new services.
- Develop and Maintain a Problem Reporting System The GCA shall develop a Reporting System that shall be used by UCs and NSPs to report outages and service degradation.
- Develop and Maintain An Automated Provisioning System The GCA shall develop a provisioning system that will keep track of SLA service commitments and NSP resources. The GCA shall use the Provisioning System to validate that the NSPs can support a new MSLA, before signing off on the MSLA. The Cooperative NSPs shall use the Provisioning System to review the GCA validation to ensure that they indeed have the resources to satisfy the MSLA. The GCA shall maintain the Provisioning System based on SLA inputs and network infrastructure updates from the NSPs.
- Publish Periodic Projections Of Traffic And Service Demands The GCA shall collect information from the User Communities and other sources to create traffic projections that shall be used by the Cooperating NSPs to plan network upgrades.

2.3 SLA Contents and Metrics

This section describes the recommended SLA metrics and contents of typical SLAs. We also describe different types of SLAs to be used to deliver network services to the missions while allowing distributed executions by NSPs. In addition, it describes sample SLA metrics values and methods for allocating performance metrics among NSDs.

2.3.1 Typical SLA Content

In the commercial services, a typical SLA may contain the following sections.

1. Description of Service
    a. Nature of traffic/Identification of Service Offering
    b. Required end points (see below)
2. Expected Performance Levels (The Service Level Specification)

3. Reporting Procedures for error conditions
4. Time-Frame for Problem Resolution
5. Process for Monitoring and Reporting
6. Process and time constraints for provisioning changes to the service
7. Consequences of Service Provider or User Community Not Meeting Obligations
8. Constraints
    a. Impact of higher precedence traffic
    b. Impact of network degradation due to enemy action
    c. Acts of God
    d. Short term and long term guarantees

2.3.1.1 Mission SLA

The main objective of the Mission SLA (MSLA) is to provide an end-to-end service agreement between the Mission Planner representing a User Community for a particular mission and the GCA In order to support a particular mission SLA, the GCA might negotiate multiple NSD SLAs with one or more Network Service Providers.

The SLS portion within the Mission SLA shall include the following information:

1. Mission ID
2. Mission Name
3. Mission Start/Stop Date/Time
4. Mission Location & size of mission area
5. Terminal Locations (Exact location information could be classified)
6. Ingress and Egress points
7. Services required for each terminal type (e.g., Voice, Data, Video)
8. End-to-End SLA metrics associated with a corresponding mission
    a. Committed Information Rate (CIR) for each service
    b. Peak Information Rate (PIR) for each service
    c. User Edge-to-Edge Delay
    d. User Edge-to-Edge Jitter
    e. Availability
    f. VOIP metrics such as blocking probability, average time to set up call

The Ingress and Egress points (item 6, above) define the end points for the traffic flow. There could be any number of such points, depending upon the nature of the services defined in the SLA. which could encompass the following different types of flows:

1. One-way point to point
2. Full-duplex point to point
3. Multicast point to multipoint
4. Mesh configurations
    a. Point to multipoint
    b. Multipoint to point
    c. Multipoint to multipoint
5. Full connectivity to all points to which the User Community is entitled. This connectivity could be used for
    a. Database access throughout the GIG
    b. Internet Access

It is possible that SLA metrics could differ for each (one-way) path of a multipath flow. In such case the SLA must provide metrics for each of the paths.

2.3.2 Single Network Service Domain SLA Metrics

This section will focus on the edge-to-edge SLA metrics in a single network service domain as shown in Fig. 4.1. The following discussion is based on a service that consists of a single one-way virtual path through a domain. The SLA will contain metrics for each direction of each path required to support the User Community.

Some general network and traffic conditions are assumed for scoping the discussion:

- The SLA does not cover the impact of delays within the edge networks operated by the User Community
- Traffic influxes in each domain are regulated and traffic bounds are satisfied.
- Each domain has sufficient network capacity for its traffic influxes
- Network components and devices in each domain are well maintained in their service life.

SLA metrics can be categorized in terms of capacity, QoS and availability. Capacity, sometimes referred to as throughput, defines the traffic rate (bits per second, or bytes per month) that must be accommodated at the Ingress to the NSD. QoS defines the packet loss, packet delay, packet jitter limits for the traffic flow, provided that the User Community traffic rate at the Ingress is less than the agreed upon throughput. Availability refers to the probability that the capacity will be available when needed.

Capacity can be specified by one or more of the following metrics, which are defined in more detail in Appendix A:

1. **Committed Information Rate (CIR)** – this is the traffic rate (in bits per second) which the NSP will accommodate either at all times, or for a specified period of time. In some services, the NSP will reject traffic in excess of this rate, measured over some agreed upon period of time. This time period, for example, could be 1 second, one minute, five minutes. In other services, the NSP could accept excess traffic, and mark it as such. The packets marked as excess could be dropped by downstream routers if they experience congestion at their output ports.
2. **Peak Information Rate (PIR)** – this measure could be used by itself or in conjunction with CIR, to indicate a traffic rate above the CIR which can be generated by the User Community for some agreed upon burst period or for some percent of the time the User Community is entitled to the CIR.
3. **Committed Burst Size (CBS)** – this measure is used in conjunction with CIR or PIR to indicated the largest burst size in bytes that can be accommodated by the NSD, without losing any packets
4. **Allowable Traffic per Month** – this measure is used to limit the amount of traffic the User Community can send per month, Traffic above this amount can be rejected by the NSP or could incur an extra charge.

Quality of Service can be defined by one or more of the following metrics which are defined in Section 8.2 of this document:

1. **Average packet delay**, which is also called IP Packet Transfer Delay (IPTD), measured in seconds.
2. **Jitter**, which is sometimes called IP Packet Delay Variation (IPDV). Jitter is a measure of the variation in delay measured over some period of time.
3. **Packet loss**, which is sometimes called IP Packet Loss Ratio (IPLR). This is the average number of packets loss per thousand packets sent, based on congestion in the NSD.

Availability can be defined by one or more of the following metrics:

1. ***Service Availability****, which is the proportion of time* the service is fully available to the User Community. This availability includes the availability of the NSD, but might or might not include the availability of the connection from the User Community to the NSD point of presence. This metric indicates the inherent reliability of the components within the NSD. The SLA could support a range of Service Availabilities, depending upon the degree of failure. For example full ca-

pability Service Availability could be specified at .99. Partial service (for example at least one half of the specified capacity) could be specified at .995.

2. ***Session Availability***, which is the probability that there is sufficient capacity in the NSD to support a session, at the time that the User Community requests resources for the session. For VOIP, Session Availability is one minus the probability that a call is blocked for lack of resources. Session Availability is often linked to other metrics that indicate session setup time. The SLA could also include a metric for the average time it takes to start a session, based on the time the request for the session was initiated. If this metric, as measured over some number of sessions, is exceeded the SLA could define a penalty for the NSP. Another metric that could be used is 95th percentile of delay is setting up a session. If session setup takes longer than the metric, the session will be considered to have failed.
3. ***Session Reliability***, which is the probability that the service will be available for the time duration that it is needed, given that is was available at the time the session commenced.
4. ***Mean Time to Repair (MTTR)***, which is time it takes the NSP to repair a failure within the NSD, measured from the time the failure is reported by the User Community
5. ***Mean TimeBetween Failure (MTBF)***, which is the average time between NSD failures.

Examples of typical military application service requirements are given in Appendix B – Service Class Definitions. Typical bounds are discussed in the following list.

1. Throughput Metrics

a. CIRs can be assigned with a range of values starting at 0 bits per second up to the physical bounds of the links within a NSD
b. Allowable traffic per month has no particular bound and will be driven by the NSP Service Offerings – typical amounts are in the megabyte per month or gigabyte per month range
c. Committed Burst Size – could vary from zero to 10s of thousands of bytes

2. QoS Measures

a. IP Packet Transfer Delay (IPTD) (one-way)
b. IP Packet Delay Variation (IPDV or Jitter)
c. IP Packet Loss Ratio (IPLR)

Packet loss depends on the type of transmission network. Fiber optic networks will have much less loss due to bit errors than would MANET networks. However, bit error is only one contributor to packet loss. Congestion is the other major contributor. Typical packet loss for fiber networks are in the order of 1 per thousand. Certain services, such as VoIP could be sold with a loss of one in a hundred, when all circuits are busy. Packet loss could change with time in a wireless domain based on weather, jamming, etc. If so, the mission requirements should provide requirements for long term and short term ranges. The packet loss ratio can be further described by its loss ratio and loss patterns.

3. Availability Measures

At a specific time frame, GIG transport services may be unavailable between specific points, because of one or more of the following conditions:

1. System or link failure due to component or software failures
2. Congestion, caused by unplanned traffic surges.
3. Network disruption, caused by intentional damage to network components or denial of service attacks

Typical availability numbers range from .95 to a highly mobile wireless network to .995 for a fiber optic network with a large number of alternative paths. Availability is mainly a long term measure. However, for domains with significant fluctuations and correlated down times, it may be necessary to provide metrics for long and short term availability.

4. **Mean Time To Repair (MTTR)**
   MTTR is the time period measured from the time the service becomes unavailable until the time the service is restored MTTR is a function of the system/device complexity and environment and could range from minutes if there is a high degree of redundancy and on-site support, to days for unattended links in locations where travel is difficult.
5. **Session Denial Probability** (e.g. service access congestion)
   This parameter indicates the probability that a request for a session is denied. Service denial could occur for several reasons. For example, one or more QoS parameters cannot be satisfied along an end-to-end data path. There are insufficient network resources. There is a security brief.

2.3.3 Single Network Service Domain Metrics Example

Table 4-1 provides an example of SLS metrics for a one-way, point-to-point video service. The SLS would also include a statement describing the end points for the service.

SLA Metric	Example Value
Committed Information Rate (CIR)	14 Megabits per second
Peak Information Rate (PIR)	19 Mbps
Business and Time of Day Considerations	Need CIR and PIR Full Time
Average IP Packet Transfer Delay	300 ms
99th Percentile of IP Packet Transfer Delay	450 ms
IP Packet Loss Ratio	.3 percent
Service Availability for Full Requirement	.995

Table 2-1 SLA Metrics and Examples

2.3.4 Multi-Network Service Domain SLA Metrics

It will be necessary to allocate QoS if a User Community requires a path through multiple NSDs. In general, there are three types of operations for multiple domains QoS metrics calculation (see example in Table 4-2):

- *Additive,* for metrics that must be summed across the multiple domains. For example, the total delay in a network is the sum of the delays in each domain.
- *Multiplicative*, for metrics that must be multiplied across multiple domains**.** For example, the total availability for a service is the product of the availability in each domain used to support the service.
- *Concave or Least Common Denominator*, for metrics whose value is the minimum of the values in each domain**.** For example the effective throughput of a multi-domain path is determined by the NSP that provides the least throughput in its domain.

An example of multi-domain performance requirement allocation is shown in Table 4-2 for a 1 Mbps virtual circuit CIR spanning two domains.

Parameter	Service Provider 1 Domain	Service Provider 2 Domain	Multi-Domain Aggregation
Committed Information Rate (CIR)	1 Mbps (full duplex)	1 Mbps (full duplex)	1 Mbps (full duplex)
Peak Information Rate (PIR)	1.5 Mbps	1.5 Mbps	1.5 Mbps
Average IP Packet Transfer Delay (IPTD)	30 ms	60 ms	90 ms
99th percentile of IPTD	90 ms	150 ms	240 ms
IP Packet Loss Ratio (IPLR)	3.0E-3	3.0E-3	~ 6.0E-3
Availability	99.9%	99.9%	~ 99.8%

Table 2-2 Multi-domain Network Performance Allocation Example

Other SLA metrics examples related to different service offerings (i.e., video conference) is shown in Appendix D – SLA Metrics Examples.

3 Appendix A – Supporting Material

3.1 Terminology and Definitions

Terms Related to the GIG Infrastructure and Its Components (with Focus on Network infrastructure and Network Level Services)

GIG Information Systems Infrastructure: The set of computers, database, and other resources (along with intelligence in associated software) used for storing and processing information for providing the GIG Enterprise Services.

GIG Network Infrastructure: The set of communication and networking resources along with embedded intelligence used to transfer bits, bytes, and packets from one device to another.

GIG Control Infrastructure: The set of processing resources associated with the **GIG Network Infrastructure** to enable near real time control of traffic that the **GIG Network Infrastructure** is subjected to and the way the **GIG Network Infrastructure** resources are used to support that traffic.

GIG Network Management Systems Infrastructure: Set of computing, database, communication resources used to help collect and process the information about the status of the **GIG Network Infrastructure** and the **GIG Control Infrastructure** and to allow human operators to take actions to help improve services provided by the **GIG Network Infrastructure** and the **GIG Control Infrastructure** and the **GIG Network Management Systems Infrastructure**.

Network Service Domain (NSD): A set of connected network resources with the following properties:

1. well defined interfaces to the rest of the GIG
2. relatively homogeneous physical and link level technologies within (including any layer between IP and link layer)
3. Common QoS mechanisms and policies within
4. Primarily administered by one organization

Network Service Region (NSR): A subset of GIG Network Infrastructure (usually a collection of Network Service Domains) using Interior Gateway (Routing) Protocols within and External Gateway (Routing) protocols (EGP) between itself and other Network Service Regions.

Network Service Region Border Routers: IP Gateways to connect **Network Service Region**s and to share information between **Network Service Region**s.

Terms Related to Entities, Organizations, and Processes

GIG User Community also called User Community (UC): A group of users representing a common mission or a business from the perspective of communication/networking**.**

Network Service Provider (NSP): An organization responsible for offering a set of packet transport services between edges of one or more **Network Service Domains (NSD)** under its control. This organization will represent edge-to-edge service offerings from each of the **Network Service Domains** in terms of standard metrics, will negotiate SLAs with the **GIG Cooperative Agency**, and will help provision and control **Network Service Domains** to meet the SLAs.

Mission Planner (MP): An organization responsible for understanding and meeting the communication needs of a GIG User Community (among its members and between its members and other GIG User Communities). Responsibilities include:

- Understanding various users, end equipment, geographical distributions, communications needs between locations by major traffic types, etc.
- Understanding performance requirements of user applications
- Summarizing the above in aggregate forms
- Working with the **GIG Cooperative Agency** to make sure that a useful set of standards are defined for creating SLAs
- Working with the **GIG Cooperative Agency** to create network SLAs by specifying traffic parameters and performance metrics consistent with its user requirements and GCA offerings
- Monitor SLA compliance and user satisfaction and work with both **GIG User Community** and (GCA) for ongoing satisfaction and improvements

- Presenting special, unplanned need of its **GIG User Community** to the **GIG Cooperative Agency**.

GIG Cooperative Agency: An organization with representation from all major **Network Service Provider**s and possibly a third party, responsible for standardization of the SLA metrics and SLA process, and for brokering the end-to-end (transport service offering) by leveraging the (service offerings) of individual **Network Service Providers**. Responsibilities include:

- Creating a set of standards for specifying traffic and performance requirements of User Communities and for specifying traffic and performance offerings of the (**Network Service Domains**) controlled by **Network Service Providers**.
- Creating methods for aggregating requirements from **GIG User Communities** (provided by **Mission Planners**) and allocating among **Network Service Domains** to be consistent with (Service Offerings) and **GIG User Community** requirements.
- Creating SLAs with (**Mission Planners**) and with (**Network Service Providers**) for end-to-end transport for each **GIG User Community**, and for each **Network Service Domain** for the aggregate traffic of **GIG User Communities.**
- Monitoring and working with **Mission Planners** and (**Network Service Providers**) for compliance and ongoing improvements
- Working with **GIG User Communities** and **Network Service Provider**s to meet special, unplanned needs, and (where appropriate) to create policies for handling unplanned situations in automated way.

End-to-end: End user to end user with one or more domains in between. Multi-domain edge-to-edge with the end user access to the ingress and egress at each end**.**

Single domain edge-to-edge: Ingress edge to egress edge across a single domain.

Multi-domain edge to edge: Ingress edge to egress edge with multiple intervening domains.

User edge-to-edge: User edge network to user edge network. It may involve one or more NSDs.

Terms Related to Data Rates, Traffic Patterns, etc.

Throughput: Throughput is the rate at which a computer or network sends or receives data.

Committed Information Rate (CIR): Average Data Rate in bits/sec between two edges. The average is over a specified interval. From the provider's perspective, CIR is the minimum value that it is committed to support. From the user perspective, it is the maximum value it can send and still receive a guaranteed service.

- The promise may be to the **GIG Cooperative Agency**, a subset of User Communities, or to another **Network Service Provider**.
- There may be multiple separate CIRs associated with different traffic types.

Peak Information Rate (PIR)

Peak Data Rate in bits/sec between two edges. For details used in SLA, see CIR.

Committed Burst Size (CBS)

Minimum Burst Length in bits at Peak Rate between two edges. For details used in SLA, see CIR.

Terms Related to Performance Metrics

Metrics Related to Packet Transfer Delay (Source: ITU-T Y.1540)

- IP Packet Transfer Delay (IPTD) is the time ($t2 - t1$) between the occurrences of two corresponding IP packet transfer reference events.

- *Mean IP packet transfer delay* is the arithmetic average of IP packet transfer delays for a population of interest.

  $\text{Delay}_{e2e} = \sum_{i=1}^{N} D(i)$

  Where, *D(i)* is the delay of each network domains

  N is the total number of domains in a user-edge-to-user-edge network

  $D(i) = \sum_{h=1}^{M} D(h)$

  Where, *D(h)* is the average hop delay within each domain.

  M is the total number of hops within each domain

IP packet transfer delay describes the average time a network takes to transfer packets between ingress and egress measurement points (MPs). IPTD limits will be crucial to the successful deployment of VoIP, videoconferencing, and real-time data applications.

IP packet loss ratio (IPLR)

- IPLR is the ratio of total lost IP packet outcomes to total transmitted IP packets in a population of interest.

IP packet loss ratio expresses the likelihood that a packet entrusted to a network at an ingress interface is not delivered to the appropriate egress point(s). IPLR must be limited to ensure intelligibility and acceptable image quality in voice and real-time video applications, and to maintain reasonable efficiency in other applications.

IP packet delay variation (IPDV) or Jitter

- IPDV is defined based on observations of corresponding IP packet arrivals at ingress and egress Measurement Points (e.g., MP_1, MP_2 in Fig. 3).

- The packet delay variation (v_k) is the difference between the *absolute IP packet transfer delay (x_k)* of the packet and a *defined reference IP packet transfer delay, $d_{1,2}$* ,

$v_k = x_k - d_{1,2}$

Or

$$\sigma = \sqrt{\sigma_1^2 + \sigma_2^2 + ... + \sigma_N^2} = \sqrt{v_1 + v_2 + ... + v_N}$$

where,

$$\sigma_k = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2}; \qquad \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

IPDV must be controlled to avoid underflow or overflow in IP routers or terminal buffers.

IP packet loss ratio (IPLR) is the ratio of total lost IP packet outcomes to total transmitted IP packets in a population of interest.

- IP packet loss ratio expresses the likelihood that a packet entrusted to a network at an ingress interface is not delivered to the appropriate egress point(s).

- IPLR must be limited to ensure intelligibility and acceptable image quality in voice and real-time video applications, and to maintain reasonable efficiency in other applications.

IP packet error ratio (IPER) is the ratio of total errored IP packet outcomes to the total of successful IP packet transfer outcomes plus errored IP packet outcomes in a population of interest.

$$\text{PLR} = 1 - \prod_{i=1}^{N}(1 - PLR_i) \approx \sum_{i=1}^{N} PLR_i$$

Spurious IP packet rate (SIPR) at an egress MP is the total number of spurious IP packets observed at that egress MP during a specified time interval divided by the time interval duration.

IPER and SIPR express the likelihood that user data delivered at an egress interface differs from the input data as a result of corruption, duplication, or misrouting in the network.

x% PD: x percentile of IPTD between two points. (move)

(IPTD): Standard Deviation of IPTD between two points.

Residual Bit Error Rate (RBER): Residual Bit Error Rate in information delivered from IP to higher layer at the receiver. (MOVE)

Service Unavailability (SU): Intervals of lengths >= 10 seconds during which no packets could be transmitted successfully divided by the total interval of time of effort to transmit.

Service Availability (Availability): 1- SU

A route connecting two specific parties consists of nodes and links in serial and/or Parallel (alternate route). The availability between the two parties may be described as follows:

P_{series} (service availability) = $\prod_{i=1}^{N} Pa(i)$ If network components connecting in series

$P_{parallel}$ (service availability) = $1-(\prod_{i=1}^{N}(1 - Pa(i))$ If components connecting in parallel

P_{e2e} (service availability) = 1- P_{series} * $P_{parallel}$

Where, $P_a(i)$ is the probability of service availability in each domain,
N is the total number of domains in a user-edge-to-user-edge network.

Service Denial Probability: The probability that end user's attempt to use a network service of the NSP is denied. The Service Denial Probability is expressed as the ratio of the number of denied service attempt to the total number of attempts. The Service Denial Probability is given by the following formula:

$$P_{e2e} \text{ (service denial probability)} = 1- \prod_{i=1}^{N} \{1 - Pd(i)\} \approx \sum_{i=1}^{N} Pd(i)$$

Where $P_d(i)$ is the service denial probability in each domain, and N is the total number of domains in an end-to-end network.

Mean Length of Down Time: Mean Length of Unavailability (larger than 10 seconds).

Mean Time Between Intervals of Unavailability: Same as Mean Time Between Failure (MTBF)

Terms Related to Requirements, Capabilities, Agreements

Applications Performance Requirements: Values of performance metrics required by an application or a service class in a given environment and at a given precedence level.

- Voice packet delay requirements may be different for C/P/S, tactical MANET, and tactical satellite access situations.
- Requirement on **Blocking Probability** may be different for different precedence levels.

User Community Performance Requirements: Values of multi-domain edge-to-edge performance metrics required by a **GIG User Community** for a set of applications.

GIG Cooperative Agency Performance Requirements: Values of edge-to-edge (for a **Network Service Domain**) network layer performance metrics required by a **GIG Cooperative Agency** from a **Network Service Domain**.

Network Service Provider Performance Requirements: Values of performance metrics promised by a **Network Service Provider** for a **Network Service Domain** it controls.

Service Offerings: Ranges of values of performance metrics offered by a **Network Service Provider** for a **Network Service Domain** it controls.

Service Class-to-Service Level Mapping: Mapping of Service Class to a **Service Offering** for a **Network Service Domain** (may be many to one).

Mission SLA (MSLA): An agreement between a **Mission Planner** and a **GIG Cooperative Agency** to meet end-to-end performance metrics provided the traffic between specified locations remains within traffic bounds specified in terms of CIR, PIR, etc.

Network Service Domain SLA: An agreement between a **GIG Cooperative Agency** and a **Network Service Provider** to meet edge-to-edge performance metrics for an **Network Service Domain** if the traffic between these edges remains within specified bounds.

Available Capacity: Current capacity available to carry additional traffic at a given range in a **Service Offering** between specified edges of a **Network Service Domain**.

Network Service Provider Performance: Values of performance metrics realized by a **Network Service Provider** on a **Network Service Domain** (for realized traffic volume parameters).

Inter-GIG User Community Traffic Profile: Values of traffic parameters realized between two locations of a **GIG User Community**

Offered Traffic Profile: Values of the parameters of the traffic submitted (measured) for transport between two edges of a **Network Service Domain**.

4 Appendix B – Service Class Examples

Table B-1 End-to-End GIG Service Class

No.	Service	Service class	Example
1	Signaling	Network Control	Routing, network management, QoS signaling
		User signaling	IP telephony signaling
2	Inelastic/Real time	Voice	IP telephony
		Video	Interactive video conferencing
		Commands, and sensor, other short transactions	Sensor-to-shooter, Unmanned Aerial Vehicle(UAV) control
		Circuit emulation	T1 over IP
		Streaming	Broadcast video, video Intelligence, surveillance and Reconnaissance (ISR) from a UAV
3	Preferred Elastic	Interactive transactions	IM
		File Transfer	Imagery, Target List management (TLM)
4	Elastic	Default	E-Mail, Web browsing, document transfers
		Scavenger	Web browsing

5 Appendix D – SLA Metrics Exemple

Services	Traffic Characteristic, Bounds, and QoS Requirements							
	Data Class	**CIR (bps)**	**PIR (bps)**	**CBS (Byte)**	**IPTD (sec.)**	**IPDV (sec.)**	**BER**	**PLR**
VoIP	Conversational	4k – 16k	32k	1K	0.1 – 0.2	0.01 – 0.8	1.0E-2	1.0E-2
T1 over IP	Circuit Emulation	256k	512	TBD	0.01	TBD	TBD	1.0E-4
Video conference	Streaming	64k – 500k	1M	500K	2	1	1.0E-2	1.0E-2
Video Surveillance	Streaming	500k	1M	500K	20	5	1.0E-2	1.0E-2
Video Target	Streaming	500k	1M	500K	0.1 – 1.0	0.5	1.0E-4	1.0E-3
Web Browsing Request	Interactive	1k	10k	2K	5	1	1.0E-5	1.0E-5
Web Browsing Downloading	Interactive	100k	500k	1M	5	1	NA	1.0E-5
Bio Sensing (sensor to satellite)	Interactive	1k	50k	100K	30	10	1.0E-5	1.0E-5
Image Sensing (satellite to sensor)	Interactive	100k	5M	1M	4	1	1.0E-5	1.0E-5
Image sensing (sensor to satellite)	Interactive	10k	500k	1M	8	1	1.0E-5	1.0E-5
Image sensing (sensor to UAV)	Background	10k	500k	1M	5	1	1.0E-5	1.0E-5
Image sensing (UAV to sensor)	Background	5M	500k	1M	3	1	1.0E-5	1.0E-5

References

[1]. Dinesh V. Verma, “Service Level Agreements on IP Networks”, http://www.research.ibm.com/people/d/dverma/papers/SLAOverview.pdf.

[2]. http://www.mitre.org/news/events/xml4bin/pdf/robinson_strategy.pdf

[3]. J. Hou, H. Tyan, and B. Wang, “QoS Extension to CBT”, IETF Internet Draft <draft-hou-cbt-qos-00.txt>, Feb. 1999

[4]. A. S. Robinson and D. Lounsbury, “Measuring & Managing End-to-End Quality of Service (QoS) Provided by Linked Chains of application and Communications Services” http://www.opengroup.org/RI/recent/papers/End_to_End_QoS.pdf.

[5]. J. Jacobs P. and Davie B., “Technical Challenges in the Delivery of Interprovider QoS", June 2005, Vol. 43, No. 6 IEEE Communications Magazine

[6]. Y. Cheng, R. Farha, A. Tizghadam, M. S. Kim, M. Hashemi, A. Leon-Garcia, and J. Won-Ki Hong, “Virual Network Approach to Scalable IP Service Deployment and Efficient Resource Management”, IEEE Communications Magazine, Oct. 2005

[7]. I. sorteberg and ϕ.Kure, “The Use of Service level Agreements in Tactical Military coalition Force Network”, IEEE Communications magazine, Nov. 2005