INTERNATIONAL CIVIL AVIATION ORGANIZATION
ASIA AND PACIFIC OFFICE

ASIA/PAC COMMUNICATION PERFORMANCE FOR ATN

First Edition - April 2005

TABLE OF CONTENTS

Related References

[1] ICAO Manual on Required Communication Performance (to be presented at OPLINKP/1 meeting, Sept. 2005)
[2] Document 9694 ICAO Manual of ATS Data Link Applications
[3] Document 9739 ICAO (CAMAL) Edition 2 Part V — Enhanced ATN Services
Chapter 2 — ATN Systems Management, 2.2.2 Performance Management Requirements

Appendix

ICAO Document 9739 (CAMAL) Edition 2
Part V Enhanced ATN Services
Chapter 2 ATN Systems Management,
2.2.2 Performance Management Requirements

EXECUTIVE SUMMARY

This document provides guidance on the aeronautical communication performance to be established. Before describing communication performance, the characteristics of aeronautical communication and performance are described.

There are many elements in the CNS/ATM system. In order to provide aeronautical communication services, both functional capability and its performance need to be identified. There are operational and technical levels, both having functional capability and its performance and the required performance. The performance requirement specifies how well the planned system should be implemented. While the required system is implemented, performance is monitored and evaluated against the required performance. The required and implemented performance cannot be justified at technical or element level without aligning them to the required performance at the operational or system level where the elements are part of the operational services or the system.

The document provides the information on the model of (operational) aeronautical communication process and communication transaction developed by OPLINK Panel and contained in the ICAO RCP Manual. Based on the performance characteristics, the RCP types are introduced in the ICAO RCP Manual. Three specific RCP types; RCP 240, RCP 120 and RCP 15 are also provided in the RCP Manual.

While determining RCP type for a planned ATM function in airspace, the communication transactions are identified and the most stringent communication performance requirement is selected as the RCP type for the required ATM function.

Most stringent communication performance requirement is selected as the RCP type for the airspace while prescribing RCP type for ATM functions in a planned airspace.

While complying with RCP type, the elements of operational communication performance requirements including technical and human factor will have to be derived. Where the prescribed RCP type is identified as the compliance target, the actual performance monitoring of each element will provide the compliance evidences of the prescribed RCP type. How and who are responsible to derive technical performance requirements are not identified in the ICAO RCP Manual. The regional planning group is expected to address this issue.

A brief description of ATN applications and ATN communication services is also provided. The performance characteristics and network topology of each ATN application and the components of ATN communication services for air-ground as well as ground-ground communication are shown in diagrams.

Relevant information regarding the ATN performance monitoring is extracted from the ICAO Document 9739 and provided in the Appendix to the document for easy reference.

1. Introduction

1.1 Background

The ATN Transition Task Force (ATNTTF) has been assigned a number of tasks to prepare the region for the introduction of the ATN. At the third meeting of the ATNTTF Working Group B meeting held in Bangkok, Thailand on 27 through 30 August 2001, a specific action item was identified to develop documentation on Performance. This document was developed in response to that action.

1.2 Scope

The document provides guidance on the aeronautical communication performance to be planned, implemented and also for operation. The ultimate goal of the document is to provide a set of performance requirements for planning, technical candidates for implementing the required performance, and performance management consideration during the operational phase of the Aeronautical Telecommunication system in the Asia/Pac Region.

Before providing such a detailed, concrete guidance, it has to be clearly understood what is meant by the performance we are discussing, and how the performance requirements is planned and implemented in the aeronautical communication systems (including hardware, software and human machine interface), and also how the planned performance is monitored and improved. Without such understandings, the performance measurements and evaluation will be wasted without any useful results.

In order the document to be useful in the Asia/Pacific region where the operational environments are considerably different within the Region in terms of the air space complexities and traffic (air traffic as well as data traffic), the guidance in the document has to be adaptable to the environment of the interest.

Performance is related to all phases of the activities in the Aeronautical Telecommunication System Life Cycle; the planning and implementation phases as well as operational phase. This implies there are various types of personnel involved as the readers of the document. It is important to identify the readers of the document, and the organization of the document for the focused readers. Since the APANPIRG is a planning group, and the planning phase has the greatest influences on the performance, the document is focused to the planning phase. Essentially the implementation and operational aspects of performance are embedded within the planning aspect. Without any planning, the monitoring and evaluating performance are not useful, since it would not be clear what is to be measured and what is the reference to be compared with and evaluated to.

2. Preliminary Considerations on Performance

2.1 Functional Capabilities and its Performance

For any systems to be implemented and operated, there are two categories of requirements; functional capability and its performance. Functional capability requirements specify the functions to be implemented together with benefits of the planned system. Performance requirements specify how well the implemented system has to be operated in terms of speed, reliability, and so forth.

In order to effectively and efficiently provide the benefits of the planned system specified in the functional capability requirements, the performance requirements have to be established and implemented.

The functional capability requirements of the planned system like Global Air Traffic Services can be found in SARPs and Regional ICDs (Interface Control Documents), but the performance requirements cannot be documented as a single body of requirements.

Even if the functional capability requirements, like the Global Air traffic Services, are the same, the required implementation details can be fairly different, since each system implementation consists of the products or services available in the planned airspace.

In airspaces, it may be planned to provide different mix of ATM functions from others, and the most stringent performance requirement will be dominant in the performance considerations.

There would be structural complexity of airspace different from other airspaces, so that it may be required to reflect such complexity to the performance considerations.

There could be the case that any implemented system satisfying the functional capability requirements do not mean that it always satisfies the needed performance.

The performance requirements have to be developed based on the ATM function type and the airspace environment where the CNS/ATM System is planned to be used. Based on such performance requirements, the implementation, monitoring, evaluation and improvement of the CNS/ATM system will follow, satisfying the required performance.

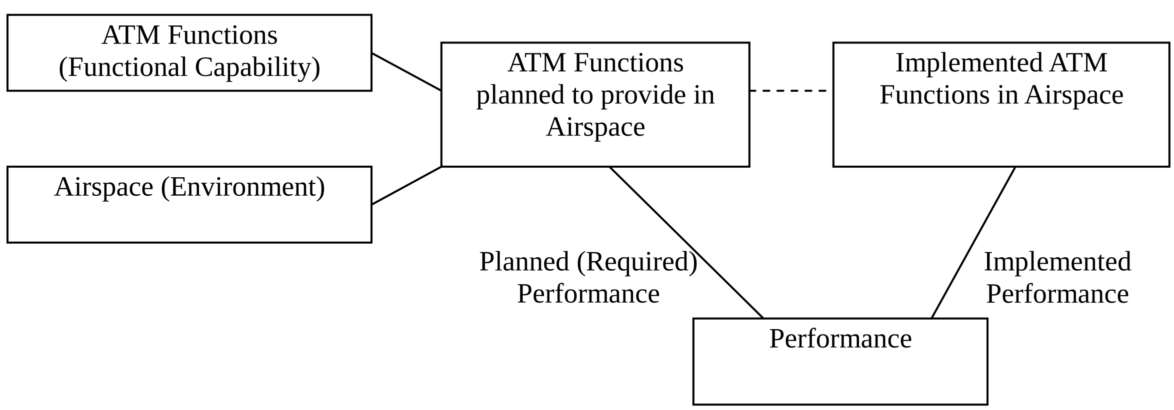

Figure 2.1 Functional Capability and its Performance

2.2 Required Performance and Implemented Performance

Performance is determined in the planning phase, and the performance is monitored and evaluated after implementation. It should be noted that in the activity for determining performance, the *required* performance is determined, while in the activity for monitoring and evaluating performance, the *implemented* performance is monitored and evaluated against the required performance to be complied with.

Obviously the monitoring and evaluating implemented performance are meaningful only if the required performance is appropriately determined.

If the required performance is determined unnecessarily high, then the cost of implementation will be unnecessarily high. If the required performance is determined too low than necessary, the operation based on the system is deteriorated or the system is useless.

The determination of required performance is not simply to give a number but it has to give rationale for determining the required performance.

2.3 Activities related to Performance

Although the planning activity for any system implementation can be taken place within a State or an organization without any coordination with other activities, it is desirable for the planning activity to be coordinated with subsequent activities, such as implementation, monitoring, evaluation and improvement, since the planned performance requirements are referred by the subsequent activities and there are many different personnel involved in these activities. Moreover, in a planned system like Global Air Traffic Services, where many States and Organizations are involved, it is imperative to coordinate among the States and Organizations.

In order to make such coordination for planning possible, it has to be agreed and shared the understanding of what is meant by 'performance', how it can be analyzed and stated. This can be called the 'understanding performance' prior to determining required performance.

a) Understanding (Required) Performance

Before engaging in the performance related activities where the coordination is needed, the meaning of the performance has to be shared among the concerned parties. This includes the definition of the performance as well as the procedures to derive and document the required performance.

b) Determining Required Performance

Applying the shared understanding of performance, a specific performance requirement has to be developed for the Air Traffic Services in the airspace where the CNS/ATM system is planned.

The performance requirements have to be developed based on the analyses of the services provided and the environment of the services provided.

c) Prescribing Required Performance

The determined required performance has to be documented in a formal way.

d) Complying Implemented Performance with Required Performance

Based on the stated performance requirements, the system has to be developed. The system has to be deployed based on the satisfied implementation of the requirements.

e) Monitoring and Evaluating Implemented Performance

The performance of implementation has to be monitored and evaluated regularly. The monitoring parameters have to be planned before the implementation.

e) Improving Performance

If there are any anomalies in the monitored performance, the actions have to be taken to improve the performance. There should be the performance reference established during the planning phase to judge the performance anomaly.

2.4 System and System Elements

Any system consists of system objectives and system elements.

For instance, the Air Traffic Management system has ATM system objectives and ATM system elements; namely Communication, Navigation and Surveillance.

The Communication sub-system in turn has the sub-system objectives and sub-system elements, namely ground sub-system elements, airborne sub-system elements and air-ground communication sub-system elements.

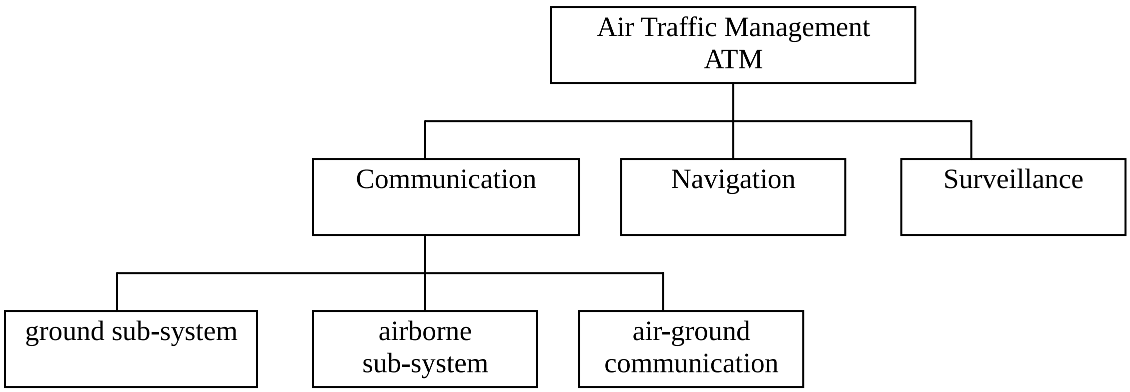

Figure 2.2 System and System Elements in CNS/ATM

Each system, sub-system and element has its own functional capability as well as its own performance characteristics. Even if one element of sub-system is implemented with potentially high performance but other elements are at lower levels of performance, the potential performance of the element is under-used, i.e. its performance is excessive and cannot contribute to the performance of the sub-system as a whole. Conversely, if one element of sub-system is at lower level of performance than other elements, then the performance of sub-system is bound by the lower level of performance of that element.

What is important here is that performance of each element is at same level and it is also at same level as the required performance of the sub-system.

In the CNS/ATM context, the Required Navigation Performance (RNP) and the required

Communication Performance (RCP) is defined, and the Required Surveillance Performance (RSP) may also be defined. Even defining the Required Total System Performance (RTSP) is suggested.

The navigation system is considered as an autonomous, while the communication system is considered as an end-to-end system of response oriented. The surveillance system may be considered as a throughput oriented as well as a response oriented system.

These three elements of the ATM have different performance characteristics and their performance parameters are currently incompatible to the others.

2.5 Operational Performance Requirement and Technical Performance Requirement

Usually systems provide some services while the systems consist of some elements. With respect to performance, there are two levels of performance requirements, one at the operational service level and the other mainly at the technical and element level. For instance, the ICAO RCP is at the operational performance level, while there are many elements in communication system with their own technical performance. As noted before, these two levels have to be aligned to each other.

One of the issues to be dealt with is how to map the operational performance requirements to the technical performance requirements of consisting elements.

Technical performance, whether it is required or implemented, cannot be justified without the matching operational performance requirements, since the ultimate goal of system is to provide services not technology.

Another issue is whether the performance requirements are regulatory or guides for implementations and operations.

3. Operational Communication and its Performance in CNS/ATM

Before discussing communication performance, it is necessary to agree with the model of communication in the CNS/ATM, especially operational communication process model. OPLINKP adopted the following model for the operational communication process.

3.1 Operational Communication Transaction

The Figure 3.1 shows the operational communication transaction as a communication process model.

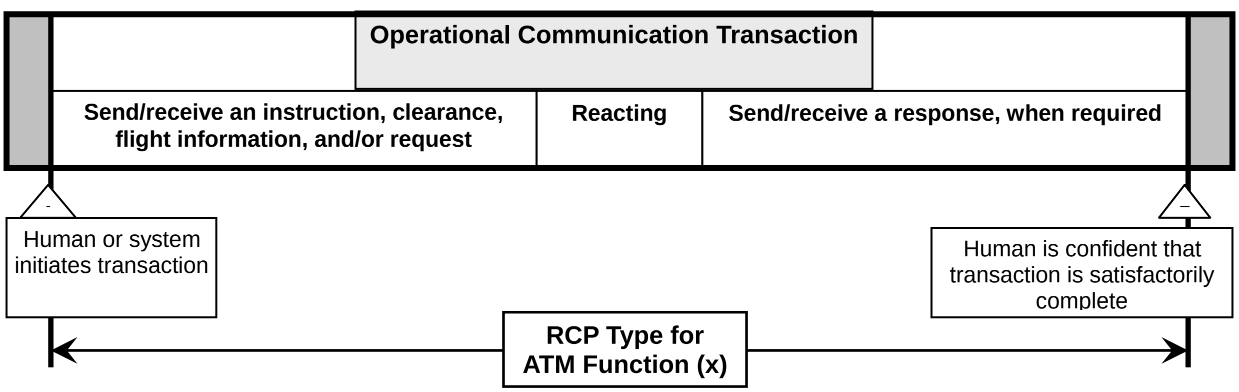

Figure 3.1 Model of Operational Communication Transaction

It should be noted that the RCP type is defined for each ATM Function.

An operational communication transaction starts when human or system initiated the transaction. It should be noted that the transaction could be initiated by a human or automatically by the system.

There are three parts in the transaction;
1) Send/receive an instruction, clearance, flight information, and/or request,
2) Reacting
3) Send/receive a response, when required.

The first part of transaction (send/receive an instruction, clearance, flight information, and/ request) includes the human machine interface (HMI) of initiator, the technical communication from initiator to responder, and the HMI of responder.

The second part of transaction is the reaction by responder. The third part of transaction is similar to the first part except that the direction of communication is from the responder to the initiator of the transaction.

The transaction includes human factors as well as technical factors. The transaction ends when human is confident of completion. It should be noted that there is always human involved, even if the transaction is initiated by system.

It should also be noted that the transaction is two-way communication when required.

In order to communicate, communication transaction has to be initiated. There is a possibility that the communication transaction may fail to be initiated. The communication transaction is either initiated or not initiated.

After a communication transaction has successfully been initiated, the communication transaction may be completed within a given time duration. There is a possibility that the communication transaction will not be completed within the given time duration, taking longer or even never terminating. The communication transaction is either completed within the given time duration or not completed within the given time duration.

If the communication transaction is not completed within the given time duration, the initiator of the communication transaction should revert to an alternative procedure.

If the communication transaction is completed within certain time duration, there may be communication transaction errors. These errors, if any, are either detected or undetected.

3.2 Performance Characterization of Operational Communication Transaction

In this section, the operational communication transaction is characterized for performance description.

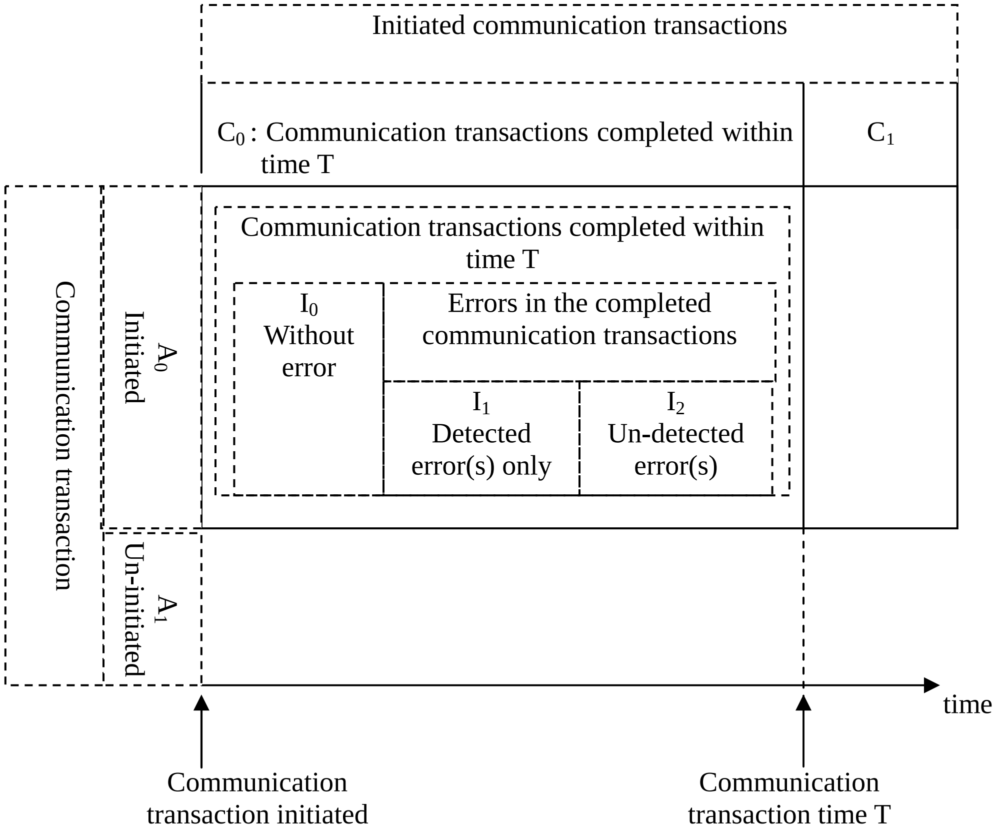

Figure 3.2 Performance Characterization of Operational Communication Transaction

The symbols in the Figure 3.2 are defined as follows.

- T: The maximum time for the completion of the operational communication transaction after which the initiator should revert to an alternative procedure.
- A_0: probability of communication transactions initiated
- A_1: probability of communication transactions not initiated
- C_0: probability of initiated communication transactions completed within T
- C_1: probability of initiated communication transactions not completed within T
- I_0: probability of completed communication transactions within T without error
- I_1: probability of completed communication transactions within T with detected error(s) only
- I_2: probability of completed communication transactions within T with undetected error(s)
- where

$A_0 + A_1 = 1$

$C_0 + C_1 = 1$

$I_0 + I_1 + I_2 = 1$

The probability of completed communication transactions within T with undetected error(s): I_2 can only be provided by manual analyses.

3.3 Required Operational Communication Performance

3.3.1 Parameters of RCP

OPLINKP has developed the Required Communication Performance: RCP. It should be noted that the

RCP developed by the OPLINKP is operational communication performance.

The performance characteristics of the communication transaction are described using the following four parameters. The set of parameters are a subset of the parameters described in Figure 3.2.

The RCP parameters are as follows.

1) Communication transaction time (T): The maximum time for the completion of the operational communication transaction after which the initiator should revert to an alternative procedure.
2) Availability (A_0): The probability that an operational communication transaction can be initiated.
3) Continuity (C_0): The probability that an initiated operational communication transaction can be completed within the communication transaction time (T).
4) Integrity (I_2): The probability of completed communication transactions within the communication transaction time (T) with undetected error(s)

3.3.2 RCP Types and Specific RCP Types

a) RCP type naming convention

In order to simplify RCP type naming convention and to make the required communication transaction time readily apparent to airspace planners, aircraft manufacturers and operators, the RCP type is specified by the value for the communication transaction time associated with the ATM function.

The values of four parameters are associated with each RCP type.

b) Specific RCP type

ICAO RCP manual specifies three RCP types;

- RCP 240,
- RCP 120 and
- RCP 15,

where the numbers; 240, 120 and 15; are the communication transaction time T in seconds.

RCP 240 is selected based on the ICAO document; PANS ATM 4444,

5.4.2.6.4 Longitudinal distance-based separation minima in an RNP RNAV environment using ADS

5.4.2.6.4.3 For aircraft cruising, climbing or descending on the same track, the following separation minima may be used:

Separation minima	*RNP type*	*Maximum ADS periodic reporting interval*
93 km (50 NM)	*10*	*27 minutes*
	4	*32 minutes*
55.5 km (30 NM)	*4*	*14 minutes*

5.4.2.6.4.3.2 The communication system provided to enable the application of the separation minima in 5.4.2.6.4.3 shall allow a controller, within 4 minutes, to intervene and resolve a potential conflict by contacting an aircraft using the normal means of communication. An alternative means shall be available to allow the controller to intervene and resolve the conflict within a total time of 10½ minutes, should the normal means of communication fail.

where '4 minutes' equates to 240 seconds.

RCP 120 is selected based on another ICAO document on 'to expedite and maintain an orderly flow of air traffic'.

RCP 15 is selected for 'separation assurance in continental en-route airspace'.

The integrity parameter value of each RCP type is selected as 10^{-3} since the communication transaction includes human.

3.4 Activities deriving RCP type

The ICAO RCP manual describes three steps of activities to determine, prescribe and comply with the RCP type. These three steps are briefly described below.

3.4.1 Determining RCP type

For a particular ATM function to be provided, the communication transactions for the ATM function are assessed, and the RCP type for each communication transaction is derived. The most stringent RCP type among the RCP types is selected for the ATM function. The assessment includes environmental factors.

3.4.2 Prescribing RCP type

An RCP type should be selected in order to meet requirements for the intended ATM function to be provided in a given airspace.

For a particular flight phase or airspace, where a mixture of ATM functions are provided, the most stringent RCP type among the RCP types of ATM functions is selected.

3.4.3 Complying with RCP type

At the complying stage, the following sub-steps are proposed,

1) establish compliance target,
2) validate assumptions, and
3) analyze actual performance.

It should be noted that the compliance activity starts to establish the target and after analyzing actual performance, evidence of compliance is documented. The actual performance is analyzed against to the compliance targets.

The following is a part of discussions at the OPLINKP with respect to compliance and approval.

There is an obligation on the part of the State and the aircraft operator to show that the procedures, aircraft equipage and airspace infrastructure comply with the RCP type. This compliance is performed as part of different approval types. The different approval types are the ATS provider approval, aircraft operator approval, and aircraft type design approval. These separate and distinct types of approvals collectively define the conceptual "ATM system approval." In cases for which there is no regulatory basis for approval of a part of the ATM system, "approval" denotes the activities that take place to show compliance with the requirements allocated to that part of the ATM system.

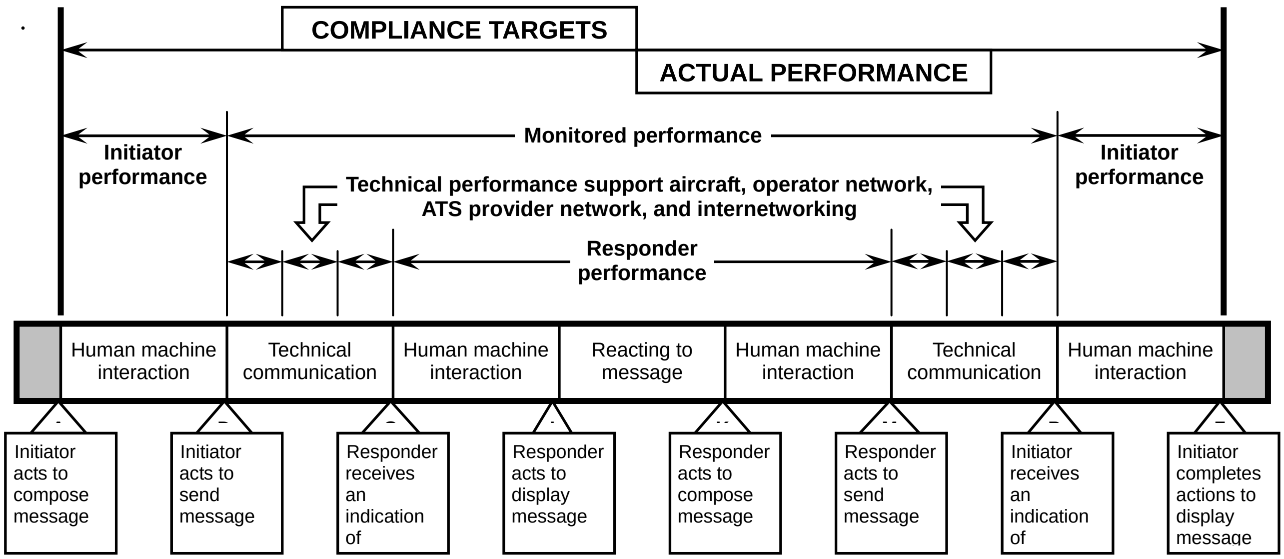

Figure 3.3 Complying with RCP type for Data Communication

a) Mapping from Operational Performance Requirements to Performance Requirements of Elements

In this stage, the compliance targets are mapped to the technical elements that comprise the communication system. Such mapping could be decided by regional group or some other organization

It should be emphasized that the sound engineering decisions have to be made in determining the RCP type and the performance requirements of consisting elements.

b) Human Factor Performance

The communication transaction includes human factors. Without monitoring human factor directly, the performance including human factors can be monitored in such a way that by monitoring the performance from point D to point P in Figure 3.3, the technical performance from point D to point G, and from point M to P, the responder’s performance from point G to point M will be provided.

3.5 Role and Responsibility of Regional Planning Group

There remain some unfinished tasks with respect to performance. Regional Planning groups may play the role of determining, prescribing, and complying with the RCP types in airspaces in the region.

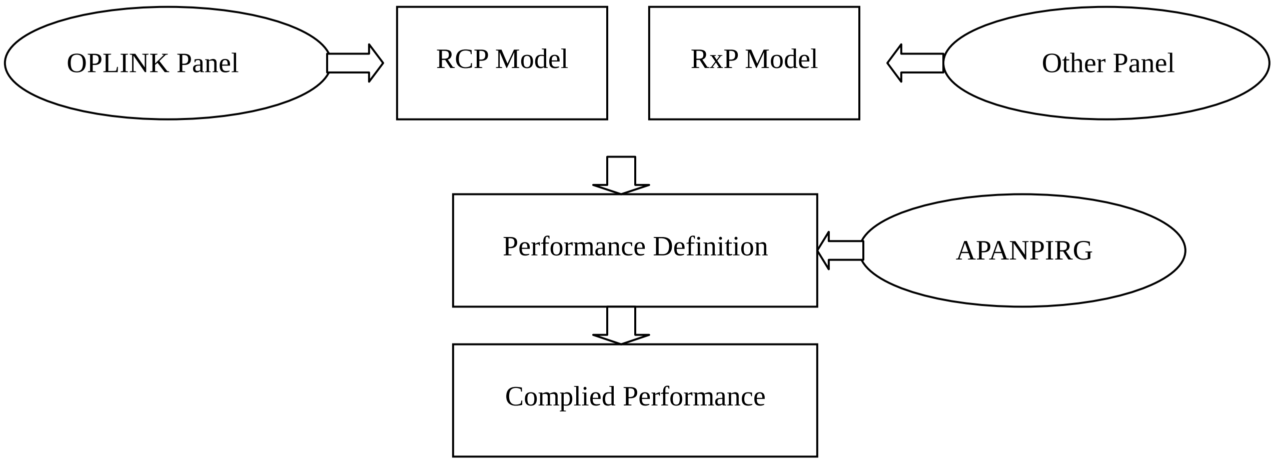

Figure 3.5 Role and Responsibility of Regional Planning Group

4. Aeronautical Telecommunication Network and its Performance

4.1 Aeronautical Telecommunication Network (ATN) in the CNS/ATM context

4.1.1 ATN Applications

The Aeronautical Telecommunication Network (ATN) is a portion of the CNS/ATM. The operational requirements (functional) for the ATS Data Link Applications are documented in the ICAO manual (DOC 9694). The Figure 4.1 depicts the position of the ATN within the CNS/ATM.

The ATS Data Link Applications documented in DOC 9694 include CPDLC, ADS (ADS-C; C for Communication), FIS, AIDC, and DLIC. They are also implemented in other communication systems as well as in the ATN as ATN Data Link Applications.

There are compatibility issues between implementations of same ATS Data Link Application in the ATN and the other communication systems.

One of the ATS Data Link Applications documented in DOC 9694 is ADS-B; B for broadcast, considered as a Surveillance capability, and is not included in the ATN Applications.

There is another ATN Application; AMHS not included in the ATS Data Link Applications. The AMHS has a corresponding legacy system; AFTN and the two communication systems co-exist via gateway functionality.

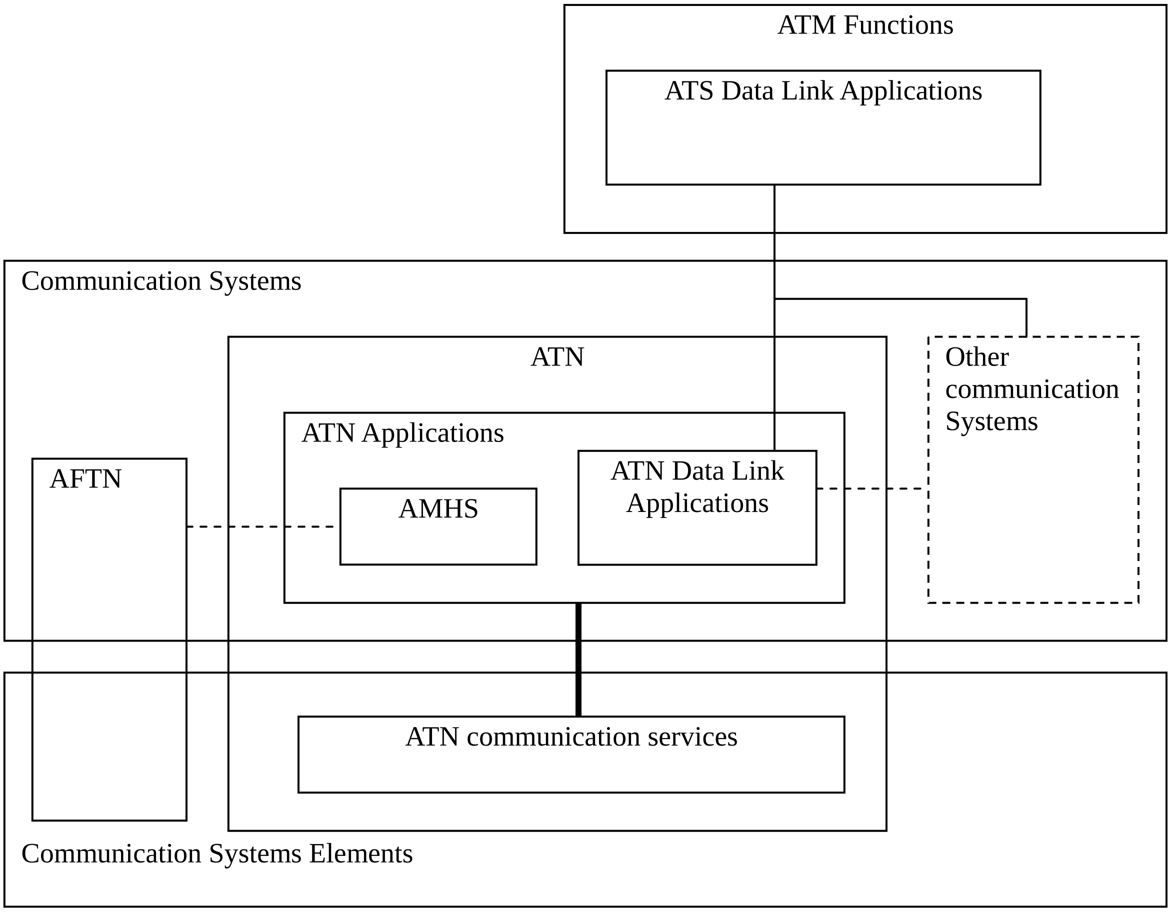

Fig. 4.1 ATN in the CNS/ATM

The ATN applications may also classified as Air-to-Ground and Ground-to-Ground. Table 4.1 depicts the classification of the ATN Applications.

Table 4.1 Classification of ATN Applications

Name of ATN Application	ATS Data Link application	Air-Ground (A-G) or Ground-Ground (G-G)
CPDLC	Yes	A-G
ADS (ADS-C)	Yes	A-G
FIS (ATIS, METAR)	Yes	A-G
DLIC (CM)	Yes	A-G
AIDC	Yes	G-G
AMHS	No	G-G

4.1.2 Characteristics of ATN Applications

There are three ATN Air-to-Ground applications; CPDLC, ADS and FIS defined as the ATN applications. Besides these three, CM (Context Management) or DLIC (Data Link Initiation Capability) is defined. In the Ground-to-Ground communication, two applications; AMHS and AIDC are defined. Each application has its own specific features and the performance requirements.

a) CPDLC

The CPDLC is the typical Air-to-Ground application that the model described in the previous sections reflects the features of CPDLC, two-way communication with human factors. The CPDLC is a type of 'Connection-Oriented-like' communication, where a 'CPDLC-Start' has to be invoked for the establishment of a CPDLC connection before any CPDLC messaging. One of CPDLC specific requirements may be the number of CPDLC connections supported at any one time, since the CPDLC will be kept connected during the flight.

b) ADS

The invocation of ADS starts at on the ground to initiate the Airborne System to generate and report the ADS messages. The ADS has less human factors where the messages are generated by the Airborne System. The ADS is a type of Connectionless-like communication, that is, the message can be sent without any 'Connection Establishment Message'. The ADS is a part of the Surveillance function, but it has a different feature from the Radar Surveillance, since the ADS transfer delay is not constant, and the ADS periodic reports from each aircraft are asynchronous. One of ADS specific requirements may be the number of ADS connections supported at any one time, since the ADS will be kept connected during the flight.

c) FIS

Currently there are two functions provided in the FIS; ATIS and METAR. The invocation of the FIS starts at the Airborne System/Pilot requesting the FIS information. The corresponding Ground System sends back the requested information, or sends back 'processing' if the requested information is not ready. Similar to the ADS, The FIS is a type of Connectionless-like communication, that is, the message can be sent without any 'Connection Establishment Message'. Currently the FIS is not a performance sensitive application, since the FIS is the planning-oriented application, but in future there may be some added functions of performance sensitive.

d) CM

The CM is not really an application in a sense of other Air-to-Ground applications. It provides the address services for the other Air-to-Ground applications. The CM is not a performance sensitive application, since it can be used only once within a specific airspace.

e) AIDC

The AIDC is a Ground-to-Ground application between two neighboring FIRs. Besides the physical connection established between two neighboring FIRs, one logical connection (i.e. association) per flight transferred between two neighboring FIRs, has to be maintained in the AIDC. One of AIDC specific requirements may be the number of the AIDC (logical) connections supported at any one time.

f) AMHS

The AMHS is a Ground-to-Ground application of the 'Store-and-Forward' mode. The AMHS is not a performance sensitive application, in fact there is no time related performance requirements for the AMHS.

4.1.3 Topological Characteristics of ATN Applications

There are three different topologies of the ATN Applications.

1) Air-to-Ground Applications Type(e.g. CPDLC, ADS, FIS, CM)

The topological scheme of End System:ES in the Air-to-Ground ATN applications is as follows.

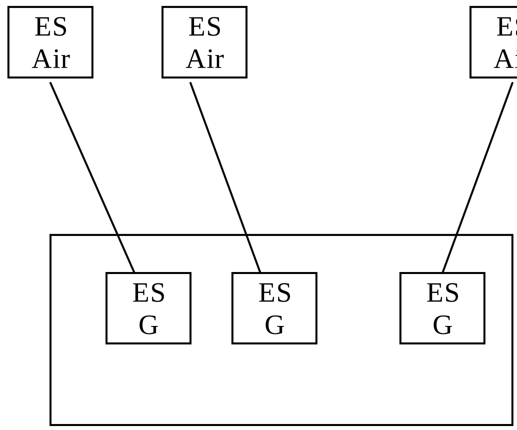

Figure 4.2 Air-to-Ground Applications Network Topology

- Two End Systems (one Airborne ES and one Ground ES) are connected by a separate connection path which includes one Air-to-Ground data link and possibly multiple Ground-to-Ground sub-networks.
- How to accommodate multiple Ground ESs of same type (e.g. CPDLC-Ground-ESs, ADS-Ground-ESs) depends on the implementation.
- Actual performance of each Air-to-Ground application, especially for Ground ES, depends on how many active connections are allowed at one time.
- It is meaningful to address the End-to-End performance, especially the response time performance of the Air-to-Ground applications for both Airborne ES and Ground ES, but it is less meaningful to address the throughput for Airborne ES.

2) Ground-to-Ground Interactive Applications Type (e.g. AIDC)

One of the topological schemes of Ground-to-Ground applications is a simple connection between two ESs (two Ground ESs) as shown below.

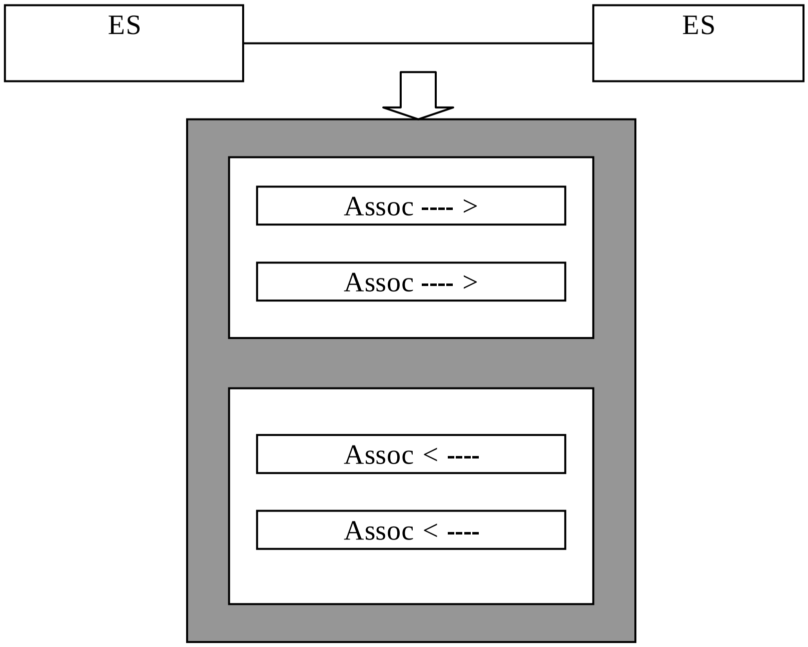

Figure 4.3 Ground-to-Ground Interactive Applications Network Topology

The connection of this type is physically simple, but logically is not simple, for instance, AIDC has a complex structure of connection.

Multiple logical connections (also called 'associations') are active at one time where one association between two ESs represents the AIDC communication with respect to each flight.

Since there exist two directions of flights between two FIRs using AIDC, there are two different classes of associations representing the flights of each direction between FIRs.

3) Ground-to-Ground Store-and-Forward Applications Type (e.g. AMHS)

Another type of topological scheme of Ground-to-Ground ATN applications is a network of servers as shown below.

Within a network of servers, where the mode of communication is 'store-and-forward', the End-to-End response time is less meaningful, although it is not useless.
Here the 'End-to-End' envisages the originator-to-destination, but the server-to-server in between can be viewed as the 'end-to-end' or 'End System to End System'.

Within a network of the store-and-forward servers, it is more meaningful to address the capability of each server, e.g. throughput, rather than the origin-to-destination aspect.

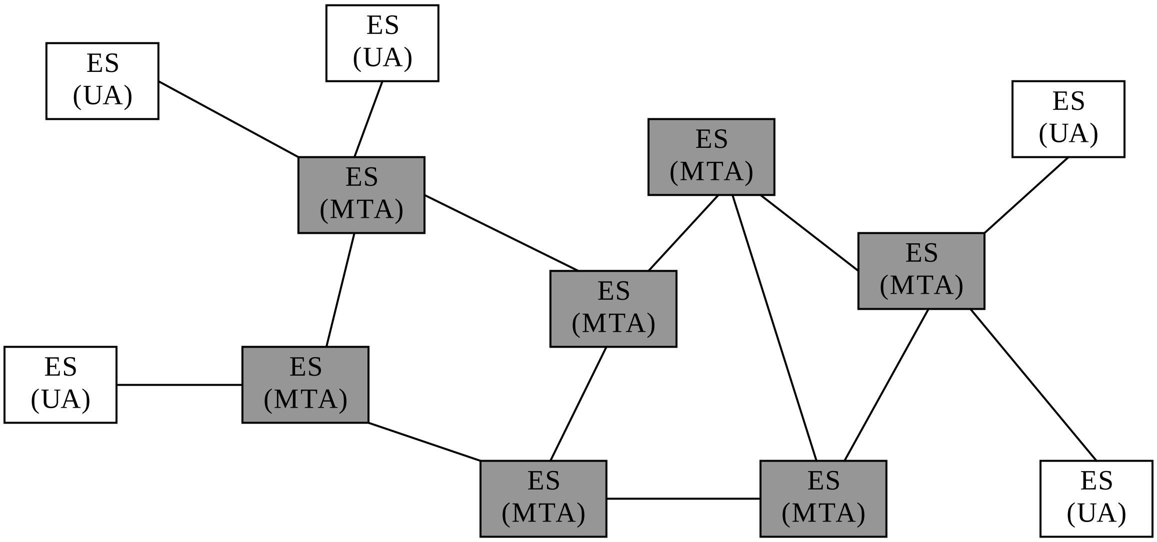

Figure 4.4 Ground-to-Ground Store-and-Forward Applications Network Topology

4.2 ATN Communication Services

4.2.1 Air-Ground ATN Application and ATN Communication Services

ATN Communication Services' elements in the Air Ground ATN Applications are depicted in the Figure 4.5.

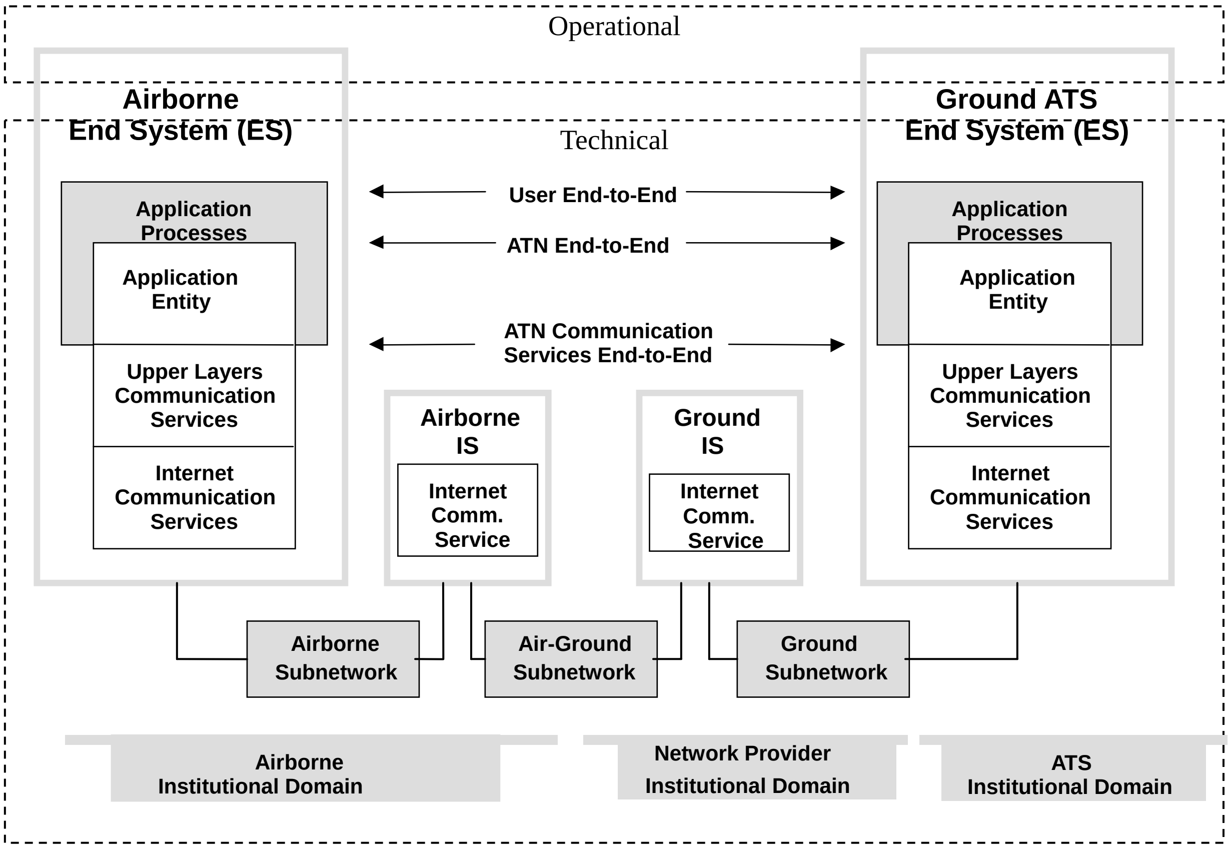

Figure 4.5 Technical Elements of Air-Ground Communication

4.2.2 Ground-Ground ATN Application and ATN Communication Services

ATN Communication Services' elements in the Ground-Ground ATN Applications are depicted in the Figure 4.6.

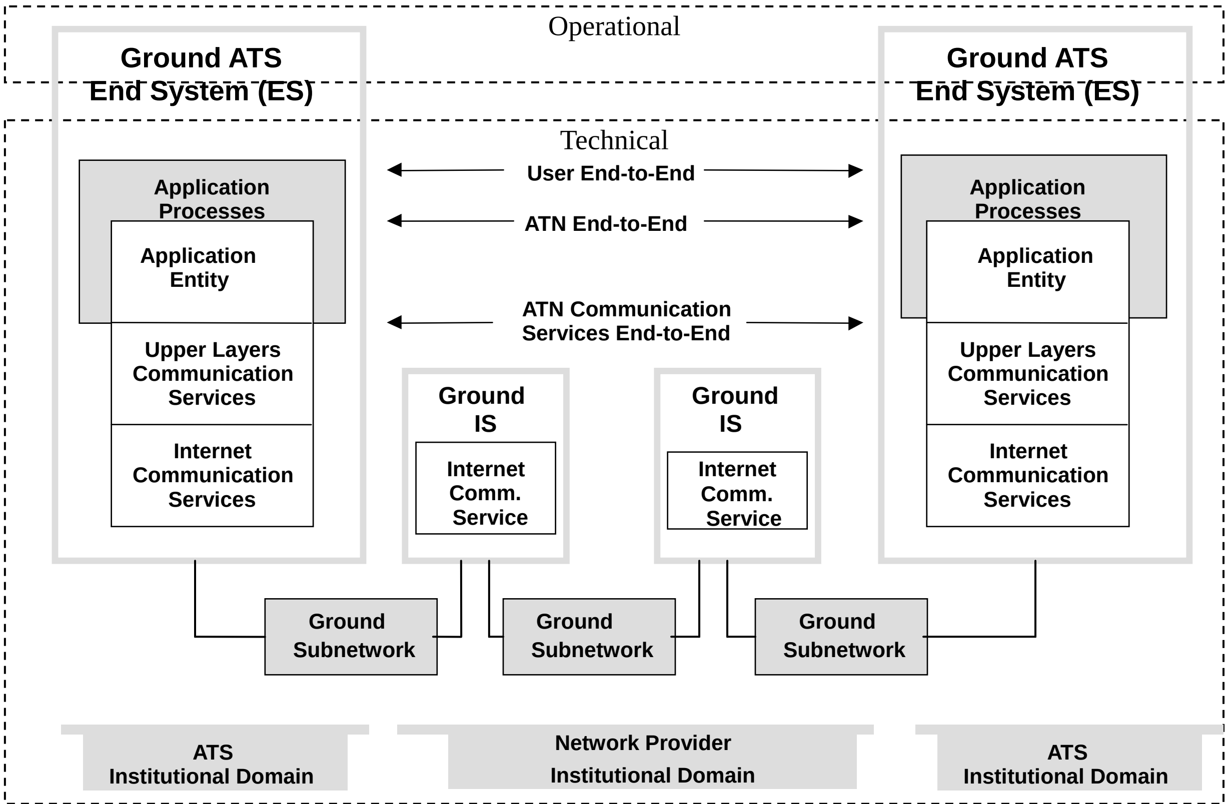

Figure 4.6 Technical Elemetnts of Ground-Ground Communication

4.3 Planning for Monitoring, Evaluating and Improving Performance

Once the Required Communication Performance is planned (i.e. derived and stated) and installed, the performance is monitored, evaluated and improved as necessary.

At the planning stage, the monitoring points, the measurement metrics and parameters are necessary to be identified.

4.3.1 Operational Performance Monitoring

As described above, there are two major components of operational performance, namely human factor performance and technical communication performance.

a) Monitoring Human Factor Performance

By monitoring operational performance and the technical performance, human factor performance is possibly being estimated.

b) Monitoring Technical Communication Performance
In the following, the Monitoring Technical Performance is discussed further.

4.3.2 Monitoring Technical Communication Performance

4.3.2.1 Performance Monitoring Points

Following the Guidance Materials of ATN System Management Services, the monitoring points, measurement metrics and parameters are listed below.

The following figure shows the possible monitoring points (a through e) in ATN Communication Services.

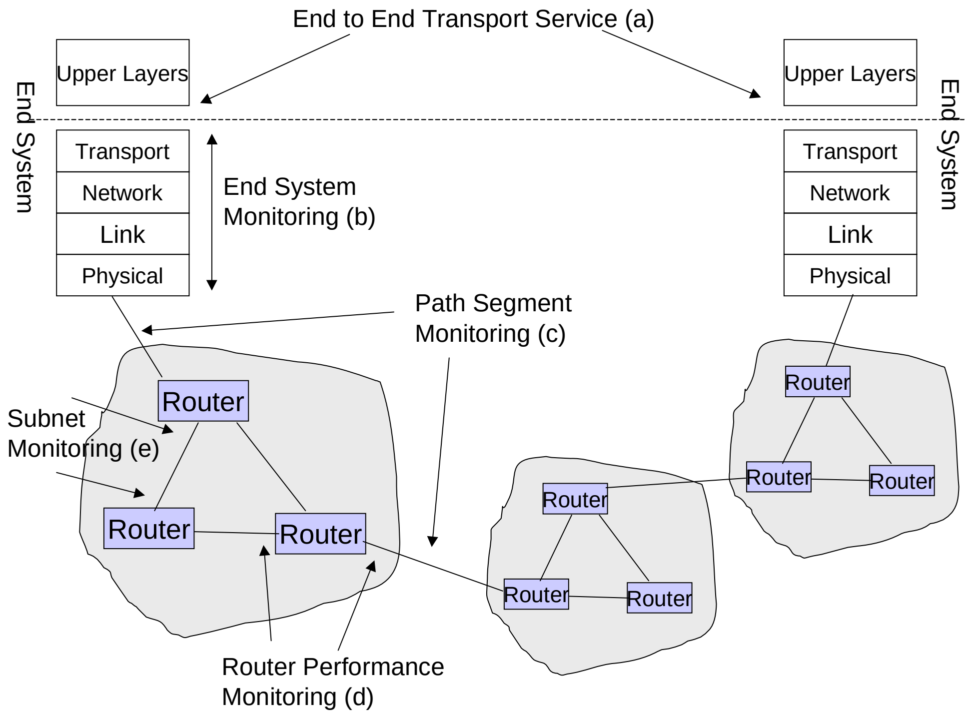

Figure 4.7 Performance Monitoring Points

Note 1; Two Performance metrics; Elapsed Time and Throughput:
There are two metrics of time related performance measurements; elapsed time and throughput.
The elapsed time is the time difference between the service requested and the service completed for a service requested. The throughput is the number of services processed during a certain time interval. The utilization of system providing services is closely related to the throughput.

The elapsed time is the metric related to the individual service request, while the throughput is the server wide metric of performance, counting the number of the completed services in a sampling duration, and the utilization is the rate of system busy during a sampling duration.

There is a close relationship between the elapsed time and the utilization of the server. If there is no other service request, i.e. zero utilization at the time, the service requested can be served immediately without any extra delay. The more services are requested, the more chances for waiting in the server queue, and the queue will be built up proportional to the rate of service requests. The higher utilization means the larger elapsed time, hence the larger delay on the average.

If D_0 denotes the delay when the network is idle, and U is a value between 0 and 1 that denotes the current utilization, the effective delay D is given by a simple formula;

$D = D_0 / (1-U)$

See, for instance D.E. Comer, 'Computer Networks and Internets', Third Edition, Prentice Hall, 2001
It implies that, if the system is experiencing the 50% utilization currently, the delay will be doubled on the average. If preoccupied with the utilization, the elapsed time will be sacrificed.

The things will be complicated further in a network with the multiple paths of services and the multiple stages of servers.

If the throughput or utilization is adopted as a performance metric, it is important to select an appropriate sampling duration of monitoring, for instance, the peak throughput or utilization is significantly different from the daily average throughput or utilization.

Note 2; Link Level Performance
The Data Link Level performance is the basic element of the network performance, but in a network like the ATN, the Data Link Level performance is less visible and possibly less relevant. The ultimate services are the End User Services, or the application level services. Also from the communication service provision viewpoint, it is important to manage the network and inter-network level performance. But the data link level within a sub-network is less visible, since the communication within a sub-network is dynamic and correlating the individual data link performance to the network level performance is a hard task, and some links of a communication path may be managed by different organizations, and it is again hard to collect the detailed performance information.

Another reason of the invisible link level performance is that the whole sub-network or communication path may be supported by some independent service providers, and the detailed data link level is again invisible from the outside.

The IDRP (Inter-Domain Routing Protocol) adopted by the ATN is based on the principle that each domain does not need to expose the routing details within a Domain to the outside. It is only necessary to show the Quality of Service Parameter values at the entry/exit points of the domain. Such a scheme makes the detailed Data Link Level Performance information invisible to the outside of the sub-network.

The monitoring points given in the Figure 4.7 can be re-organized to the following Figure 4.8 in terms of a hierarchy of services and communication elements.

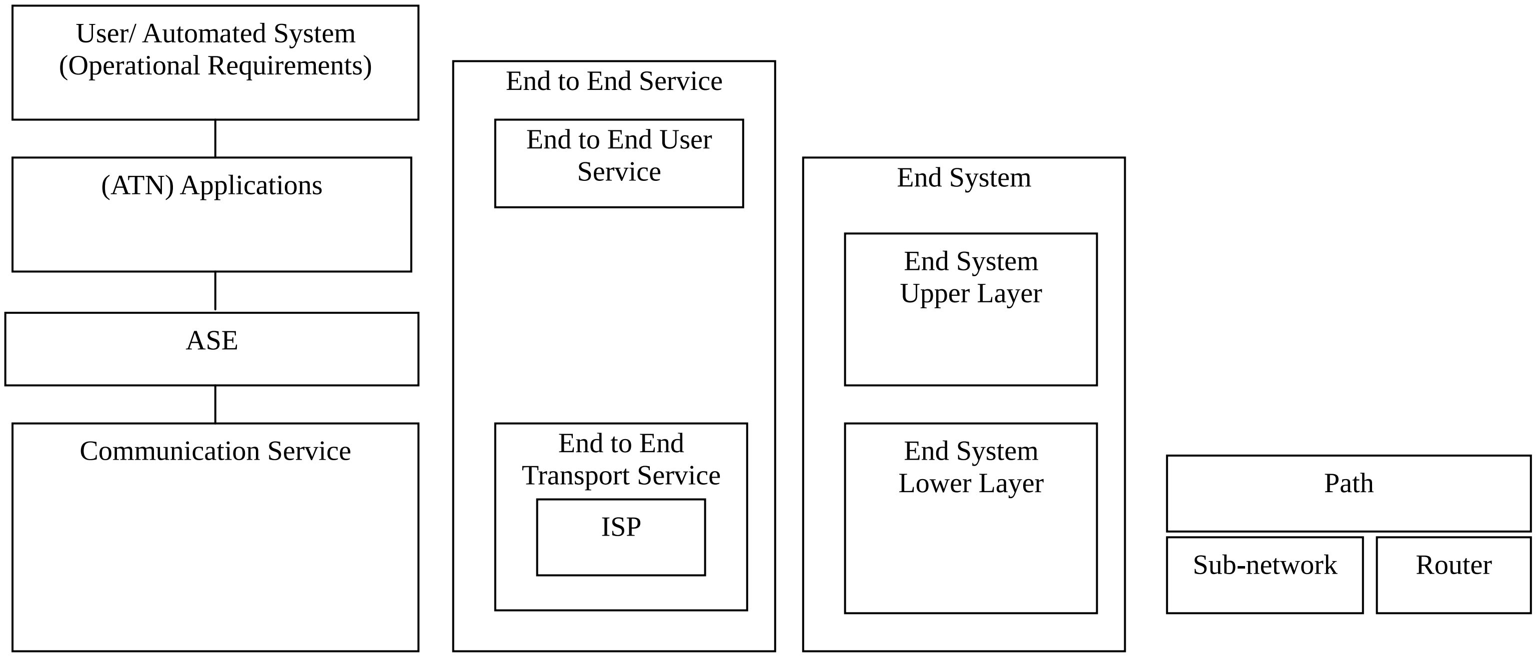

Figure 4.8 Performance Monitoring Points Diagram

The monitoring points then are listed as in the Table 4.2 below.

Table 4.2 Performance Monitoring Points

<table>
<tr><td></td><td>End to End Service</td><td>End System</td><td rowspan="4"></td></tr>
<tr><td>User/Automated System</td><td>End to End User Service (Operational Requirements)</td><td></td></tr>
<tr><td>(ATN)Application</td><td rowspan="2"></td><td rowspan="2">End System Upper Layers</td></tr>
<tr><td>ASE</td></tr>
<tr><td rowspan="2">Communication Service</td><td rowspan="2">End to End Transport Service (Service Level Agreement with ISP)</td><td rowspan="2">End System Lower Layers</td><td>Path Segment</td></tr>
<tr><td>Sub-network, Router</td></tr>
</table>

4.3.2.2 Performance Monitoring Parameters

Based on the Performance Monitoring Points listed above, the Performance Metrics for each Monitoring Points are listed as follows;

Table 4.3 Performance Metrics for each Monitoring Point Type

Monitoring Points		Performance Metrics
End to End Service		• Availability • Reliability • Continuity • End-to-End Transfer Delay • Integrity • Throughput • Connection Establishment Delay
Operational Requirements	General Requirements	(Air-Ground) • Probability of non-receipt of a message • Probability that non-receipt of a message will fail to be notified • Probability that a message will be misdirected
	Application Specific Requirements	See the following Table 4.4 as an example
	Transfer Delay	See the following Table 4.5 as an example.
End System(Lower and Upper Level)		• Internal Performance of Systems • Availability • Reliability • Continuity • Transfer Delay • Integrity • Throughput • Connection Establishment Delay
Path Segment		• Reliability • Transit Delay • Integrity • Throughput
Sub-network		• Availability • Reliability • Transit Delay • Continuity • Integrity • Throughput • Connection Establishment delay • Data loss rate
Router	Packet Forwarding	• Packets successfully forwarded • Percentage of packet discarded

	Routing Updates	• Availability • Reliability • Continuity • Time elapsed from a change in the network topology to this change being reflected in all affected routing tables
	Air-Ground Routers	• Number of Air-Ground virtual circuits supported at any time • Number of IDRP adjacencies
Service Level Agreement with ISP		• Reliability • Transit Delay • Throughput

Table 4.4 Application-specific Performance Requirements [ICAO DOC 9694]

Application	Availability	Integrity	Reliability	Continuity
CM / DLIC	99.9%	10^{-6}	99.9%	99.9%
ADS	99.996%	10^{-7}	99.996%	99.996%
CPDLC	99.99%	10^{-7}	99.99%	99.99%
D-FIS	99.9%	10^{-6}	99.9%	99.9%
AIDC	99.996%	10^{-7}	99.9%	99.9%

Table 4.5 Transfer Delay Performance Requirements [ICAO DOC 9739]

Performance Levels (ATSC Class)	Mean End-To-End Transfer Delay (sec)	95% End-To-End Transfer Delay (sec)	99.996% End-To-End Transfer Delay (sec)
A	0.5	0.7	1
B	1	1.5	2.5
C	2	2.5	3.5
D	3	5	8
E	5	8	12.5
F	10	15	22
G	12	20	31.5
H	15	30	51
I	30	55	90
J	60	110	180

It should be noted that the parameters in the Table 4.5 are not compatible to the RCP type parameters described above.

The following Table 4.6 shows the detailed parameters for the Performance Management Objectives identified.

Table 4.6 Metrics/Parameters for ATN Performance based on the ATN Performance Management Requirements

No	A/G G/G	Objectives	Metrics (Metrics in the previous table to be applied)	Detailed Measurement Parameters
1	A/G & G/G*	End User QoS Measurement Note: G/G* mainly interactive G/G applications	Availability	• Successful attempt to use the service • Failed attempt to use the service • Connection re-tried
			Reliability	• Number of messages sent to a given destination • Number of those received without errors by that destination • Residual error not detected by the destination(by manual analysis)
			Continuity	• Service interrupted • Corresponding resumption of service
			End to End Transfer Delay	• Transmission time and reception time of each message • Transport round trip delay
			Integrity	• Message delivered successfully • Message delivered with errors not detected • Message with errors in fact message is correct • Probability of errors being mis-detected • Message rejected
			Throughput	• Number of messages sent(size and time)
			Connection Establishment Delay	• Time to establish a transport connection • Time to establish a connection using Dialogue Service • Time to establish an application specific relationship

2	A/G & G/G	End System(Upper & Lower Layer) Monitoring Note: There are four types of ES 1) A/G(Air)ES 2) A/G(Ground)ES 3) G/G(Interactive) ES 4) G/G(S&F)ES	Internal Performance of System	• Number and size distribution (mean and max) of packets sent and received at the internet level, for each priority and ATSC class, and possibly ISP(Internet Service Provider). • Transit delay of a packet through the End System
			Availability	• Successful attempt to use the service • Failed attempt to use the service • Connection re-tried
			Reliability	• Number of messages sent to a given destination • Number of those received without errors by that destination • Residual error not detected by the destination(by manual analysis)
			Continuity	• Service interrupted • Corresponding resumption of service
			Transfer Delay	• Relay time of each message(G/G-S&F-ES)
			Integrity	• Message delivered successfully • Message delivered with errors not detected • Message with errors in fact message is correct • Probability of errors being mis-detected • Message rejected
			Throughput	• Number of messages received/sent(size and time)
			Number of Active Connections	• Number of Active Logical / Physical Connections (A/G-Ground-ES and G/G-Interactive-ES)
			Storage Capacity(G/G-S&F-ES)	• Number of messages stored

3	A/G & G/G	Sub-network Service Monitoring	Availability	• Successful attempt to use the service • Failed attempt to use the service
			Reliability	• Number of messages sent • Number of messages received
			Transit Delay	• (Connection Oriented mode) Can use acknowledgement mechanism; it is not suitable for end-to-end sub-network and real time. • (Connectionless mode)Using special link level test packet with time-stamp or using CLNP echo for the round trip delay; it needs a special care to segregate the sub-network from multiple hops and to segregate the internal processing time in the system(Router/ End System);
			Continuity (depends on sub-network type)	• Service Interruption o (Connection oriented mode) un-commanded loss of a sub-network connection or o inability of the sub-network to accept data for sending on the established connection o (Connectionless mode) transmission failure • Resumption of Service o (Connection Oriented mode) Successful re-establishment of connection o (Connectionless mode) The time of the next successful transmission attempt

			Integrity	• (Connectionless mode) CLNP header checksum; it is not mandatory in the ATN SARPs for End System to generate. • Deflate checksum (whole packet checksum); it is optional. • Number of packets sent over the sub-network • Number of packets received over the sub-network • Number of packets rejected by the sub-network • Number of packets received with a failed CLNP header checksum or Deflate checksum failure
			Throughput	• Number of messages sent (size and time) over each communication path; offline analyses required.
			Connection Establishment Delay	(Connection Oriented mode only) • The time of each sub-network connection initiated • The time of successful connection establishment
4	A/G & G/G	Measurement of Path Segment performance	Reliability	• Number of packets on entry and exit
			Throughput	• Number of packets on entry and exit
			Transit delay	• Each individual packet on entry and exit, or • Sample packet on the link or • Periodic echo packet on the data flow
5	A/G & G/G	Monitoring for Excess Capacity	Throughput	• Peak throughput
6	A/G & G/G	Planning Future Capacity		• Network Design Models • Data volumes for each sub-network on a point to point basis and during each sample period • Data volumes(both bytes and packets) handled by each Router during each sample period
7	A/G & G/G	Dynamic (Sub-network) Capacity Management	Indicating the need for Additional Capacity	• Queuing delay for packet transmission; average queue length during a sample period(or absolute queue delay per packet)

			Indicating When Additional Capacity is no longer Needed	• Queuing delay for packet transmission; average queue length during a sample period(or absolute queue delay per packet)
			Adding and Removing Additional Sub-network Capacity	• It needs System Management Action
8	A/G & G/G	Router Forwarding Measurement	Integrity	• Packets successfully forwarded • Percentage of packet discarded • Reasons packets discarded
9	A/G & G/G	Router Key Parameters measurement	Internal Performance of System	• System parameters(e.g. memory utilization) • The number of entries in the Forwarding Information Base(FIB)
10	A/G & G/G	Measurement Route Convergence Times	Internal Performance of System	• Routing events • Clock synchronization accuracy
11	A/G	Air-Ground Router Overhead Measurement		• Establishment and termination of adjacencies with Airborne Router • Establishment and termination of sub-network connection with Airborne Router • Maximum number of adjacencies and sub-network connections
12	A/G & G/G	Analysis of Data Stream Performance (depends on the Service Level Agreements with ISP(Internet Service Providers)	Reliability	• Number of packets on entry and exit
			Throughput	• Number of packets on entry and exit
			Transit delay	• Each individual packet on entry and exit, or • Sample packet on the link or • Periodic echo packet on the data flow
13	A/G & G/G	Monitoring of Service Provider Compliance by Aggregate Data Flow (depends on the Service Level Agreements with ISP(Internet Service Providers)	Reliability	• Same as others but the aggregated values
			Throughput	• Same as others but the aggregated values
			Transit delay	• Same as others but the aggregated values

14	A/G & G/G	Monitoring of User Message Delivery	Integrity	• Mis-delivered user message • Not delivered to the intended recipient(lost or mis-delivered) with notification to the originator • Not delivered to the intended recipient(lost or mis-delivered) with no notification to the originator
15	A/G & G/G	Monitoring of User Message transfer Times (depends on the types of applications)	Transfer Delay	(Conformed Service, e.g. LACK in CPDLC) • Round trip time based on the acknowledge procedure of applications; it needs to factor out the processing time in the system. (Unconfirmed Services) • Measurement based on timestamp, clock synchronization, or system management echo function. (For both cases) • Factor out human response time, system processing time.

Note; It should be noted that the monitoring points/parameters listed above are not guaranteed to be technically or practically feasible. For the details, consult to the original document; ICAO Doc 9739, Part V, section 2.2.2

5. Summary

The current document provides the guidelines for the performance related considerations. In the document, preliminary considerations on the performance, the RCP developed by the OPLINKP, the ATN performance and monitoring its performance are described.

Regarding the definition of communication performance, the OPLINK Panel is finishing to develop the ICAO Manual on the RCP. The ICAO RCP manual will provide the detailed descriptions and explanations related to the Operational Communication Performance Requirements: RCP.

In future, the specific RCP parameter values (or the ranges of values) for a specified ATS functions and the airspaces may be given, which currently are partially determined.

Appendix

ICAO DOC 9739 (CAMAL) Edition 2 Part V Chapter 2 ATN System management Services 2.2.2 Performance Management Requirements, p.p. V-2-35-67

Performance Management Requirements

Note 1. This chapter is concerned with the derivation of Performance Management requirements for the ATN Internet and Upper Layers, including applications. A top down analysis is presented, first identifying the objectives for Performance Management and then going on to look at how these objectives are met, from which system and tool requirements can be derived.

Note 2. It should be noted that this analysis does not attempt to determine how difficult or costly each identified requirement is to implement.

Derivation of Performance Management Objectives

1.1 ISO/ITU-T standards define performance management as follows:
Performance management enables the behavior of resources in the OSI Environment and the effectiveness of communication activities to be evaluated. Performance management includes functions to:

1. gather statistical information;
2. maintain and examine logs of system state histories
3. determine system performance under natural and artificial conditions; and
4. alter system modes of operation for the purpose of conducting performance management activities.

1.2 In an ATN context, there are three basic purposes for which the above procedures will be conducted:

a) The end-to-end performance will need to be measured in order to ensure that the operational requirements are being met.

b) Users will wish to monitor the performance of ATN Internet Service Providers (ISPs) and subnetwork providers, and, in particular, to assess their compliance with service level agreements (this could be for both Operational and Financial reasons).

c) ISPs will wish to monitor the performance of the network, to ensure that they are maintaining service level agreements, to ensure that the network capacity will match future requirements (Capacity Planning), and to ensure that the service that their users require is being delivered, and in the most cost effective manner.

1.3 It should be noted that the end-to-end service may be provided by multiple ISPs operating in serial.

1.4 In order to determine the Performance Management objectives, it is thus necessary to look at each of the above purposes, to identify the performance monitoring points, and to consider the different relationships between ISPs. In particular, service agreements will relate to path segments though the ATN Internet and a given path segment may be maintained by more than one ISP in series. The performance management models that result from this scenario need also to be considered.

1.5 Performance Monitoring Points

1.5.1 Figure 2.2-1 illustrates the various performance monitoring points (PMPs) that are required in the ATN Internet Communication Service (ICS). These are:

a) **The End to End Transport Service** measures the end-to-end service provided to a TS-User. Performance monitoring is required here in order to measure the overall performance characteristics of the ATN ICS.

b) **End System Monitoring (Lower Layers)** is required to determine the part of the end to end ICS overhead attributable to the End System itself. This should be a small component of the overall end-to-end figure, but this still needs to be demonstrated.

c) **Path Segment Monitoring** is required to monitor the performance of ISPs and to measure the contribution to the end-to-end overhead of the routers and subnetworks operated by an End User.

d) **Router Performance Monitoring** is required to measure the performance of individual routers and to determine their contribution to the end-to-end overhead.

e) **Subnetwork Performance Monitoring** is required to monitor the compliance of subnetwork providers with performance level agreements and to measure the contribution of each subnetwork to the end-to-end overhead.

1.5.2 In addition, PMPs are required in the upper layers of the ATN, which includes the application ASEs and ASE-Users:

a) **End System Monitoring (Upper Layers)** is required to determine the part of the end to end overhead attributable to that part of the system which uses the ICS. Note that some aspects of upper-layer performance characteristics may vary dynamically depending upon the particular implementation architecture. For example, a distributed architecture may at times be subject to LAN congestion delays in the local application system.

b) **The End to End ASE Service** measures the end-to-end service provided to application software which uses the communication services of the complete ATN stack. Performance monitoring is required here in order to measure the technical performance of the standardised application protocols.

c) **The End to End User Service** measures the end-to-end service as perceived by the human user, or in the case of automated functions, at the highest level of the automated processing thread. Performance monitoring is required here in order to verify that the end to end service meets the operational requirements.

1.6 End to End Service Monitoring

1.6.2 An end user of the ATN (typically a CAA or Airline) is interested in ensuring that the required end-to-end Quality of Service is being maintained. At the same time, they will also have an Accounting Management objective to monitor the cost of using the ATN, and, whilst this is a different subject, the data capture requirements will be similar.

1.6.3 The end-to-end quality of service (QoS) can be broken down into the following parameters:

a) **Availability**: "The ability of a system to perform its required function at the initiation of the intended operation. It is quantified as the proportion of the time the system is available to the time the system is planned to be available".

b) **Reliability**: "The probability that the system will deliver a particular message without errors."

c) **Continuity**: "The probability of a system to perform its required function without unscheduled interruptions during the intended period of operations."

d) **End-to-End Transfer Delay**: "The period elapsed from the time at which the originating user initiates the triggering event until the time the transmitted information has been received by the intended recipient."

e) **Integrity**: "The probability that errors will be mis-detected. This may be when a correct message is indicated as containing one or more errors, or when a message containing one or more errors is indicated as being correct."

f) **Throughput**: "The quantity of data (e.g. measured in characters) that can be sent during a given period."

g) **Connection Establishment Delay**: "The time from initiating an end-to-end connection to its successful establishment (i.e. when messages can be sent).

Note. The definitions of Availability, Reliability, Continuity, Transit Delay and Integrity are taken from the ADSP Manual [8].

1.6.4 The operational requirements for performance management require the ability to measure each of the above Quality of Service metrics on an end-to-end basis i.e. between an application user and each other end user of the ATN with which it communicates. [Objective 1]

1.7 End System Monitoring

1.7.3 End Users typically own and operate the End Systems that form the first part of the chain of end to end communications. They will need to monitor the internal performance of these systems. [Objective 2]

1.7.4 This is so that the actual performance provided by the ISP(s) used for end-to-end communications can be calculated, and to enable performance trends in the local system to be monitored for capacity planning purposes.

1.8 Subnetwork Service Monitoring

Note. An ATN ISP or end user may also operate subnetworks and hence wish to monitor their internal operation. However, the internal monitoring of subnetworks is outside of the scope of this analysis.

1.8.3 An ISP or End User will need to monitor the operation of each subnetwork they use in order to ensure that the expected/required Quality of Service is being maintained. [Objective 3]

1.8.4 The Quality of Service metrics to be measured are very similar to those for the end-to-end service and will always include: Availability, Reliability, Continuity, Transit Delay, Integrity and Throughput. In addition, there will be subnetwork specific metrics. For example. a connection mode subnetwork (e.g. X.25) will have a connection set-up delay metric, while a Frame Relay circuit has only a maximum guaranteed throughput, with the possibility of data loss, and hence the data loss rate will need to be monitored and compared with offered traffic load.

1.9 Path Segment Monitoring

1.9.3 An ISP or End User will need to monitor the overall performance of each path through their segment of the ATN Internet [Objective 4]. The metrics measured will be those specific to a connectionless service i.e. Throughput, Reliability, Integrity and Transit Delay.

1.9.2 Performance achievement will also need to be measured against the theoretical maximum so that excess capacity can be removed [Objective 5]. Trend analysis will also need to be performed so that future growth can be predicted and hence future capacity growth planned and installed when needed [Objective 6].

1.9.3 ISPs and End Users may also wish to manage their capacity dynamically. This make take the form of bringing on additional capacity to meet busy hour requirements, according to a defined schedule, or simply be reacting to reports of congestion by opening up new circuits. [Objective 7].

1.10 TN Router Performance Monitoring

1.10.3 General ATN Router performance may be considered in two separate parts: forwarding and route updates. There are also special considerations for Air/Ground Router performance.

1.11 Packet Forwarding

1.11.1 An ISP or End User will need to monitor the following packet forwarding metrics [Objective 8]:

a) Packets successfully forwarded during a given measuring period (e.g. per second)

b) Percentage of packets discarded.

1.11.3 As the service level requirements can vary by both ATSC Class and priority, these metrics will need to be broken down by both of these parameters.

1.11.4 Key parameters of the router that affect forwarding need also to be measured. These include the size of the Forwarding Information Base (FIB). [Objective 9].

1.12 Routing Updates

1.12.1 An ISP or End User will need to monitor the route convergence rate. That is the time elapsed from a change in the network topology to this change being reflected in all affected routing tables. [Objective 10].

1.11.2 This metric is important in the maintenance of Availability, Reliability and Continuity targets as packet loss can occur during this period. The target convergence rate may be derived empirically or from simulation models.

1.13 Air/Ground Routers

1.13.1 Ground/Ground Routers typically support a limited number of fixed adjacencies. However, Air/Ground Routers will support a large and variable number of adjacencies with aircraft, with each adjacency supported by one or more virtual circuits. The performance of the router may be affected by the number of such adjacencies maintained, and performance will thus need to be measured against the adjacencies and virtual circuits supported. It will thus be necessary to monitor the number of air/ground virtual circuits supported at any one time, and the number of IDRP adjacencies. [Objective 11].

1.14 Service Level Management

1.14.1 Service guarantees will need to be put in place to ensure that the end-to-end Quality of Service required to meet the operational requirements is achieved. When this service is provided via multiple ISPs in serial, there are two possible models for maintenance of the end-to-end Quality of Service:

a) The end user(s) have separate service level agreements with each ISP that supports (in serial) the end-to-end path.

b) The end user(s) have a single service agreement with one ISP who is also responsible for the service provided by any other ISP en route.

1.15 Separate Service Level Agreements

1.15.1 Whilst this model is, on first appearances, easier for the ISPs to operate, it pushes the complexity onto the end users, who have to be aware of the paths that their data takes, and who are required to negotiate with each ISP en route. This potentially makes for many small contracts between users and ISPs. It may thus be more costly to implement because of this contractual overhead.

1.15.2 In this model, each ISP will need to be able to demonstrate to an End User that they are meeting the service requirement for the path segment under their control. It will thus be necessary to monitor incoming packets from either users or other ISPs and analyse them by sending end user. Similarly, the ISP will need to analyse traffic at each exit point. [Objective 12].

1.15.3 Analysis of traffic by source is potentially a very costly exercise as each packet has to be metered.

1.16 Single User Service Agreements

1.16.1 This model can be operated in two different ways, which can be labeled as "microscopic" or "macroscopic". In the microscopic view, one ISP simply becomes the "agent" of an end user, and organizes the separate contracts with each other ISP en route. In order to monitor compliance, each data flow will still need to be identified as above, and the end user's agent will collate the information together and monitor the overall performance, as above.

1.16.2 In the macroscopic view, an ISP will aggregate together the traffic from all of its users and thence identify the traffic volumes it will exchange with each other ISP. On the basis of agreed traffic volumes, the ISPs may then guarantee minimum service levels to each other. A given ISP may then offer a service guarantee to its end user, and base this guarantee on the knowledge of which ISPs lie on the route and by summing the guaranteed minimum service levels.

1.16.3 The macroscopic view is simpler to achieve because it requires only that each ISP monitors its own part of the ATN Internet (as in 2.2.2.1.9, Path Segment Monitoring above). However, it will probably need to be coordinated through an industry forum that will also receive, collate and monitor the overall service provided to end users. [Objective 13].

1.17 Operational Requirements

1.17.1 Monitoring the end-to-end performance of the ATN service provided to application users is a stated requirement. The objective is to enable evaluation of the operational effectiveness of the communication resources by means of statistical information and logs of system state histories.

1.17.2 The workload of the applications can be measured, globally or on a connection basis, by the accounting management parameters.

1.17.3 The ADSP Manual defines in Part I, Chapter 3 Appendix A three types of communication systems performance requirements related to the ATS data link applications:

a) General Performance Requirements

b) Application Specific Performance Requirements

c) Transfer Delay Requirements

1.18 General Performance Requirements

1.18.1 The general performance requirements on the air-ground applications are:

a) the probability of non-receipt of a message will be equal to or less than 10^{-6},

b) the probability that non-receipt of a message will fail to be notified to the originator will be equal to or less than 10^{-9}, and

c) the probability that a message will be misdirected will be equal to or less than 10^{-7}.

1.18.2 These performance requirements can be designed into a system, and bench-tested before operational use. However, there is still a need to monitor whether the requirements are being met during operational use. [Objective 14].

1.19 Application Specific Performance Requirements

1.19.1 The performance requirements specific to each air-ground application are also defined in the ADSP Manual [8], as shown in Table 2.2-3.

Table 2.5-3. Application-specific Performance Requirements

Application	Availability	Integrity	Reliability	Continuity
CM / DLIC	99.9%	10^{-6}	99.9%	99.9%
ADS	99.996%	10^{-7}	99.996%	99.996%
CPDLC	99.99%	10^{-7}	99.99%	99.99%
D-FIS	99.9%	10^{-6}	99.9%	99.9%
AIDC	99.996%	10^{-7}	99.9%	99.9%
ADS-B	99.996%	10^{-7}	99.996%	99.996%

1.19.2 It is necessary to make assumptions as to what period of time these statistical figures are to be measured over. For example, if averaged over a week then the achieved performance in one particular minute could have been well outside of these requirements.

1.19.3 Again, there is a need to monitor whether these requirements are being met during operational use of the ATN by applications. This objective is covered by 2.2.2.1.6, End to End Service Monitoring above.

1.20 Transfer Delay Requirements

1.20.1 The performance available to each air-ground application is also defined in the ADSP Manual, as shown in Table 2.2-4.

Table 2.5-4. Transfer Delay Performance Requirements

Performance Levels (ATSC Class)	Mean End-To-End Transfer Delay (sec)	95% End-To-End Transfer Delay (sec)	99.996% End-To-End Transfer Delay (sec)
A	0.5	0.7	1
B	1	1.5	2.5
C	2	2.5	3.5
D	3	5	8
E	5	8	12.5
F	10	15	22
G	12	20	31.5
H	15	30	51
I	30	55	90
J	60	110	180

1.20.2 The requested ATSC class (A through H) can be selected by the application user on dialogue initiation, and is not known in advance.

1.20.3 These transfer delay figures are intended to represent the total transfer time from human user input at one side to human user perception at the other end. As the actual transit delay is indirectly dependant on the requested class of communication service, this parameter should be made available with the transit delay measurements. [Objective 15].

1.21 Derivation of QoS Characteristics

1.21.1 It is not a straightforward matter to translate from the operational performance requirements described above into the QoS characteristics of a communications subnetwork. In fact it would be misleading to attempt to do so. The reason is that the end-to-end delay has to be distributed between all the elements (human, automation and communications) involved in the information exchange. There are several discrete entities which require a performance "budget" to be allocated to them, within the overall performance constraints. The following illustration relates to an air-to-ground message:

a) Human User (Air) - thinking time and data input budget
b) Avionics automation (processing) and local data transfer (between equipment) budget
c) Airborne communications processing budget (end system)
d) Air/ground subnetwork budget
e) Ground/ground subnetwork budget
f) Ground communications processing budget (end system)
g) Ground Automation (Data Processing) budget (e.g. FDPS)
h) Local distribution (communications) budget (FDPS to controller positions)
i) Local display automation (processing) budget
j) Human User (Controller) budget

1.21.2 The budget allocation for ground to air messages can be similarly expressed.

1.21.3 When it comes to matching responsibilities for achieving the required performance to the decomposition, the full complexity becomes apparent:

a) Human factors influence the rate at which the users can input or access messages at the HMIs

b) The aircraft manufacturer, airline and avionics industry all play a part in constraining the performance capabilities of the avionics sub-systems and also the on-board communications processors and sub-network interfaces (radios)

c) Theoretical air-ground sub-network performance is governed by the technical solutions (standards) that are invoked, but practical performance is also influenced by congestion and interference levels in a particular environment

d) Theoretical ground sub-network performance is influenced by technical solutions and network design (topology and sizing), but practical performance is also influenced by congestion and network component availability

e) Service providers may be involved in the ground communications processing

f) ATSOs own the FDPS and will therefore influence the effectiveness with which the automation processes data messages to the correct local destination

g) ATSOs also own the internal centre architecture, and specify local communications bandwidths to controller positions

h) Control position design influences the time taken for a message to be presented, and human factors influence the time taken for the recipient to recognise, understand and action the message

1.21.4 As the overall communications characteristics are the result of the aggregation of (possibly several) subnetwork performance and processing performance, it may be useful to develop the concept of "Achievable Communications Performance" (ACP), based on these parameters and calculated at the sub-network level for different traffic and environmental scenarios. The ACP for any given component (e.g. the air-ground subnetwork) can then be compared with the "budget" for that component, to determine whether overall performance target is likely to be met.

1.21.5 During such an analysis, it would also be possible to "trade" budgets between components (use a higher performance but higher cost ground sub-network, in order to accommodate a lower performance, but cheaper, air-ground sub-network.) This kind of analysis constitutes "End to End systems engineering".

2. Fulfillment of Objectives

2.1 As presented below, there is a general requirement to log events locally for later, offline, analysis [Req 1]. Unless stated otherwise, all requirements to log or record events given below imply a local log, rather than a standard SM log that is visible to an SM Agent.

2.2 *Objective 1: End User QoS Measurement*

2.2.1 The end user will need to be able to monitor each of the QoS metrics identified in 2.2.2.1.6, End to End Service Monitoring above.

2.2.2 Availability

2.2.3 In order to determine the percentage availability according to the above definition, the end user will need to record each successful attempt to use the service [Req 2] and each failed attempt [Req 3]. Analysis of such a record can then determine the percentage availability achieved over a given time period. [Req 4]

2.2.4 It is not always straightforward to identify a failure of a connection oriented service, as there may be instances where the connection responder deliberately rejects the connection attempt as part of the normal user protocol operation. Also, the connection may be re-tried by intermediate protocol layers that are not visible to the end-user.

2.2.5 Reliability

2.2.6 In order to determine the probability that a message is delivered without errors it will be necessary to record both the number of messages sent to a given destination [Req 5] and those received without errors by that destination, analyzed by sender [Req 6]. Comparison of these two records may then determine the percentage reliability. [Req 7]

2.2.7 There may still be residual errors which are not detected by the destination system. The only way to detect such errors would be an offline comparison of the actual messages logged by the sending system and by the receiving system. This could be undertaken on a statistical sampling basis, to avoid the otherwise prohibitive overhead of having to replay every single application message.

2.2.8 Continuity

2.2.9 In order to determine the continuity, it will be necessary to record each service interruption and each corresponding resumption of service [Req 8]. For the connection-oriented end-to-end service, a service interruption event can be equated to the uncommanded loss of a transport connection, and service resumption to the successful re-establishment of such a transport connection. Analysis of this log can determine the continuity level achieved. [Req 9]

2.2.10 End-to-end Transit Delay

2.2.11 Note that this can vary according to message priority, and ATSC Class.

2.2.12 End-to-end transit delay could be measured by recording the transmission time of each message, and the reception time of each message, at each end of the connection. Comparison of these two records can then determine the actual transit delay provided that both end users have synchronized clocks.

2.2.13 Synchronised clocks can be an impractical requirement, although time synchronization to some degree of accuracy for ATN Routers seems an inevitable requirement (see Objective 10: Measurement of Route Convergence Times). Logging the transmission of each message is expensive and will result in large log files. Some applications have timestamps built-in to certain protocol elements. However, the semantics of the timestamp information are ill-defined and application-specific. Fortunately, there is a more readily available metric from which end-to-end ICS transit delay can be estimated; this is the transport round trip delay.

2.2.14 The connection mode transport protocol (TP4) can readily measure the round trip delay from transmission of a data TPDU to reception of the corresponding AK TPDU. Simply dividing this figure in two gives an estimate for the transit delay, assuming that the transit delay is the same on average in both directions.

2.2.15 In practice, a TS user message (TSDU) may be split into several TPDUs, and transmitted with overlapping acknowledgements. However, the transit delay for the total message can still be estimated by measuring the delay from the transmission of the first TPDU of the TSDU to the reception of the AK TPDU for the last TPDU making up the TSDU, and then subtracting from this the estimated transit delay for a single TPDU. The result is the estimated transit delay for the whole message.

2.2.16 This is an estimate, and each estimation is subject to an error. However, the errors should balance out over several messages, although the potential for systematic errors due to asymmetrical paths in each direction does exist. The transport protocol should record the estimated transit delay for each message, transmission time and the size of each message [Req 10]. Later analysis of the record can then provide an accurate estimate of the achieved end-to-end transit delay, analyzed by message size, and transmission time [Req 11].

2.2.17 Integrity

2.2.18 A high degree of message integrity is built-in to the ATN by design, e.g. by employing the error detection capabilities of the TP4 protocol. However, it is necessary to monitor the probability of errors being mis-detected in order to verify that the theoretical integrity performance targets are in fact being maintained.

2.2.19 By definition, the integrity of the ATN ICS cannot be measured by the ATN Internet because to measure it implies that the Internet can detect its own mistakes - in which case it could then correct them. Integrity has to be measured by the end user either by comparing messages sent and received at each end of the communications path, or by adding an additional message integrity check on transmission, and recording, on reception, the number of messages received correctly and the number received with errors. In the former case, analysis requires access to logs at both end of the communications path while, in the latter case, access is required to only one log. Routine offline comparison of sent and received message logs can thus be used to provide the integrity metric. [Req 12]

2.2.20 Note that the integrity measurement includes not only messages delivered with errors which are not detected at the time of processing, but also error indications which are given when the message was in fact correct (e.g. failure of the checksum algorithm). Thus it is necessary to log not only the content of successful messages, but also (a statistical sample of) rejected messages.

2.2.21 In addition, security provisions are expected to require the use of digital signatures for authentication purposes. A digital signature provides a high quality integrity check. In practice, a likely reason for authentication failure is an undetected network error. When such security mechanisms are implemented, authentication failures should be logged as well as numbers of messages successfully received [Req 13]. The integrity achieved can then be computed from such a record - assuming that real security violations can be filtered out [Req 14].

2.2.22 Throughput

2.2.23 Throughput achieved can be simply measured provided that a log of each message sent (size and time of transmission) over each communications path is maintained [Req 15]. Throughput achieved can then be determined by analysis of the log [Req 16].

2.2.24 Connection Establishment Delay

2.2.25 In the ATN, there are at least three possible measures of connection establishment delay. One is the time taken to establish a transport connection, another is the time taken to establish a connection using the Dialogue Service, and a third is the time taken to establish an application-specific relationship (e.g. to complete the CPDLC-Start service). The latter may be significantly longer than the former, as it usually requires an exchange of messages after the transport connection has been established - although this can be avoided in some cases.

2.2.26 The time taken to establish a transport connection is the proper measure for the performance of the ATN Internet. However, the end user only sees dialogue service connection establishment time. On the other hand, the problem of using this as a metric is that it includes the response time of the remote application. It therefore is not a correct estimate of the performance of the ATN Internet.

2.2.27 Arguably, both metrics need to be recorded and analyzed. Therefore, an ATN End System will need to log both dialogue service D-STARTs [Req 17] and transport connection connect requests [Req 18], and the time at which the request was made. Similarly, an ATN End System needs to log the time at which the connection was successfully established [Req 19]. Later analysis can then determine the mean for both logs [Req 20].

2.2.28 Note that some application services will establish a transport connection, but not a Dialogue, by sending an application response via a negative D-START response primitive. Thus, apparently unsuccessful Dialogue establishments must also be logged.

2.3 *Objective 2: End System (Lower Layers) Monitoring*

2.3.1 In order to monitor End System performance and to quantify the performance of ISPs, the operator of an End System will need to know:

a) At the Internet level, the number and size distribution (mean and max) of packets sent and received, analysed by priority and ATSC Class [Req 21]. When the End System is connected to multiple ISPs, the above will need to be broken down by ISP as well.

b) The transit delay of a packet through the End System (incoming and outgoing) [Req 22]. Note that this need not be measured all the time, but could be measured during system testing and then taken as a system characteristic.

2.3.2 This information will enable the computation of the traffic load on the End System and the load applied to each ISP [Req 23]. The per packet transit delay needs to be subtracted from the end-to-end transit delay in order to compute the ISP transit delay.

2.4 *Objective 3: Subnetwork Service Monitoring*

2.4.1 It is not possible to deal with all aspects of subnetwork monitoring without considering each type of subnetwork. This analysis seeks only to establish the generic requirements. Specific subnetwork monitoring requirements will need to be determined on a case by case basis.

2.4.2 Availability

2.4.3 In order to determine the percentage availability of a subnetwork, the subnetwork user will need to record each successful attempt to use the service and each failed attempt [Req 24]. Analysis of such a record can then determine the percentage availability [Req 25].

2.4.4 It is not always straightforward to identify a failure of a connection oriented subnetwork, as there may be instances where the connection responder deliberately rejects the connection attempt as part of the normal user protocol operation.

2.4.5 Reliability

2.4.6 In order to determine the probability that a packet is delivered without errors it will be necessary to record both the number of packets sent and received at both ends of a subnetwork [Req 26]. Comparison of these two records may then determine the percentage reliability [Req 27].

2.4.7 Continuity

2.4.8 In order to determine the continuity, it will be necessary to record each service interruption and each corresponding resumption of service [Req 28]. For connection mode subnetworks, a service interruption event can be equated to the uncommanded loss of a subnetwork connection, or the inability of the subnetwork to accept data for sending on an established connection. Service resumption is then the successful re-establishment of the connection and the resumption of data transfer. For connectionless subnetworks a service interruption can be equated to a transmission failure, with service resumption being the time of the next successful transmission attempt. (Note that on some connectionless subnetworks, it may not be possible to detect transmission failure).

2.4.9 Analysis of the logs can then determine the number of interrupted and uninterrupted service invocations and hence the continuity of service provided by the subnetwork [Req 29].

2.4.10 End-to-end Transit Delay

2.4.11 There is no general purpose mechanism to measure transit delay over subnetworks. A connectionless network, such as an Ethernet, has no acknowledgement mechanism that would allow round-trip times to be measured. In such cases, transit delay can only be measured by using special link level test packets to convey a time-stamp (assuming clock synchronization) or to "echo back" a response and thereby derive the transit delay from the round trip time.

2.4.12 On the other hand, a connection mode subnetwork, such as X.25, does provide an acknowledgement mechanism, but this is not necessarily end-to-end across the subnetwork and has no real-time constraints. It is thus unsuitable for measurement of transit delay.

2.4.13 The CLNP Echo is potentially available to provide an estimate of round trip delay between a pair of Routers or a Router and an End System attached to the same subnetwork. However, the processing delay within the target system is potentially significant and will need to be independently measured and factored out. Furthermore, CLNP echo is of limited value when more than one subnetwork joins the same pair of systems, as it is not possible to guarantee which subnetwork is used to transfer the Echo packet.

2.4.14 A strategy for measurement of transit delay will need to be developed for each type of subnetwork. For connectionless subnetworks, or wide area networks involving multiple hops, this will need to be based on some kind of subnetwork specific echo packet. For single hop connection mode networks this may use the subnetwork's own acknowledgement procedures.

2.4.15 For example, in VDL Mode 2, the air/ground data link uses a connection mode variant of HDLC (the AVLC) which could be used to measure round trip time and hence to estimate transit delay.

2.4.16 Integrity

2.4.17 Data integrity problems can be measured by two ATN Internet mechanisms:

a) The CLNP Header Checksum. This only covers the CLNP header, and its generation by a sending End System is not mandatory in the ATN SARPs. It is therefore of limited value.

b) The Deflate checksum. This provides an integrity check over the entire packet and is therefore a very good check on integrity. However it is only available when the optional Deflate-based compression is used.

2.4.18 A log should be kept of the number of packets sent and received over a subnetwork, those rejected by the subnetwork, and those received with a failed CLNP Header checksum, or a Deflate checksum failure [Req 30]. Later analysis of the log, compared with errors detected by the subnetwork user, can then provide an estimate of the subnetwork undetected error rate [Req 31].

2.4.19 Throughput

2.4.20 Throughput achieved can be simply measured provided that a record of each packet sent (size and time of transmission) over each communications path is maintained [Req 32]. Offline analysis of this record may then be used to determine throughput. [Req 33]

2.4.21 Connection Establishment Delay

2.4.22 For connection mode subnetworks, the time at which each subnetwork connection is initiated and the time of successful establishment should be recorded [Req 34]. Later analysis of the record can then be used to determine the subnetwork connection establishment delay [Req 35].

2.5 *Objective 4: Measurement of Path Segment Performance*

2.5.1 The requirements for measuring path segment performance are related to the performance management model adopted. They are hence discussed in Objective 12: Analysis of Data Stream Performance below.

2.6 *Objective 5: Monitoring for Excess Capacity*

2.6.1 Excess capacity occurs when there exist underused data links. These could potentially be removed or replaced by less performant and lower cost data links.

2.6.2 Two metrics are required to determine excess capacity: the actual throughput over each subnetwork (peak) and the available throughput. When the peak throughput is significantly lower than the available throughput then the data link may be eliminated if a suitable (and also underused) alternative exists. Alternatively, it could be replaced by a lower cost/capacity subnetwork. The throughput threshold when a lower cost/capacity subnetwork is realistic needs also to be known.

2.6.3 Subnetwork throughput is determined from subnetwork monitoring (see Throughput 2.2.2.2.4.19). The available throughput and the threshold for downgrading to a lower cost/capacity subnetwork are required to be known a priori. A monitoring tool should be available to monitor subnetwork utilisation and report on candidates for downgrading [Req 36].

2.7 *Objective 6: Planning Future Capacity*

2.7.1 Capacity planning requires the development of a network design model [Req 37]. This will comprise the Routers, subnetwork interconnections and predictions of data flows through the internetwork, with both normal and busy hour profiles required. From this model, the capacity requirements of subnetwork connections and Routers can be predicted.

2.7.2 In the ATN, the Network Design Model is complicated by the existence of ATSC Class which restricts the options for data flow, as well as other types of traffic such as AOC. It will also be necessary to use the model to predict behaviour during outages. For this reason, more than a simple static model may be required, in order to simulate the impact of subnetwork connection and Router loss, and to demonstrate that there is sufficient capacity to maintain the required Quality of Service to high priority applications during such failures.

2.7.3 The model will require good quality information on the expected traffic flows if it is to be useful. Furthermore, accurate predictions of future traffic levels will be necessary if it is to be used to plan the growth of the network. This requires capture of information on current network loading and analysis of historical network performance information in order to perform trend analysis and hence to predict future growth.

2.7.4 The following data is thus required to be recorded during daily operations and kept for historical analysis:

a) Data volumes for each subnetwork on a point to point basis and during each sample period [Req 38].

b) Data volumes (both bytes and packets) handled by each Router during each sample period [Req 39].

2.7.5 The above needs to be analyzed by priority and ATSC Class.

2.8 *Objective 7: Dynamic Capacity Management*

2.8.1 There are three requirements that flow from this. The first is a need for an indication that the packet flow rate over a given subnetwork has increased beyond some threshold that implies a need for additional capacity. The second is a corresponding indication that the additional capacity is no longer required, and the third is the means to bring on new capacity.

2.8.2 *Indicating the need for Additional Capacity*

2.8.3 The first indication that the applied load is reaching the limits of the subnetwork will be an increase in the queuing delay for packet transmission over that subnetwork. This could be measured as either the absolute queuing delay per packet, or, perhaps more easily, as the average queue length during a sample period. As queue length is proportional to queuing delay, this should be sufficient.

2.8.4 An alternative mechanism might be to measure the actual throughput over some period and compare it against the maximum achievable. However, this requires knowledge of the maximum achievable throughput and, as this is not necessarily a constant (if the subnetwork includes an element of statistical multiplexing), this can be difficult to predict accurately. This can be further complicated if more than one subnetwork is already used to support a router to router adjacency.

2.8.5 Average queue length is thus the preferred mechanism for determining when an event should be generated requesting additional capacity. As the quality of service requirements can differ by priority, there may need to be different reporting thresholds for different priority bands [Req 40].

2.8.6 *Indicating When Additional Capacity is no longer Needed*

2.8.7 The inverse of the above may be readily adopted as the signal that indicates when capacity may be reduced i.e. when the average queue length drops below some threshold, However, the normal queue length (i.e. when load and capacity are in balance) will be between zero and one. Therefore, the measure of average queue length will have to be a real number if it is to be useful [Req 41].

2.8.8 Alternatively, if throughput levels are used to determine when the additional capacity is to be provided, then throughput monitoring (as in Objective 5: Monitoring for Excess Capacity) should also be used to determine when the additional capacity is to be withdrawn.

2.8.9 *Adding and Removing Additional Subnetwork Capacity*

2.8.10 There is a need for Systems Management Actions to be provided to bring up new subnetwork connections or to remove existing subnetwork connections in response to the above events. While these actions could be invoked by a remote manager, the simplest implementation is probably to have a local response (i.e. within the Systems Management Agent) [Req 42].

2.9 *Objective 8: Router Forwarding Measurement*

2.9.1 In addition to the data forwarded by a Router, as required above, a complete assessment of the Router's performance will also require that packet discards are recorded and analyzed by discard reason [Req 43].

2.9.2 If the reason for packet discard is "congestion", then this may be indicative of a lack of network capacity. There is thus a need to report, to a Network Manager, when the number of such discards exceeds a given threshold during a reporting period [Req 44].

2.10 *Objective 9: Router Key Parameters Measurement*

2.10.1 The key parameters of the Router that affect forwarding must also be measured. This may include system specific parameters [Req 45] (e.g. memory utilization) and the number of entries in the Forwarding Information Base (FIB) [Req 46].

2.11 *Objective 10: Measurement of Route Convergence Times*

2.11.1 The measurement of Route Convergence Time requires that each Router records each routing event for collection and later analysis [Req 47]. Furthermore, so that routing events in different routers can be correlated, each must have a synchronised clock [Req 48]. The accuracy of the synchronization is for further study.

2.11.2 An offline tool will be required to perform the analysis. This will analyse the logs provided by each Router and correlate routing events between routers. This tool will be required to determine the time taken from each change in the network topology to a stable routing set being reached in every router, including the elimination of any false routes that may have been introduced as a consequence of the change.

2.11.3 Trend analysis of route convergence rates will also be necessary against historical data in order to predict the need for higher performance routers. This is so that the convergence time does not increase to the point at which the end-to-end Quality of Service falls below acceptable levels [Req 49].

2.12 *Objective 11: Air/Ground Router Overhead Measurement*

2.12.1 An Air/Ground Router will need to log:

a) Establishment and termination of adjacencies with Airborne Routers [Req 50].

b) Establishment and termination of subnetwork connections with Airborne Routers [Req 51].

2.12.2 A tool will be required to analyze such logs and thus to determine the numbers of such adjacencies and subnetwork connections at any one time and to compare this against the monitored performance of the router. The maximum number of adjacencies and subnetwork connections that can be maintained without significantly impacting performance can then be determined, and hence the need for future capacity planned [Req 52].

2.13 *Objective 12: Analysis of Data Stream Performance*

Note. The need for this objective depends upon the model for service level agreements developed by the industry.

2.13.1 If the performance assessment mechanism adopted by the ISP requires that data flows are individually monitored, then the ISP will need to demonstrate compliance with Quality of Service metrics for a given user's data flows. In turn, this requires identification and metering of data flows on both entry to and exit from the ISP.

2.13.2 The metrics to be monitored are those appropriate to a connectionless internet i.e. Transit Delay, Integrity, Throughput and Reliability [Req 53].

2.13.3 The metering of packets (both by number and size) on entry and exit can be used to assess both throughput and reliability. Note that as different service levels may apply to different priorities and ATSC class, data flows may require separate meters by priority and ATSC Class, as well as other traffic types such as AOC.

2.13.4 Throughput can then be measured by direct computation of the metered values. An assessment of reliability can be gained by comparing counts of packets that entered the ISP's network with those on exit [Req 54].

2.13.5 Measurement of transit delay is not quite so straightforward. The "brute force" approach would be to log each individual packet on entry and exit and to record the time of entry and exit. Offline analysis could then compute the transit delay for each packet and hence the average transit delay, etc. Alternatively, time-stamped user messages could be sent. Both of these would require synchronized clocks at each end.

2.13.6 However, this approach would be computationally expensive and is not readily justifiable. Instead two approaches are possible:

a) A sample of such packets is logged. The sampling procedure could use some algorithm based on the packet identifier in order to select packets for logging and this would guarantee the same packets were selected at each end of the link. The approach is basically the same as the brute force approach, but it does reduce the amount of data to be logged and processed.

b) Periodic echo packets are sent on the data flow. These can be used to directly compute the round trip delay and hence derive the transit delay.

2.13.7 As the computational effort is much lower, the use of periodic echo packets for transit delay monitoring is to be preferred. The echo packets will need to be sent at various ATSC classes and priorities in order to determine the service levels for each ATSC class and priority [Req 55].

2.14 *Objective 13: Monitoring of Service Provider Compliance by Aggregate Data Flows*

Note. The need for this objective depends upon the model for service level agreements developed by the industry.

2.14.1 In the macroscopic view, the ISP is interested only in the total data flow at each entry and exit point. There is no analysis to the level of individual data flows. In general, this objective requires less computational effort than Objective 12 and hence it may be preferred.

2.14.2 The metrics are still the same i.e. Throughput, Transit delay and Reliability. Again, they may need to be metered by ATSC Class and priority. However, the meters are on aggregate values and not on individual data flows [Req 56].

2.14.3 Total throughput can be calculated from this information as can an overall assessment of reliability. Transit delay will have to be determined by use of periodic echo packets between all entry points and all exit points. The echo packets will need to be sent at various ATSC classes and priorities in order to determine the service levels for each ATSC class and priority [Req 57].

2.15 *Objective 14: Monitoring of User Message Delivery*

2.15.1 In monitoring the general performance objectives for user message delivery, it is necessary to define, for every application, at what point a "message" is considered to be submitted to the communication subsystem, and at what point it is considered to be received. There is a finite possibility of user messages being lost or corrupted in the local end-systems, or of failure to resolve a recipient address correctly after the message is deemed to have been submitted. There is also the possibility of message submission being refused at source (e.g. due to lack of local resources), but such messages would not be included in the non-receipt statistics.

2.15.2 For ATN applications, a "User" message can be considered as submitted or received when it crosses the ASE service boundary. Although there is no requirement for the abstract ASE service to be physically realized, there would seem to be a requirement, from a performance management standpoint, that an event is logged whenever a user message crosses the notional ASE service boundary. [Req 58].

2.15.3 The requirement is to detect user messages which were submitted and then either:

a) mis-delivered (i.e. delivered, but to a peer user other than the intended recipient),

b) not delivered to the intended recipient (lost or mis-delivered), with notification to the originator,

c) not delivered to the intended recipient, with no notification to the originator.

2.15.4 Two approaches are possible:

a) In the direct end to end approach, user messages submitted and received are logged at each peer user system and the logs are later compared in order to detect mis-matches.

b) Alternatively, the probability of message loss occurring within the peer end systems could be computed by analysis of local traces, and the probability of message loss or mis-direction by the ICS computed separately, and then the figures combined.

2.15.5 In any case, it is not possible to monitor the non-receipt statistics in real-time. The analysis must be performed off-line, probably by statistical sampling rather than analysis of every user message. The non-receipt statistics must then be aggregated over an agreed period of time. [Req 59].

2.16 *Objective 15: Monitoring of User Message Transfer Times*

2.16 1 There is no general purpose mechanism to measure transit delay between application users. An unconfirmed application message such as an ADS Report, has no acknowledgement, and in such cases the transit delay can only be definitively measured by use of time-stamps.

2.16.2 On the other hand, a confirmed application service, such as CM-Logon, does provide an acknowledgement, but this has no real-time constraints and is thus unsuitable for measurement of transit delay. Other application services have optional responses, such as CPDLC LACK messages, and these could potentially be utilized to measure round-trip times when present. However, the processing delay within the target system is potentially significant and will need to be independently measured and factored out.

2.16.3 A strategy for measurement of transit delay will need to be developed for each type of application. For unconfirmed application messages, this will need to be based on timestamps and synchronized clocks, or else some kind of systems management echo function will have to be defined. For confirmed application services this may use the application's own acknowledgement procedures.

2.16.4 To measure transfer times directly would require time-stamping of all application data at the moment it is submitted to the communications subsystem for sending. Then, assuming synchronized clocks across the ATN, the receiving system can directly compare the times of sending and receiving. If clocks are not synchronized, it would be possible to estimate the time differential between pairs of systems by the exchange of calibration messages.

2.16.5 However, there is no systematic time-stamping of application APDUs in the currently-defined ATN applications. Time stamps are available in:

a) All CPDLC uplink and downlink messages, where the date/time can be set, with 1 sec resolution, by the CPDLC-User.

b) All ADS reports sent air-ground, with 1 sec resolution.

2.16.6 Use should be made of application time-stamps, where available, to estimate the end to end transfer delays of user messages. [Req 60].

2.16.7 Where application messages are not time-stamped and/or there is no time synchronization, the "typical" transit delay can be estimated via a round trip exchange. The measure then includes "budgets" for human response times, system processing times and communication transit delays. If no dialogue was in place, the delay includes the connection establishment delay and the transfer delay for the two messages. Otherwise, the delay includes the data transfer delay for the two messages only.

2.16.8 For each confirmed application service, the mean and maximum round trip delay values can be measured directly. [Req 61].

2.16.9 For all applications, including unconfirmed services, the transfer delay can be estimated by factoring out the delay imposed by the ATN internet and upper layers from the delay imposed by the human response time and system processing time. The latter will be fairly static, while the ATN internet transit delays can be determined as described in 2.2.2.2.2.10: End to End Transit Delay.

2.16.10 The objective is to maintain a log of user message transfer times vs. requested ATSC class for offline statistical analysis. Over a given period, the mean, 95% and 99.996% end to end transfer delays can then be computed for received user messages for each ATSC class. [Req 62].

2.16.11 Note that it is not necessarily an error when an individual message exceeds one of the transfer time thresholds, since a finite percentage of messages are allowed to exceed the delay thresholds. The question remains whether or not to:

a) signal to the peer users when an individual message exceeds one of the transit time thresholds, or

b) notify the SM manager when an individual message exceeds one of the transit time thresholds.

2.3 Summary of Performance Management Requirements

The following requirements have been derived from the preceding analysis. There is also an implicit requirement [Req 63] to be able to selectively enable and disable event logging where a requirement to log an event is identified.

General

All ATN systems are required to keep local event logs for the recording of designated systems management events. Some mechanism must also be provided to transfer these logs to an offline processor. [REQ 1].

All ATN Systems are required to log each successful and each unsuccessful attempt to establish a connection over a connection mode subnetwork. [REQ 24].

All ATN Systems are required to log the number of successful and unsuccessful attempts to send a packet over a connectionless subnetwork. [REQ 24].

All ATN Systems are required to log the number of packets sent and received over each subnetwork or subnetwork connection, and to count the volume of data sent and received, analyzed by priority and ATSC Class. [REQ 26, REQ 32, REQ 38].

All ATN Systems are required to log the time of uncommanded loss of a subnetwork connection. [REQ 28].

When possible, ATN systems are required to log the time of each failure to transmit a packet over a connectionless subnetwork, and time of the next successful transmission attempt. [REQ 28].

All ATN Systems are required to keep a count of the number of packets received with a CLNP header checksum failure. [REQ 30].

When Deflate is implemented, ATN Systems are required to keep a count of the number of packets received with a Deflate checksum failure. [REQ 30].

On each connection mode subnetwork, an ATN system is required to log the time at which each connect request is sent and the time at which the connection is successfully established. [REQ 34].

All ATN Systems are required to monitor average queue length analyzed by priority during each sampling period, and to generate a notification when the average queue length exceeds a set threshold (high watermark) or drops below another set threshold (low watermark). [REQ 40, REQ 41].

When subnetworks or subnetwork connections can be dynamically managed, Systems Management actions are required to be available to activate and deactivate them. [REQ 42].It should be possible to selectively enable and disable event logging where a requirement to log an event is identified. [REQ 63].

End System

The user application is required to record, in a local log, each successful attempt to establish an end-to-end connection [REQ 2].

The user application is required to record, in a local log, each unsuccessful attempt to establish an end-to-end connection [REQ 3].

The transport layer is required to record, in a local log, the number and total size of user messages sent on each transport connection, in a measurable time period. [REQ 5, REQ 15].

The transport layer is required to record, in a local log, the number and total size of user messages received, with and without errors, on each transport connection, in a measurable time period. [REQ 6, REQ 15].

The transport layer is required to record, in a local log, each connect request, and the time at which the connect request was issued. [REQ 18].

The transport layer is required to record, in a local log, the time of each successful connection establishment. [REQ 19].

The transport layer is required to record, in a local log, each uncommanded transport connection loss. [REQ 8].

The transport layer is required to record the measured round trip delay between transmission of a TPDU and its acknowledgement together with an indication of whether the TPDU marks the end of a TSDU. [REQ 10].

When authentication is implemented, authentication failures are to be logged. [REQ 13].

The Dialogue Service is required to record, in a local log, each connect request, and the time at which the connect request was issued. [REQ 17].

The Dialogue Service is required to record, in a local log, the time of each successful connection establishment. [REQ 19].

The number, average and maximum size of CLNP packets sent and received during a reporting period are required to be logged. These are to be analyzed by ATSC Class and priority, and by each data link. [REQ 21].

The mean transit delay of packets through an End System is required to be measured under various loading conditions. [REQ 22].

An event is required to be logged whenever a user message crosses the notional ASE service boundary. [REQ 58].

Use should be made of application time-stamps, where available, to log the end to end transfer delays of user messages. [REQ 60].

For each confirmed application service, the round trip delay between request and confirmation messages should be logged. [REQ 61].

Router

ATN Routers are required to keep counts of packets forwarded and data volumes, analyzed by priority and ATSC Class. [REQ 39].

ATN Routers are required to log packet discards by discard reason. [REQ 43].

When the rate of packet discards due to congestion in a given period exceeds a defined threshold, then a notification should be sent to a network manager. [REQ 44].

System specific parameters that affect forwarding performance should be logged [REQ 45].

Changes to the number of entries in a Router's FIB should be logged. [REQ 46].

ATN Routers are required to log each route received and each route advertised to another router, recording the time received/advertised. [REQ 47].

ATN Routers are required to maintain synchronized clocks, to a specified accuracy, for event logging purposes. [REQ 48].

An Air/Ground Router should log the establishment and termination of adjacencies with Airborne Routers [REQ 50].

An Air/Ground Router should log the establishment and termination of subnetwork connections with Airborne Routers [REQ 51].

Depending on the performance assessment model adopted, ATN Routers may be required to meter each data flow, counting number of packets and data volumes for each identified data stream, where a data stream is identified by a unique combination of source, destination, priority and ATSC Class. [REQ 53].

Depending on the performance assessment model adopted, ATN Routers may be required to meter the number of packets and data volumes received from each identified source and sent to each identified destination. Separate meters are required for each ATSC Class and priority. [REQ 56].

Offline Analysis Tools

A tool is required to process logs of connection establishment successes and failures in order to determine the service availability. [REQ 4].

A tool is required to process logs of messages sent and received on individual transport connections and to correlate the logs at both ends of the same transport connection, in order to determine reliability. [REQ 7].

A tool is required to analyze transport layer logs for uncommanded transport connection disconnects and the later successful re-establishment to the transport connection to the same destination, if any. This is to measure service continuity. [REQ 9].

A tool is required to analyse TPDU round trip delay logs in order to estimate the average transit delay per user message. [REQ 11].

A tool is required to analyse the user messages recorded (both airborne and ground). The tool would correlate individual messages and compare them for any loss of data integrity.
[REQ 12].

A tool is required to analyze logs of security authentication failures, in order to assess the data integrity achieved. [REQ 14].

A tool is required to analyze logs of messages sent and received per transport connection, correlated with application type, in order to determine the achieved throughput for each application. [REQ 16].

A tool is required to analyze End System logs to determine the load placed on the network by the End System. [REQ 23].

A tool is required to analyze subnetwork access logs in order to derive the percentage availability of each subnetwork. [REQ 25].

A tool is required to analyze subnetwork usage logs, to correlate the logs of all users of a given subnetwork, and hence to determine subnetwork reliability. [REQ 27].

A tool is required to analyze subnetwork logs in order to determine the continuity of service. [REQ 29].

A tool is required to analyze packet received counts, and CLNP Header and Deflate checksum failures in order to assess subnetwork integrity. [REQ 31].

A tool is required to analyze logs of data volumes sent over each subnetwork in order to determine the achieved throughput. [REQ 33].

A tool is required to analyze logs to determine the connection establishment delay for subnetworks, transport and dialogue services. [REQ 35, REQ 20].

A tool is required to analyze logs of subnetwork usage in order to compare achieved utilisation against available capacity and hence to identify where excess capacity may exist. [REQ 36].

A Network Design Model is required for capacity planning purposes. [REQ 37].

A tool is required to analyze router logs of packets forwarded and discarded and hence to assess router throughput. [REQ 43].

A tool is required to measure route convergence rates and to predict future trends in route convergence rates. [REQ 49].
A tool is required to analyze the impact on an Air/Ground Router of the number of Airborne Router adjacencies and subnetwork connections that it supports. [REQ 52].

Depending on the performance assessment model adopted, a tool may be required to analyze data stream meters in order to determine throughput and reliability as provided to each service user. [REQ 54].

A tool is required to generate echo packets in order to measure roundtrip delay and hence transit delay for each identified data stream. [REQ 55, REQ 57].

Depending on the performance assessment model adopted, a tool may be required to analyze meters at entry and exit points in order to determine throughput and reliability as provided by the Service Provider. [REQ 56].

A tool is required to compare end system logs of user messages sent versus those delivered, non-delivered and mis-delivered. The non-receipt statistics must then be aggregated over an agreed period of time. [REQ 59].

A tool is required to analyze logs of user message transfer times vs. requested ATSC class. Over a given period, the mean, 95% and 99.996% end to end transfer delays should be computed for received user messages for each ATSC class. [REQ 62].

Open Issues

A strategy is required on a per subnetwork basis to measure transit delay over that subnetwork.

Subnetwork specific monitoring criteria need to be developed.

There is the potential need for an industry group to collate together ISP performance statistics and to plan development of future capacity.

The question remains whether or not to signal to the peer users and/or notify the SM manager when an individual user message exceeds one of the transit time thresholds.

A strategy for measurement of transit delay will need to be developed for each type of application.

"User" Communications Performance Over ATN

Communications performance (per Doc. 9705, Table 1-1, ATSC class specifying 95% one-way transit delay) is specified by "user" systems in the initial connection request for e.g. CPDLC, ADS, or CM.

The sender of a user message thus requests that a particular ATSC class be assigned to that message. The ATN Internet will use best endeavors to honor that request. If the requested ATSC class is unavailable, the first known ATSC class higher than that requested or, if there is no such route, the first known route of lower ATSC class than that requested is provided.

The recipient of a user message knows the ATSC class requested for the current dialogue, but has no knowledge of the ATSC class(es) of the route actually used. (For example, for CM, the ground "user" system has no means of knowing the ATSC class available or used for the CM Logon request, i.e. the ground "user" system is unaware of the sub-network used, or the communication performance available, at time of log-on).

For example, when a "user" system requests a connection, it specifies an ATSC communication class C or D, e.g. within a CPDLC_connect_request issued by a ground system. If the requested 'C' is not available, and there is no better class available, the next best would be offered until 'no preference' is reached. Neither the requesting nor confirming "user" is notified of the class eventually used.

If communication via any path, at any class, is or becomes completely unavailable, the "user" is notified by a provider abort (i.e. in the example case, TP4 notifies the Dialogue Service, which notifies the CPDLC ASE, which notifies both the air and ground "user" systems via the CPDLC-provider-abort service).

The ATSC class established and made available for a connection is dependent upon sub-network availability (e.g. VHF, Mode-S, Satellite), and performance subsequently provided to the "users" can therefore change after connection is established if, for example, a sub-network becomes available or is temporarily lost. This can happen between "user" messages, or within the same "user" message, and will result in a performance increase or degradation over that previously observed for these "users". The requesting "user" is not notified of any such changes, but must infer it from observed performance.

Systems Management (SM) functions could report e.g. excessive delays, based on monitoring done by protocol layer (e.g. transport, CPDLC ASE). The output of such SM functions could for example be communicated to technical watch facilities, or the "user" systems, if desired. This has not yet been specified. Such notifications could then be used for example to terminate communications if the offered performance went too low, though this would ignore the fact that any degradation could be only temporary. SM functions are designed for communications system monitoring, and not for "user" service observation. Therefore, use of SM functions may provide information to the "users", but should not be viewed as a replacement for operational datalink Service level monitoring.

Recommendations for initial ATN datalink operations have been made in other fora as follows:

a) End Users should identify requirements for the SM Manager specifications to enhance, but not replace, Service level monitoring. This could include e.g. aborting the link based on performance observed for other "users" and this aircraft (e.g. AOC exchanges), or aborting the link if the available sub-networks are reduced to only those known to be incapable of meeting the originally specified class of service based on static definitions.

b) ATN "user" systems should consider implementing timer monitoring functions for CPDLC connection requests, and connection triggering messages for ADS, to establish the current performance being offered.

c) ATN "user" systems should consider implementing timer monitoring functions for other regular exchanges, at the Service Level, to identify changes in performance levels. The aim would be to terminate the connection if performance drops below that operationally required for datalink in the area concerned.

For those End User systems which use Logical Acknowledgements, CPDLC round-trip monitoring could be based on LACK message

– – – – – – – – –