

Quality of Service Management and Interoperability

Christian Kissling and Tomaso de Cola
*German Aerospace Center (DLR)
Oberpfaffenhofen,
Germany*

1. Introduction

Within this chapter quality of service management strategies are assessed with respect to their applicability and efficiency in the ATM context. In particular addressing the service demands of ATM communication, such as strict latency and loss limitations is considered herein. This also covers the architecture for selection of links for data transmission and the interaction between technology independent and technology dependent components in the networking architecture by means of standardized communication protocols such as IEEE 802.21 and ETSI BSM extensions.

The term Quality of Service (QoS) is used in a variety of different ways and often depends also on the context that it is used in. One notion of QoS denotes the performance of a service from the users view. A measure for the grade of QoS is how good the performance attributes of a service match with the demands made on it. The kind of attributes which are relevant and need to be fulfilled depends thus naturally on the context of the service. While for many other services perceived or qualitative QoS measures are applicable, the ATM communication environment envisaged here makes high and precise demands on different attributes, presented later on in detail. For common internet applications such as HTTP, email and VoIP a lot of work has been spent to map the quality perceived by the user to networking parameters which can be measured and controlled. (ITU-T, G.1010) provides for instance a model for multimedia QoS categories from the end user perspective. Tables with technical requirements are provided for eight different application categories such as audio, video and data services. In the literature a large number of publications are available dealing with different aspects of providing QoS in the terrestrial Internet but also in satellite networks. (Marchese, 2007) provides a good overview and summary of different aspects of QoS provision in the context of heterogeneous networks, such as present in the ATM environment considered here.

The provision of QoS in an operational and safety critical aeronautical environment is however considerably different from the applications and demands in the Internet. Service parameters such as defined in (ITU-T, G.1010) thus cannot be directly applied here. The most intuitive reason for this is that a violation of QoS attributes in Internet applications results in a reduced service quality, which is naturally undesirable and bothersome for users, but has not necessarily implications on operational events and safety of life. In the aeronautical domain, for the management of air traffic this is decisively different. Late

arrival of e.g. directive commands issued by the controller for the pilot can have catastrophic effects. Also corrupted messages or multiple receptions of messages can have such serious consequences, affecting the safety of the airplane and the passengers. For this reason it is not sufficient if the QoS mechanisms for ATM communication try to achieve the requirements as far as possible but it is necessary that the requirements are definitely met.

1.1 Network design

Since IPv6 is the unification point in the SANDRA network (SANDRA, 2011), there is the need of the design and adaptation to an aeronautical internet. Main focus within this task is the handling of the network management and also of the resource management. Additionally, effort is spent for the development of new and efficient handover and mobility management algorithms and concepts, respectively. Also an IPv6 based naming and addressing architecture will be provided. Due to the high degree of mobility on a global scale and the heterogeneous network environment (i.e. short-range and long-range terrestrial as well as satellite access technologies), work on a network mobility (NEMO) based IPv6 protocol started in contrast to the ICAO chosen Mobile IPv6 protocol supporting only host mobility.

For the SANDRA Terminal as shown in the aircraft segment of Fig. 1, the lower layer (data link and physical layers) functions are provided by an on-board Integrated Modular Radio (IMR) consisting of heterogeneous radio access technologies.

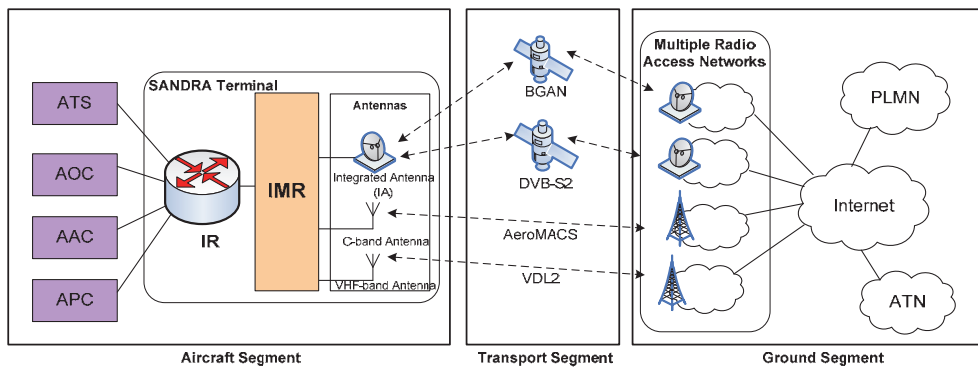


Fig. 1. SANDRA Network Architecture (Ali, 2011).

The upper layer (layer 3 and above) functions are managed by an Integrated Router (IR). The following chapter will describe in detail the realization of the connection of these two entities.

2. Quality of service management and interoperability

2.1 QoS definition for aeronautical networks

In a joint study of EUROCONTROL and the Federal Aviation Administration (FAA), potential future communication technologies which are suitable to provide the necessary safety and regularity of flight have been investigated and requirements for the future application services have been derived. The results of this study have been published in the so called "Communications Operating Concept and Requirements for the Future Radio

System (COCR)" (EUROCONTROL, 2007). Within this study the concepts of ATM have been analyzed from an operational point of view and the expected technical requirements have been formulated, also for services which are not yet deployed but are expected to be deployed in the future. The results in the COCR provide information for all operational services with respect to their periodicity, volume and technical requirements. The main QoS requirements for the services are the following ones:

- **Transmission delay (TD₉₅):** The TD₉₅ represents the one-way latency requirement for every Operational (OP) message which 95% of all messages of a service have to arrive within. It is defined per flight domain (i.e. Airport (APT), Terminal Maneuvering Area (TMA), En-Route (ENR) and Oceanic, Remote and Polar (ORP)), per service type (ATS and AOC) and for each Class of Service (CoS).
- **Expiration Time (ET):** In case the TD₉₅ is not met due to various reasons (e.g. packet loss) the COCR sets a so called Expiration Time within which the packets have to arrive which failed the TD₉₅ requirement. To be compliant with the requirements, the percentage of messages indicated by the continuity requirement has to arrive within the ET.
- **Continuity:** Denotes the probability that a transaction will be completed having met specified performance. With respect to the ET, this probability represents the percentage of the transmitted messages which arrive within the latency performance requirement set by the ET.
- **Integrity:** Denotes the acceptable rate of transactions that are completed with an undetected error. This requirement refers to packets which are considered to be received correctly but actually contain false information, e.g. caused by undetected bit errors
- **Availability:** Denotes the probability that the equipment comprising the system is operational and conforms to specifications (excluding planned outages and logistics delays). It is further distinguished into
 - *Availability of use:* Probability that the communication system between the two parties is in service when it is needed.
 - *Availability of provision:* Probability that communication with all aircraft in the area is in service.

The COCR specifies these QoS requirements per service, but also for aggregated Classes of Service (CoS). For the definition and evaluation of the QoS architecture, the three main impacting requirements are thus the **TD₉₅**, the **ET** and the **Continuity** requirement. Table 1 shows an excerpt from the COCR, specifying the ET, TD_{95-FRS} and Continuity (C_{UIT-FRS}) for the different defined CoS.

Within the COCR, the different application services are then also mapped to the CoS listed in Table 1.

It should be noted that these requirements are impacted by the QoS architecture, but not entirely defined by it. Primarily the requirements are dependent on the underlying link technology which set boundaries for latency, packet loss etc. with the available data rate, propagation delay, retransmission mechanisms and forward error correction (FEC) methods. Clearly a QoS architecture cannot ensure compliance with the requirements if the underlying link and physical layer are not capable to transport the data sufficiently. On the other hand, in case the underlying link technology is providing sufficient transmission capabilities, the QoS architecture has to ensure that these abilities are efficiently used so the requirements are met. One challenge of the SANDRA design is thus to define a QoS

architecture which allows meeting the requirements, provided that the underlying link technology provides sufficient performance (in terms of throughput, latency and packet loss rate).

CoS	ET [s]	TD _{95-FRS} [s]	C _{UIT-FRS}	Service Type
DG-A	Reserved	9.8	Reserved	NET Data
DG-B	1.6	0.74	0.99999992	ATS A/G Data
DG-C	5.0	1.4	0.9996	
DG-D	7.8	2.4	0.9996	
DG-E	8.0	3.8	0.996	
DG-F	12.0	4.7	0.996	
DG-G	24.0	9.2	0.996	
DG-H	32.0	13.6	0.996	
DG-I	57.6	26.5	0.996	
DG-J	Not available	13.6	Not available	
DG-K		26.5		
DG-L		51.7		

Table 1. Excerpt from the COCR CoS definitions.

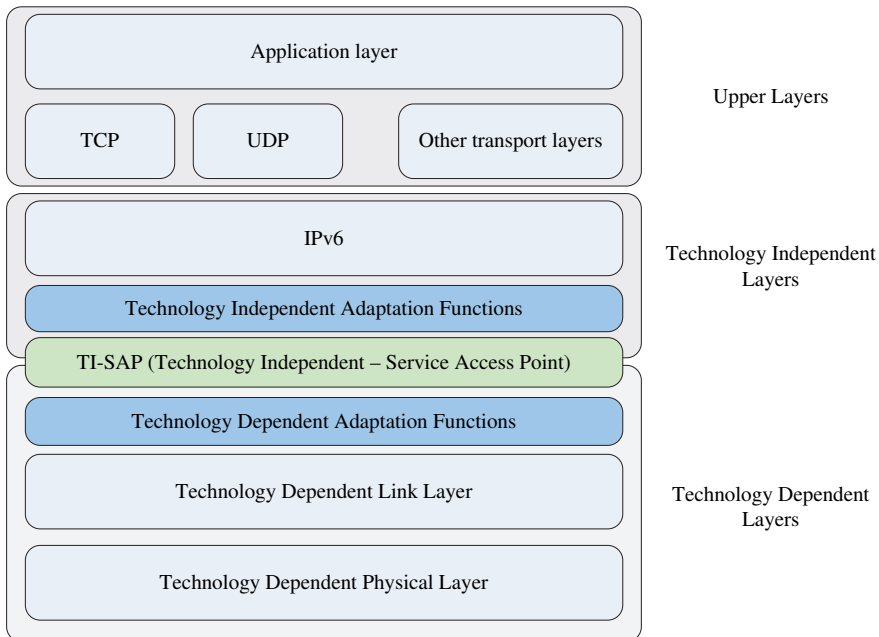


Fig. 2. Functional interaction between technology independent higher layers and technology dependent lower layers.

For the SANDRA QoS design, the additional problem is addressed how different communication links can be integrated into a seamless network and which mechanisms and approaches are suitable to allow provision of the required QoS. SANDRA hereby focuses on the network layer QoS mechanisms mainly. Fig. 2 illustrates the general approach. One requirement for the layer 3 QoS mechanisms is that they must be interoperable and independent of the type of used link. Going beyond this, also the uniform interfaces (denoted Service Access Points, SAP in the following) to the technology dependent Layer 2 are in the scope of SANDRA and discussed hereafter in more detail.

2.2 QoS mapping in the SANDRA architecture

As straightforward from the considerations drawn in the previous section, the necessity for the SANDRA architecture is to simultaneously manage different QoS traffic profiles and transmission technologies over which different services have to be handled, translate into a QoS mapping problem. Beside the technical challenges that arise in selecting the Layer 2 queues to which the traffic has to be forwarded depending on the QoS requirements (scheduling and QoS mapping problem), a particular attention has to be reserved to the characteristics of the QoS architecture, being embedded in SANDRA. Apart from the specific QoS model being adopted (IntServ or DiffServ as sketched in the following sections), some attention has to be addressed to how Layer 3 and Layer 2 intercommunicate, by preserving the QoS requirements specified in the Service Level Specifications (SLS) of the specific traffic service. In this respect, different approaches can be applied. Ad-hoc solutions can be deployed, by extending for instance the functionalities and the related primitives already available from the ISO/OSI protocol stack. Given the scope of the SANDRA framework, it is instead better to have a model in line with architectures currently or going to be standardised. In this perspective, the features offered by the ETSI BSM protocol architecture are worth being considered. The main peculiarity consists in the definition of the SI-SAP interface, virtually separating the upper layer (Satellite Independent, SI) from the lower layers (Satellite Dependent, SD) and providing dedicated primitives to efficiently manage QoS, Address Resolution and Multicast functionalities over satellite. The overall ETSI BSM protocol architecture is depicted in Fig. 3, where the main components are:

- SI layer: it implements the upper layer and in particular the IP protocol (versions 4 or 6). It also incorporates the Satellite Independent Adaptation Function (SIAF) module, which is responsible for adapting the SI functions to the characteristics of the lower layer specification, through dedicated primitives.
- SD layer: it implements the lower layer, in particular the datalink and the physical ones. It also implements the Satellite Dependent Adaptation Functions (SDAF) module, which interacts with the aforementioned SIAF through dedicated primitives.
- SI-SAP interface: it logically separates the SI from the SD layers, providing a set of dedicated primitives, exchanged between the SIAF and SDAF modules, responsible for QoS, address resolution and multicast functionalities.

In this light, it is reasonable to extend the principles of the ETSI BSM protocol architecture for application in the SANDRA framework, to particularly address the QoS requirements of aeronautical networks (Plass,2011).

In fact, two main “ingredients” of the SI-SAP interface can be re-used and properly extended to match the requirements of the SANDRA functional architecture: the Queue Identifier (QID) and the QoS primitives. The former is defined in the ETSI BSM protocol architecture as identifier of the Layer 2 physical queues, so to allow an efficient QoS mapping between Layer

3 and Layer 2 queues, through the dedicated QoS primitives. The latter, in turn, allows actually implementing the QoS mapping algorithms and offering the essential tool to perform the resource allocation, based on the requests coming from the upper layers. The QoS problem in the SANDRA network involves not only resource allocation issues but also transmission technology selection, thus requiring the extension of the current SI-SAP interface functionalities along with the use of the IEEE 802.21 architecture in terms of the Media Independent Handover (MIH) functions. In practice, the QID has to be conceptually extended in a way that it incorporates both queue and link identifiers. Besides, the integration and the interaction of the ETSI BSM and the IEEE 802.21 architecture is of primary importance to perform the communication of the link selection to the upper layer and perform the resource allocation based on the requirements notified from the higher layers (e.g., application protocol or management plane). To this end, the SI-C-QUEUE primitives will be conveniently extended in their scope so to also include the new functionalities, thus allowing the different components to interwork properly according to the SANDRA network characteristics.

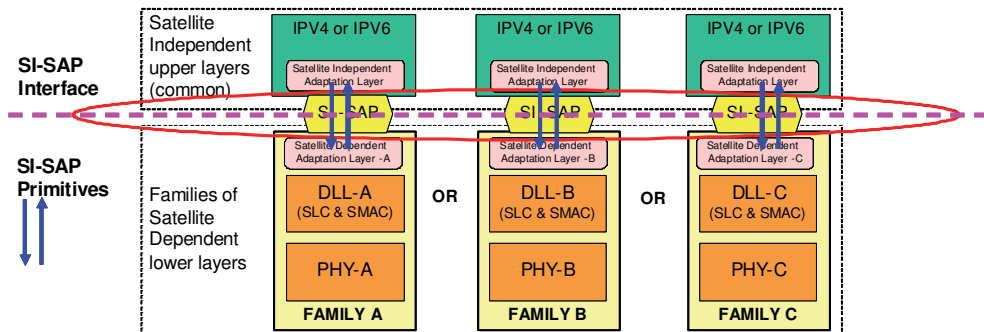


Fig. 3. ETSI BSM protocol architecture and SI-SAP interface definition.

At this point, the final point to be addressed is the way the described protocol architecture integration (ETSI BSM and IEEE 802.21 namely) can be finally embedded in the real architecture of the SANDRA network. In this respect, a particular attention has to be reserved to the IR and IMR interaction. Although the SI-SAP interface has been conceived to logically separate the upper from the lower layers within a satellite terminal, it can be easily extended to physically separate two different components, by distributing the implementation of the primitives. This can be done by re-thinking the SI-SAP interface as separating IR and IMR; these, in turn, will implement the related QoS primitives, thus working as the SIAF and SDAF modules in the original ETSI BSM architecture.

The overall system function can be then summarised in the following operations:

- In case the QoS requirements are constrained to a specific link by the upper layer, the IR will signal the selected transmission technology along with QoS request in a dedicated QID to the IMR, which in turn will forward the forthcoming data traffic to the specified transmission link. The availability of the transmission link is known after the start-up phase, which is accomplished by suitably combining the SI-C-QUEUE-open primitives with the MIH functionalities.
- In case no link-constrained request is performed by the upper layer, the IR simply signals the IMR about the QoS requests. In turn, the IMR will be responsible for running the link selection algorithm to identify the transmission technology most appropriate to match the received QoS requests. Also in this case the signalling is performed through

real exchange of the SI-SAP primitives; in particular, in this case the QID will basically contain an identifier for the QoS request and a default value of the transmission technology, being it not explicitly selected by the upper layers.

- In case a link was no longer available or its availability was reduced (upon notification through the specific MIH functions), the IMR would in turn notify it to the IR through the corresponding enhanced SI-C-QUEUE primitives to trigger a new resource allocation. The IR in turn will run a new resource allocation request to match the new link configuration, by modifying or demanding the assignment of a new QID.

The overall interaction between the SANDRA components is represented in the following picture.

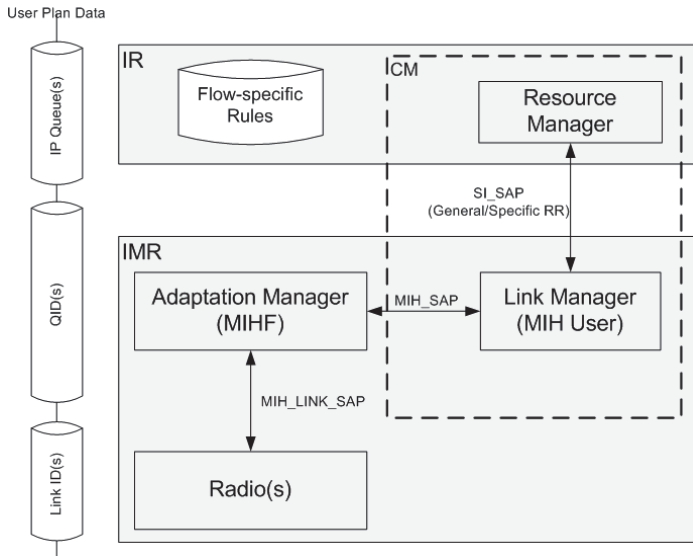


Fig. 4. Interaction between IR and IMR modules within the SANDRA network.

A particular attention has to be reserved to the interaction between IR and IMR in terms of message exchange, performed through primitives' generation and reception according to the architecture above described. In more detail, as it was introduced in the previous paragraphs, the overall IR-IMR system behaviour can be regarded as a sort of Master-Slave interaction, where either the IR or the IMR play the role of master and slave respectively, depending on the specific case being dealt with. In case the application is requesting specific link and QoS profiles, the IR plays the role of master, implying that the IMR will attempt to match the IR requests in terms of link allocation and resource management. On the other hand, when the link selection is forced by the IMR (which plays the master in this situation), the IR is basically responsible for forwarding data through the appropriate logical interfaces to the IMR, without taking any decisions in terms of data filtering and QoS policing/shaping.

The overall interaction can be described as a block diagram, where the two entities (IR and IMR) take decisions based on their role and the functions they are implementing.

The block diagram is essentially composed of the IR and IMR state machines:

- IR: It does not perform any operations unless the request of new radio resource is either issued by the IMR or by other external entities, such as application requests.

- IMR: It does not perform any operation unless a link is available (link label) or the allocated resources need to be updated.

Starting from these two states for IR and IMR, respectively, it is possible to exemplify the dynamics of the overall SANDRA system in presence of constrained and unconstrained services.

As far as the former is concerned, the IR will specify a new radio resource with a specified radio technology. This will be then notified to the IMR through proper primitives, which will be responsible for checking the availability of the requested resources as part of the radio resource management operations. In case the resource are not available, a loop of message exchange between IMR and IR is then initiated to agree on a different resource request, thus possibly ending up with the data forwarding operations.

As far as the latter is concerned, the radio resource request issues without specifying any radio technology, which will be instead selected by the IMR. Accordingly, the IMR is then in the position to setup the selected radio technology and performs the bandwidth allocation upon resource availability, following the same procedure reported before.

An additional case, independent of the specific service being handled, worth being considered is imminent handover or available resource change event. They are both handled by the IMR, which informs the IR through the appropriate primitives. In turn, the IR will update the radio resource assigned to a given service, by issuing a new request to the IMR; in order to match the current resource availability. It could be also the case that when no agreement is reached about the resources that a service should require (e.g., no alternative links are available), the service could be not admitted (or dropped) in (from) the SANDRA network.

The block diagram is sketched in the following picture, Fig. 5.

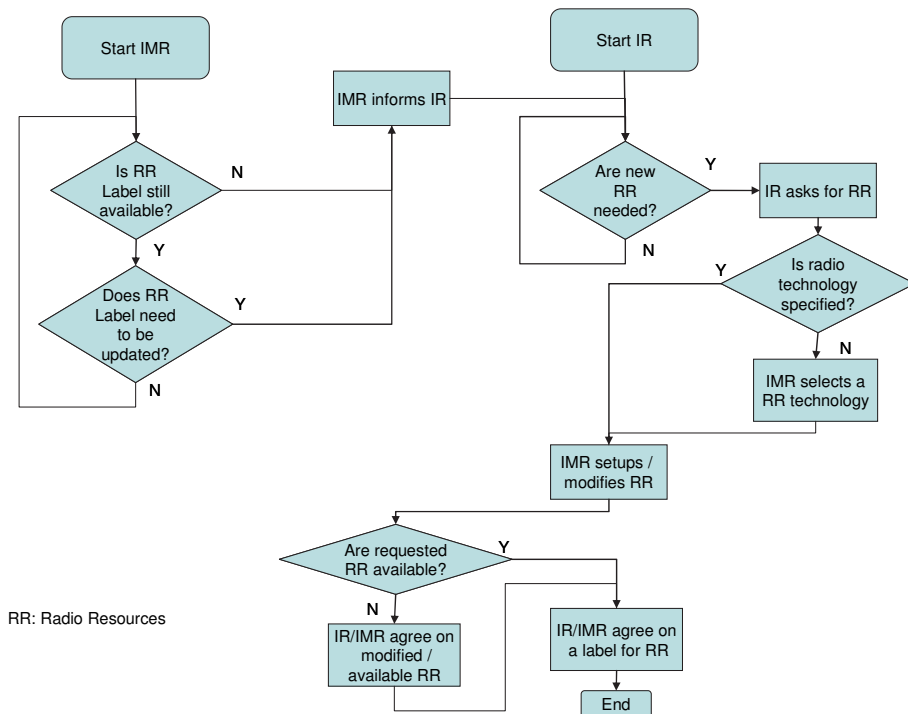


Fig. 5. IR/IMR interaction.

2.3 QoS management architecture

In contrast to QoS architectures which are deployed in the internet, the QoS design in the aeronautical scenario has to comply with a range of security and safety requirements which limit the freedom of choice for a QoS architecture considerably. The selected QoS management architecture should also rely on well established and standardized solutions. From today's perspective, one of the major design constraints is the strict separation of operational (ATS and AOC) and non-operational (AAC and APC) services within the network due to safety. While this separation is a real requirement nowadays, in SANDRA an all-integrated, seamless network is envisaged for the far future, which integrates also operational (OP) and non-operational (NON-OP) services and provides the required safety at the same time. Naturally this has also an impact on the QoS architecture.

Within the SANDRA context the challenge of integrating different communication links into a single common network architecture creates the need to deploy adequate QoS management functionalities. The QoS disciplines which have to be considered in particular for such a QoS architecture design include the following:

- **Connection Admission Control (CAC):** Technique used to decide which traffic is admitted into the network. Going back to the Asynchronous Transfer Mode it is defined as "the set of actions taken by the network during the call set-up phase (or during call re-negotiation phase) in order to determine whether a connection request can be accepted or should be rejected (or whether a request for re-allocation can be accommodated)" (Hitoshi, 1998). In the SANDRA context, the rejection of an OP connection request is clearly not an option. In the scenario where OP and NON-OP domains are fully separated CAC is thus not applicable. When looking into the fully integrated scenario however, CAC is a technique which can be applied to the NON-OP domain to control the amount of traffic admitted from NON-OP sources that is injected into the overall network with the purpose to avoid disadvantageous impact on the OP services. The notion of "connection" can hereby refer to different aspects, e.g. to acceptance/rejection of users entering the system, of TCP connections, SIP connections or general data flows. The use of CAC techniques is supposed to increase the QoS perceived by the users since, e.g. the interruption of a voice call is perceived worse than a rejection of the call in the first place. For these reasons the application of CAC techniques should here be limited to NON-OP traffic.
- **Congestion Control (CC):** In case too many packets are present in a network the performance in terms of delay and loss rate (e.g. due to buffer losses) degrades. This situation is commonly called congestion (Tannenbaum, 2002). While for moderate levels of traffic load (i.e. injected packets) the packet delivery increases proportionally with the load, at some point the message processing is no longer able to cope with the packets, queue sizes first increase (at the cost of increased delay) and finally packets are dropped due to buffer overflow. When retransmission mechanisms without control are present, the packet drop will result in an even higher offered traffic load which in turn results in more dropped packets. Congestion control defines techniques which have the purpose of controlling the occurrence of congestion and ensuring that the network is able to carry the offered traffic. In contrast to flow control techniques, congestion control is a global issue involving all involved nodes. Within SANDRA, the network must be able to cope with the traffic offered by the OP services in any case. In other words the network must be sufficiently dimensioned so congestion due to OP traffic cannot occur. For a scenario where OP/ NON-OP networks are separated, CC is thus

not applicable for the OP part. In the all-integrated scenario, however CC is applicable to the NON-OP traffic in order to ensure that the NON-OP traffic cannot cause congestion in the network which is adversely affecting the OP services.

- **Scheduling / Queuing:** Packets are stored in queues until they can be transmitted on the link. The scheduling algorithm then determines the order in which the queues are served, i.e. the order according to which packets of the different queues are sent. For DiffServ architectures in the Internet, commonly known queues are Expedited Forwarding (EF), Assured Forwarding (AF) and Best Effort (BE). Many different scheduling algorithms are known from the literature, such as First In First Out (FIFO), Round Robin (RR), Weighted Round Robin (WRR), Weighted Fair Queuing (WFQ), Deficit Round Robin (DRR), Stochastic Fair Queuing (SFQ) or Worst Case Weighted Fair Queuing (WF²Q). For wireless communication, other scheduling algorithms taking the wireless nature of the medium into account are defined, such as for instance Idealized Wireless Fair Queuing (IWFQ), Wireless Packet Scheduling (WPS), Channel Condition Independent Fair Queuing (CIF-Q), Wireless Fair Service Algorithms (WFS) or Proportional Fair Queuing (PF). These lists are clearly not exhaustive but shall provide only a fundamental overview. Within the aeronautical scenario, the queuing and scheduling has especial importance since the priority of a packet is directly impacted by it. The COCR defines a range of different Classes of Service (CoS) which also refer to the priority of a packet (e.g. measured in terms of TD₉₅). Proper scheduling techniques ensure that packets belonging to a higher priority service are also transmitted earlier over the link. The scheduling thus addresses the design requirements, which state that the different services within a service category (i.e. for instance DG-C and DG-E within the ATS service category) can be prioritized. Furthermore the scheduling ensures that the different service categories can be prioritized among each other, i.e. for instance ATS over AOC. Finally the scheduling has a significant importance to ensure that OP services (i.e. ATS and AOC) are always prioritized over NON-OP services (i.e. AAC and APC). Since the queue size is in reality always limited, situations can occur where the buffers overflow, e.g. in situations where the link rate is lower than the arrival rate, the buffers fill up and finally overflow. In such a situation where the buffer is full but new packets arrive a decision has to be made on which packet needs to be discarded. There are three basic and intuitive possibilities:

- Drop a random packet in the queue
- Drop the packet at the first position in the queue
- Drop the packet at the end of the queue (tail dropping)

In the context of OP services, the queue management policy may improve the QoS. Here applying a tail dropping policy is not necessarily a good approach, for instance in situations where a packet further in front in the queue is already outdated (e.g. due to a long waiting time in the queue) and the later arriving packet already contains the most recent information. In the case of applying a drop-tail policy the packet with the recent up-to-date information would be dropped whereas the previous packet with the outdated information is sent, since it is already in the queue. This is contra-productive to the goal of providing timely information. On top of this the interaction with higher layer transport protocols such as TCP is relevant. For instance dropping the first packet in the queue may trigger the TCP congestion avoidance algorithm already earlier (which is beneficial), but on the other hand may introduce unnecessary retransmissions

of later packets (which is undesirable). For this reason the selection of queuing policies is of particular interest for OP services when deploying a network.

Additionally, queuing policies try to address the issue of congestion control by applying so called active queue management (AQM). Here the queue length is continuously measured and, when exceeding a threshold, incoming packets are marked (to indicate an imminent congestion situation) according to a probability which is a function of the queue length or are directly dropped with this probability. The original purpose of this AQM was to support the behaviour of TCP and avoiding catastrophic congestion.

- Link selection strategies / routing decisions. Within the future aeronautical communication network, it is expected that many aircraft will have more than one data link technology. Besides legacy links such as the VHF based VDL-2, new link technologies, named Future Radio Systems (FRS) in COCR terminology, will arise. Examples for FRS are Aeromacs, LDACS, or future satellite communication links. For exchanging data a decision has then to be made which of the available links shall be used for transmission. The decision which link is favorable for the data exchange can depend on several criteria, such as cost of link usage, time before outage (e.g. due to leaving the coverage area or a handover), provided QoS and regulatory policies. The link selection strategy must on one hand collect information about the status of the different links and on the other hand try to find the best possible selection which is compliant with the requirements while at the same time minimizing the cost.

2.3.1 Separation of Operational and Non-Operational Domains

From today's perspective, one of the major design constraints is the demand for strict separation of operational (OP) and non-operational (NON-OP) services. This separation can be achieved on different layers:

- Separation at physical layer: Most rigorous form of separation. Here the OP and NON-OP services use different radio frequencies (RF) for transmission and remain entirely separated throughout the protocol stack up to the application layer.
- Separation at link layer: OP and NON-OP services use the same physical RF. Separation is achieved here by means of Link Layer segments, e.g. restricting that within a GSE Layer 2 cell only fragments of OP or only of NON-OP packets must be encapsulated.
- Separation at network layer: OP and NON-OP services may use the same physical RF frequency and also share Layer 2 cells. The separation is achieved here by different IP datagrams which are not shared among operational and non-operational services.

Fig. 6 illustrates the separation between OP and NON-OP service domains as expected for the near future.

As can be seen here, the domains are entirely separated down to the physical layer. The ATC and AOC services are connected to one mobile router, whereas the AAC and APC are connected to a different one. The strict separation of operational and non-operational services has far ranging consequences on the QoS architecture, especially with respect to Connection Admission Control (CAC), Congestion Control (CC) and traffic shaping as was explained earlier. The architecture shown in Fig. 6 was considered as the expected near term situation within the NEWSKY Project (NEWSKY, 2009).

In contrast to this, the more visionary approach which is also investigated in SANDRA is to have a full integration of different service domains into one network and to provide the

needed safety and security among OP and NON-OP services by means of networking techniques.

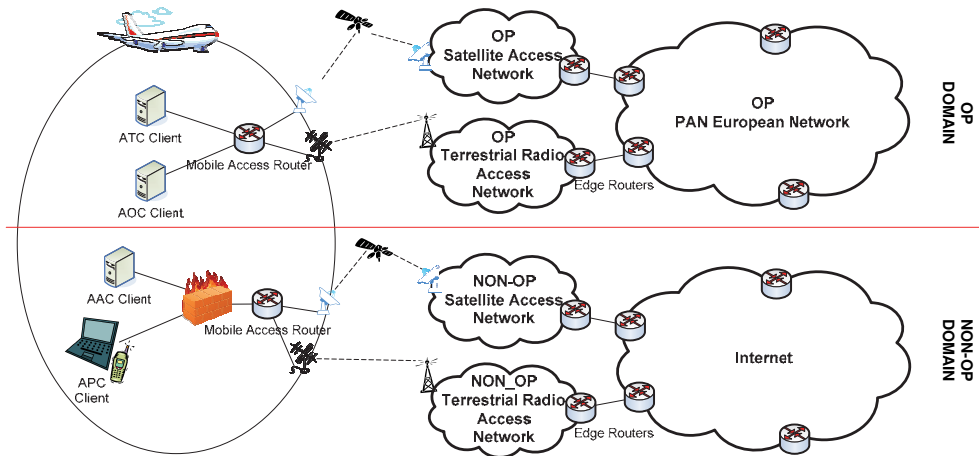


Fig. 6. Service domain separation for near future.

Fig. 7 illustrates an intermediate integration, where on the airborne side OP and NON-OP services are integrated. The division into OP links and networks and NON-OP links and networks still exists here.

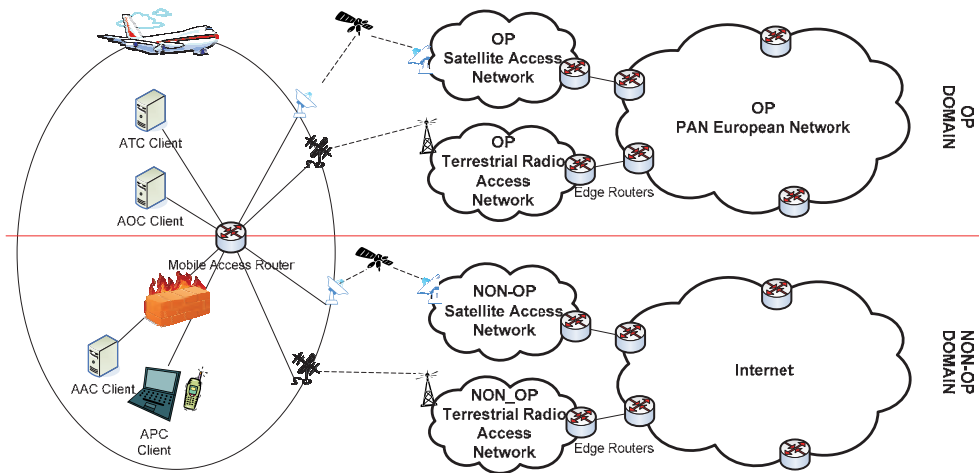


Fig. 7. Service integration in Mobile Access Router, separation at network domain level.

Here besides saving the additional equipment on board of the aircraft (the mobile router for the NON-OP domain) in principle the mobile access router has the freedom to route data over the same links or restrict due to policies the usage of some links, e.g. restricting the use of OP certified links for transporting NON-OP data. As long as the OP access networks on

ground are not interconnected with the NON-OP domain, sharing links between OP and NON-OP services is of course not very meaningful.

The relevance of the integration gets even more clear when looking into a fully-integrated scenario as shown in Fig. 8.

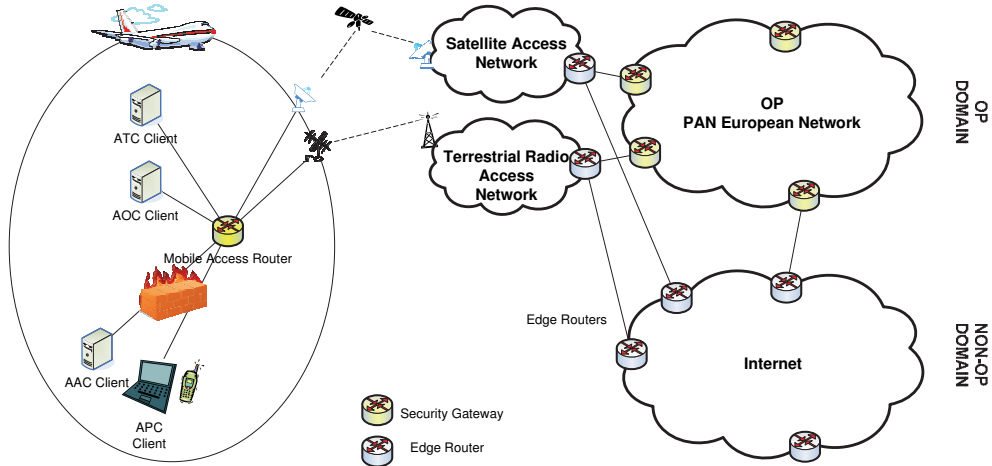


Fig. 8. Full integration of services, links and networks.

In this case the available links (SatCom and terrestrial radio in Fig. 8) may transport OP and NON-OP applications. The edge routers of the access networks then route the data to the other core networks, i.e. the OP PAN European Network domain or the public Internet. The edge routers of the OP PAN European Network domain additional have to provide security functionalities to avoid intrusion and corruption of incoming data. In principle a direct connection of the PAN European Network and the public Internet is conceivable, but not necessarily existing. It is clear that such an architecture creates a strong demand for strong and safe security mechanisms to protect the OP network, otherwise such an architecture will remain unacceptable due to safety concerns; as of now it is disallowed by regulation.

2.3.2 Underlying QoS approach

For provision of QoS different approaches are known from the literature. The suitability of the most well known ones, Integrated Services (IntServ) and Differentiated Services (DiffServ) for application in the aeronautical scenario is briefly reviewed in the following.

2.3.3 IntServ QoS approach

The IntServ architecture (Wroclawsky & Braden, 1997), (Zhang *et al.*, 1997) was developed for supporting specific QoS for end-to-end sessions across networks. In this approach, single flows (representing a stream of packets) are identified and treated individually. Every packet is checked for the resources it is entitled to receive. For this purpose the state of all flows in the network has to be periodically signalled among the routers in the end-to-end path of each flow. The Resource ReSerVation Protocol (RSVP) (Zhang *et al.*, 1997) was designed for this purpose. IntServ also has connection admission control mechanisms as an

integral part of its functionality which admits new traffic to the network only if sufficient resources are available. By doing all this IntServ can guarantee hard upper bounds for packet delays and packet loss caused by buffer overflow. Moreover IntServ can rely with RSVP on an existing and well deployed signalling protocol. The per-flow treatment also allows Multi-Level-Priority-Preemption (MLPP) which can be beneficial to differentiate ATM messages according to their priority and urgency.

While these IntServ features match very well with the QoS requirements in the ATM environment, the application of IntServ would have several major drawbacks. As is the case for all IntServ architectures, the main drawback is the scalability of the system and the signalling overhead. The traffic profile of ATM message exchange as predicted in the COCR consists of mainly small messages in the order few bytes, reaching at maximum several kilobytes in single cases. In the downlink for instance (i.e. aircraft to ground in ATM terminology) the maximum message size is 2763 bytes for the FLIPINT service. Estimations on the traffic profile have shown that the maximum message arrival rate hereby is slightly below 1 msg/s per aircraft at maximum, having an average of less than 0.1 msg/s per aircraft. This means in practice that either for every message a dedicated IntServ flow would have to be initiated and signalled, or an IntServ flow needs to be setup and kept alive for a longer time without being used most of the time, and accepting the overhead caused by the periodic keepalive messages necessary for this. Besides the volume overhead of the IntServ signalling also the time required for session initiation is an important overhead, considering that some messages have latency requirements as low as 0.74 s (Class of Service DG-B) and 1.4 s (Class of Service DG-C). For GEO satellite links already the session initiation would consume a considerable fraction of the maximum latency. Finally the heterogeneous and highly mobile environment, consisting of different link technologies and the belonging different access networks and the need for intra- and inter-technology handovers causes path changes. A change in the end-to-end path would then result also in additional IntServ session re-establishment overheads.

2.3.4 Differentiated Services (DiffServ)

DiffServ (Nichols *et al.*, 1998), (Blake *et al.*, 1998) is the second well known QoS architecture specified by the IETF. In contrast to IntServ no individual flows can be distinguished but only different aggregated classes of traffic. Instead of a guaranteed forwarding behaviour for every flow, DiffServ defines the per-hop forwarding behaviour for the aggregate classes. For identification of the aggregate, the Traffic-Class field in the IPv6 headers are used. Since in DiffServ only traffic aggregates are treated instead of single flows, no hard guarantees for the availability of resources and the end-to-end QoS performance can be given. An overdimensioning of resources is thus necessary here in order to meet the QoS requirements. The overdimensioning affects for instance the buffer sizes in the schedulers to avoid packet drops due to buffer overflow but also the available datarates on the links. While in theory the definition of one DiffServ aggregate per COCR Class of Service (CoS) would be possible (resulting in 12 aggregates), in practice a smaller number of DiffServ aggregates improves the scalability and reduces the complexity. In this case the application CoS need to be mapped by a classifier into the suitable DiffServ aggregates. Since all COCR CoS have different demands for maximum latency, an aggregation into fewer DiffServ aggregates implies also an increase of the required bandwidth, since the latency of the most demanding service in a DiffServ aggregate has to be met since DiffServ is not distinguishing within an aggregate. In other words services which could tolerate a longer latency need to

be transmitted in fewer time (i.e. the time of the most demanding service) what results in a higher demand in terms of data rate. For a DiffServ QoS approach also appropriate estimation and dimensioning of the network capacities is essential and requires a good model for the prediction of the amount of traffic to be transported including an additional buffer for unexpected traffic bursts. Such an (over)dimensioning on the other hand can also mean a waste of resources if capacity is strictly allocated per aggregate class and cannot be shared among different aggregates and considering the highly bursty traffic profile.

On the other hand a DiffServ architecture has significant advantages over an IntServ approach which outweigh the aforementioned drawbacks. Most important of all the issues with scalability do not exist here since only aggregates have to be treated instead of single flows. DiffServ is such much more suitable for the highly populated global ATM network under consideration with respect to this. Moreover a change of the end-to-end path, as can happen due to intra- and inter-technology handovers in this highly mobile scenario is not an issue here since no re-establishment of the RSVP tunnels is required anymore. Also the signalling overhead of IntServ for session initiation and keepalive can be saved while saving also the time for flow establishment which is beneficial for the overall delay profile.

2.4 Flow Identification

As was shown in other work (NEWSKY, 2009), routing decisions should be taken per flow, not per packet, e.g. due to problem of different latencies when messages are sent over different links, passing of packets, impact on TCP retransmission mechanism and reordering as well as load oscillations.

To identify the flow that a Layer 3 packet belongs to, the flow session identifier shall check the 5-tupel consisting of the IP source and destination address, source and destination transport layer ports and transport protocol.

In contrast to IPv4, which only allowed the identification of a traffic aggregate by the DSCP field or a particular flow, indicated by the 5-tupel, IPv6 additionally allows marking of single or aggregate flows via the flow label header field. Since also safety critical messages need to be exchanged in the aeronautical scenario, also security mechanisms such as IPSec may be applied. While encryption (IPSec Encapsulated Security Payload) may not be applied in all cases, means for authentication (IPSec Authentication Header) may be present. Considering the possibility to use IPSec also in tunnel mode, the flow identification can be done based on either inner or outer header (w.r.t. the tunnel) and before or after IPSec processing. Fig. 9 shows IPSec ESP tunnel mode for IPv6 datagrams.

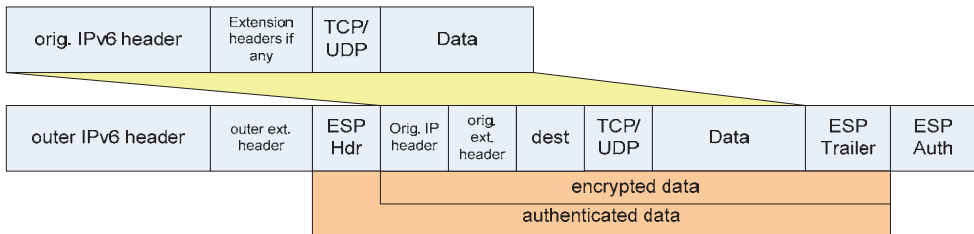


Fig. 9. IPv6 IPSec in ESP-tunnel mode.

In IPSec tunnel mode the inner header fields are not accessible in ESP mode since they are encrypted. Identification of the 5-tupel is not possible in these cases since also the UDP and

TCP headers, which are part of the 5-tuple, are located in the encrypted part. Though encryption is currently not envisaged for operational messages it is beneficial to do the flow identification before the IPSec processing since here identification of the 5-tuple can be done in any case.

In the case of dedicated Security Gateways (SG), the flow label assignment in the inner header must be done there, since after processing by the SG inner header fields must not be changed anymore. In case the SG is not implementing flow classification abilities, the flow label identifier in the router can only do a classification in case the inner header fields are visible (i.e. not encrypted) and only assign a flow label to the outer header.

In case the inner header fields are not visible no flow identification based on the original 5-tuple is possible.

For IPSec tunnel endpoints in the end systems (ES), it is the ES responsibility to set the correct values of the traffic class and flow label. As in the case of dedicated SGs, the subsequent routers can only do a classification in case the inner header fields are visible and flow label assignment can only be applied to the outer header fields.

The flow identifier also has to assign packets coming from the non-operational domain (AAC/APC) accordingly for a non-operational flow so the routing decision functionality can treat these packets separately. The differentiation between operational and non-operational domain can be accomplished either IP address based or based on the physical interface:

IP address

In this case the OP and NON-OP traffic is distinguished only by the 5 tuple in the packet headers. This is however imposing a risk for spoofing attacks where these header fields are maliciously modified by an attacker.

Physical interface

In this case the IR has different physical connector interfaces to the OP and NON-OP domain. Due to the physical separation, it is ensured that NON-OP data can in no case interfere with OP data, since a NON-OP packet is always unambiguously identified and treated.

For assigning the correct aggregate class, the flow identifier additionally needs management information in form of DiffServ tables to map packets correctly to code points and flows IDs. These tables are specified in the management plane and allow configuration of the mapping

3. Conclusions on QoS architecture

In summary the following observations for the QoS architecture in an ATM can be made from the aspects briefly presented before:

A flow-oriented architecture such as IntServ would have the feature of guaranteeing a certain end-to-end behaviour, but is not suitable w.r.t. the bursty traffic profile, having only spurious transmission of single messages which have also only small size. The signalling overhead is considerable w.r.t. the small message payloads and also the additional time demand for a session initiation is considerable w.r.t. the latency requirements. A flow-oriented QoS architecture such as IntServ is thus no preferable solution for application in an ATM.

The alternative QoS architecture matching better with the given scenario is thus DiffServ. For deployment of a DiffServ QoS architecture several design parameters have to be kept in

mind, in particular the correct dimensioning of the resource trunks, mapping of application CoS into aggregate classes and priority scheduling. The main benefits here are the scalability also for a large and global ATM network. Also a change in the network point of attachment, e.g. due to a handover are not an issue here. The data volume and signalling delay overheads of IntServ can be saved here as well. For an integration of operational with non-operational services in the same network, however further specification of the mechanisms ensuring a safe separation of these two domains is required as well as deployment of mechanisms for CC, CAC and flow control of the NON-OP services.

Independently of the selected QoS model, the aeronautical QoS framework requires a solid and mature signalling framework, which can be easily derived from the experience acquired in ETSI BSM and IEEE 802.21 standardisation bodies. In particular, the extension of the SI-SAP primitives to match the aeronautical service requirements and the IR/IMR interaction are expected to be promising to help develop a fully QoS-oriented aeronautical architecture. On the other hand, the joint use of the aforementioned ones and the MIH framework should also guarantee an important support to efficiently manage the available transmission links and perform their selection accordingly.

4. Acknowledgments

The research leading to these results has been partially funded by the European Community's Seventh Framework Programme (FP7/2007-2013) under Grant Agreement n°233679. The SANDRA project is a Large Scale Integrating Project for the FP7 Topic AAT.2008.4.4.2 (Integrated approach to network centric aircraft communications for global aircraft operations). The project has 31 partners and started on 1st October 2009.

5. References

- Ali, M.; Xu, K.; Pillai, K. & Hu Y. F. (2011). Common RRM in Satellite-Terrestrial based Aeronautical Communication Networks. In: *Third International ICST Conference on Personal Satellite Services (PSATS 2011)*, Malaga, Spain, February 2011
- Blake, S.; Black, D.; Carlson, M.; Davis, E.; Wang, Z. & Weiss, W. (1998). RFC 2475: An Architecture for Differentiated Services, *IETF*, 1998
- EUROCONTROL (2007), FAA, Communications Operating Concept and Requirements for the Future Radio System (COCR), EUROCONTROL/FAA, 2007
- Hitoshi, S. (1998) *Teletraffic Technologies in ATM Networks*, Artech House.
- ICAO (2009), International Civil Aviation Organization: Manual for the ATN using IPS Standards and Protocols (Doc 9896). February 2009, 1st edition, Unedited Advance version
- ITU-T Recommendation G.1010 (2001) End-user multimedia QoS categories, Nov. 2001.
- Marchese, M. (2007), *QoS over Heterogeneous Networks*, John Wiley & Sons Ltd, 2007.
- NEWSKY (Networking the Sky for Aeronautical Communications) (2009) - www.newskey-fp6.eu
- Nichols, K.; Blake, S.; Baker, F. & Black, D. (1998) RFC 2474: Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers, *IETF*, 1998.
- Plass, S.; Kissling, C.; De Cola, T. & Van Wambeke, N. (2011): The SANDRA communications concept - Future aeronautical communications by seamless

- networking. In: Proc. Third International ICST Conference on Personal Satellite Services (PSATS 2011), 17.-18. Feb. 2011, Malaga, Spanien.
- SANDRA (Seamless Aeronautical Networking through integration of Data links Radios and Antennas) (2011) - www.sandra.aero
- Tannenbaum, A. (2002) Computer Networks, Prentice Hall Professional Technical Reference, 2002.
- Wroclawsky, J. & Braden, R. (Ed.) (1997). RFC 2210: The Use of the Resource Reservation Protocol with Integrated Services, *IETF*, 1997.
- Zhang, L.; Berson, S.; Herzog, S. & Jamin, S. Braden, R. (Ed.) (1997). RFC 2205: Resource ReSerVation Protocol, (RSVP) --- Version 1 Functional Specification, *IETF*, 1997



Future Aeronautical Communications

Edited by Dr. Simon Plass

ISBN 978-953-307-625-6

Hard cover, 378 pages

Publisher InTech

Published online 26, September, 2011

Published in print edition September, 2011

There are well-founded concerns that current air transportation systems will not be able to cope with their expected growth. Current processes, procedures and technologies in aeronautical communications do not provide the flexibility needed to meet the growing demands. Aeronautical communications is seen as a major bottleneck stressing capacity limits in air transportation. Ongoing research projects are developing the fundamental methods, concepts and technologies for future aeronautical communications that are required to enable higher capacities in air transportation. The aim of this book is to edit the ensemble of newest contributions and research results in the field of future aeronautical communications. The book gives the readers the opportunity to deepen and broaden their knowledge of this field. Today's and tomorrow's problems / methods in the field of aeronautical communications are treated: current trends are identified; IPv6 aeronautical network aspect are covered; challenges for the satellite component are illustrated; AeroMACS and LDACS as future data links are investigated and visions for aeronautical communications are formulated.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Christian Kissling and Tomaso de Cola (2011). Quality of Service Management and Interoperability, Future Aeronautical Communications, Dr. Simon Plass (Ed.), ISBN: 978-953-307-625-6, InTech, Available from: <http://www.intechopen.com/books/future-aeronautical-communications/quality-of-service-management-and-interoperability>

INTECH

open science | open minds

InTech Europe

University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China

Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

© 2011 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the [Creative Commons Attribution-NonCommercial-ShareAlike-3.0 License](#), which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited and derivative works building on this content are distributed under the same license.