

REPORT ITU-R M.2114

**Key technical and operational characteristics for access technologies
to support IP applications over land mobile systems**

(Question ITU-R 223/8)

(2007)

1 Introduction

With the high demand for Internet Protocol (IP) applications growing consistently at such high rates, the need to develop more efficient systems able to support new internet features will be a necessity.

It is anticipated that users' demands for mobile services are increasing with greater diversity and complexity for the mobile communications market. Diverse and complex services are expected to have quite different traffic characteristics and QoS levels in comparison to voice traffic or text. In addition, increasing demands on multimedia services causes further more complexity in designing and developing the future mobile systems. Future mobile technologies should support diverse services with different traffic characteristics which are pervasive regardless of air interface technologies. IP applications are accepted as one of solutions to provide affordable and compatible services as used in wired communication systems. Therefore, IP applications over mobile systems will be important in the future due to global interoperability.

IP applications over mobile systems are supported by existing standards that use IP to send data. However, to support enhanced IP applications over mobile systems such as seamless delivery of multimedia data services, several significant technical characteristics in the radio interface and access networks should be considered.

The All IP Network (AIPN) is an IP-based network providing common capabilities that are independent to the type of service being provided and the access system being used. Convergence to IP technology within the AIPN system design should be considered for the system as a whole with minimum duplication of functionality.

During the initial stages of AIPN introduction it is likely that earlier systems and terminals will exist in parallel with AIPNs.

Therefore, the AIPN should be able to support access networks containing Circuit Switched (CS) terminals and accommodate access systems based on CS.

Interworking and interconnection with CS networks should be provided.

The AIPN should be designed to enable efficient coexistence with earlier PS domains.

2 Scope

This Report defines the essential technical and operational characteristics needed to support IP applications over mobile systems.

3 References

3.1 Related ITU-R Recommendations and Reports

Recommendation ITU-R M.1079 – Performance and quality of service requirements for IMT-2000 access networks

Recommendation ITU-R M.1741 – Methodology for deriving performance objectives and its optimization for IP packet applications in the mobile-satellite service.

Recommendation ITU-R S.1711 – Performance enhancements of transmission control protocol (TCP) over satellite networks (pre-published)

Recommendation ITU-R M.1645 – Framework and overall objectives of the future development of IMT-2000 and systems beyond IMT-2000

Report ITU-R F.2058 – Design techniques applicable to broadband fixed wireless access (FWA) systems conveying Internet protocol (IP) packets or asynchronous transfer mode (ATM) cells.

3.2 IETF

The following references provide supporting information and do not refer to specific implementations.

IETF RFCs are available from the IETF web site, <http://www.ietf.org/rfc>.

RFC 791 Internet Protocol, Sept. 1981.

RFC 793 Transmission Control Protocol, September 1981.

RFC 1035 Domain Names – Implementation and Specification, November 1987.

RFC 1661 The Point-to-Point Protocol (PPP), July 1994.

RFC 2002 IP Mobility support, October 1996.

RFC 2068 Hypertext Transfer Protocol -- HTTP/1.1, January 1997.

RFC 2131 Dynamic Host Configuration Protocol, March 1997.

RFC 2460 Internet Protocol, Version 6 (IPv6) Specification, December 1998.

RFC 3095 Borman, et al, “RObust Header Compression (ROHC): Framework and four profiles: RTP, UDP, ESP, and uncompressed”, July 2001.

RFC 3545 Koren, et al, “Enhanced Compressed RTP (CRTP) for links with High Delay, Packet Loss and Reordering”, July 2003.

4 Definition

AIPN: A collection of entities that provide a set of capabilities for the provision of IP services to users based on IP technology where various access systems can be connected. The AIPN provides a set of common capabilities (including mobility, security, service provisioning, charging and QoS) which enable the provision of services to users and connectivity to other external networks. An AIPN requires one or more connected access systems to allow users to access the AIPN.

Mobile IPv4: Provides mobility support for IPv4.

Mobile IPv6: Provides mobility support for IPv6.

5 Abbreviations

AAA	Authentication, authorization, and accounting
AIPN	All IP network
BR	Border router
CN	Correspondent node
DHCP	Dynamic Host Control Protocol
E2E QoS	End-to-end quality of service
HA	Home agent
HAAA	Home AAA
HDB	Home data base
IETF	Internet Engineering Task Force
IMS	IP multimedia subsystem
LAC	Link access control
MAC	Media access control
MIPv4	Mobile IPv4
MIPv6	Mobile IPv6
MMD	Multi-media domain
MS	Mobile station
PDF	Policy decision function
PDSN/AGW	Packet data serving node/access gateway
PDU	Protocol data unit
PPP	Point-to-Point Protocol
SLA	Service level agreement
VAAA	Visited AAA
VDB	Visited data base
P-CSCF	Proxy-call session control function
PDSN	Packet data serving node
S-CSCF	Serving-call session control function
RAN	Radio access network

6 Basic capabilities

6.1 Key characteristics

6.1.1 Technical characteristics

Essential key characteristics to support IP applications over mobile systems include:

- Compatibility

- Transparency
- Scalability and efficiency
- Security
- High speed burst traffic
- Low cost per bits
- Seamless delivery services
- Various grade of services.

IP applications over mobile systems must be able to remain compatible to all levels used for the standard non-mobile IP applications. IP applications over mobile systems should therefore remain “invisible” for higher level protocols and applications. Higher layers should continue to function normally even while the mobile has altered its point of attachment to the network. As new features are developed, the sub-network (subnet) should also function with maximum efficiency and minimum complexity.

As the user rates of mobile communications increase, IP applications over mobile systems should be scalable over the large numbers utilizing the Internet. These subnets should also be able to provide support for advanced Internet features as Multicasting and quality of service (QoS). Identification during attachment, authentication and authorization are required in order to protect against remote attacks. As information is transmitted from node to node, the location of a mobile node must be authenticated to maintain security.

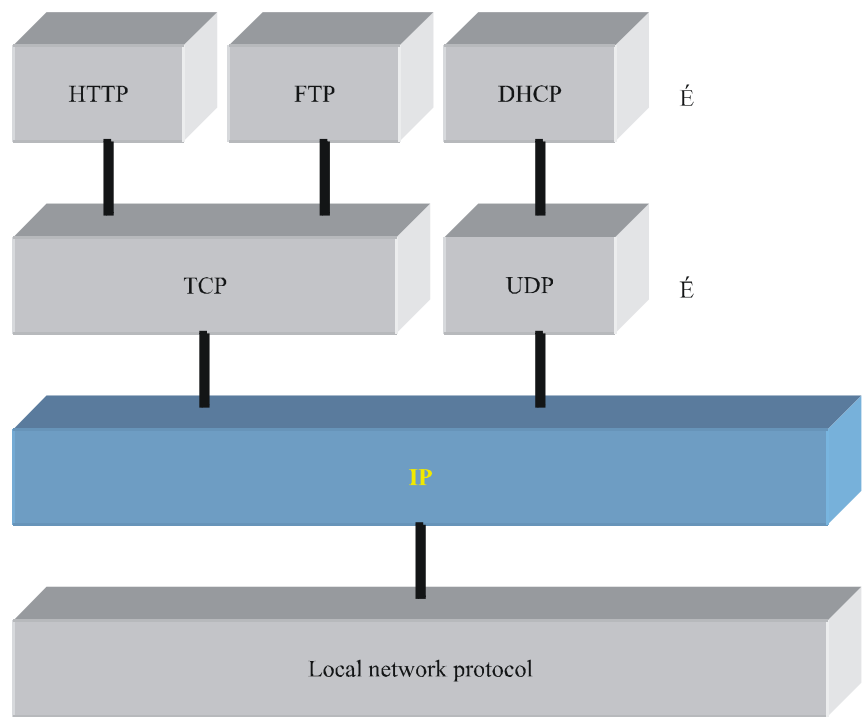
6.1.2 Operational characteristics

The AIPN should be under the control of the operator of the AIPN. The AIPN should provide common mechanisms for the AIPN operator to control access to and usage of AIPN resources.

6.2 Encapsulation

The Internet Protocol Version 4, IPv4, is described in IETF RFC 791. Internet Protocol Version 6, IPv6, is described in IETF RFC 2460. Figure 1 illustrates the protocol hierarchy.

FIGURE 1
Protocol hierarchy



Rap 2114-01

All IP application related data is exchanged between hosts in IP packets. For example, the Hypertext Transfer Protocol (RFC 2068) protocol data unit, PDU, is encapsulated in a Transmission Control Protocol (RFC 793) PDU, which is encapsulated in an IP PDU, and then finally in a local network PDU. In the context of mobile systems, the local network may be considered the radio interface.

6.3 Address management

To participate effectively in an IP network, a device must possess an IP address. IP packets are routed based on the destination address field within the packet header.

IPv4 addresses are 32-bit binary numbers. Most commonly, they are expressed by considering the 32 bits as four octets, converting each octet to decimal, then separating these numbers with dots to create dotted decimal notation. For example, the address 17.112.152.32 is the dotted decimal notation of:

```

00010001011110000100110000010000
00010001011110000100110000010000
  
```

IPv6 addresses are 128-bit binary numbers. If these addresses are expressed in human readable form they are most commonly expressed in hexadecimal with colons to aid legibility, for example:

fe80:0000:0000:0000:020a:95ff:fef3:2e91.

In order to interwork between Internet and wireless systems, the mapping between IP addresses and MAC addresses should be considered. For example, IP address of 255.255.255.255 is mapped to the MAC address of ff:ff:ff:ff:ff:ff for data broadcasting in wireless LAN systems.

6.4 Maximum transmission units (MTUs) and IP fragmentation

IPv4 packets (datagrams) vary in size, from 20 bytes (the size of the IPv4 header alone) to a maximum of 65 535 bytes. Subnets need not support maximum-sized (64 kB) IP packets, as IP provides a scheme that breaks packets that are too large for a given subnet into fragments that travel as independent IP packets and are reassembled at the destination. The maximum packet size supported by a subnet is known as its maximum transmission unit (MTU).

6.5 Multicasting/Broadcasting

The Internet model includes “multicasting”, where IP packets are sent to all the members of a multicast group. Multicast is an option in IPv4, but a standard feature of IPv6. IPv4 multicast is currently used by multimedia, teleconferencing, gaming, and file distribution (web, peer-to-peer sharing) applications, as well as by some key network and host protocols (e.g. RIPv2, OSPF, NTP). IPv6 additionally relies on multicast for network configuration (DHCP-like autoconfiguration) and link-layer address discovery (replacing ARP). In the case of IPv6, this can allow autoconfiguration and address discovery to span across routers, whereas the IPv4 broadcast-based services cannot without ad-hoc router support.

Alternatively MBMS (multimedia broadcast and multicast service) has been developed in mobile systems. The broadcast mode in MBMS differs from the multicast mode in that there is no specific requirement to activate or subscribe to the MBMS in broadcast mode. The broadcast mode is intended to efficiently use radio/network resources e.g. data is transmitted over a common radio channel. Data is transmitted in the broadcast service area as defined by the network. The reception of the traffic in the broadcast mode is not guaranteed.

6.6 Bandwidth on demand (BoD) subnets

Some subnets allow a number of subnet nodes to share a channel efficiently by assigning transmission opportunities dynamically. Transmission opportunities are requested by a subnet node when it has packets to send. The subnet schedules and grants transmission opportunities sufficient to allow the transmitting subnet node to send one or more packets (or packet fragments). These subnets are referred to as Bandwidth on demand (BoD) subnets.

6.7 Bandwidth asymmetries

Some subnets may provide asymmetric bandwidth (or may cause TCP packet flows to experience asymmetry in the capacity) and the Internet protocol suite will generally still work fine. However, there is a case when such a scenario reduces TCP performance. Since TCP data segments are “clocked” out by returning acknowledgments, TCP senders are limited by the rate at which ACKs can be returned. Therefore, when the ratio of the available capacity of the Internet path carrying the data to the bandwidth of the return path of the acknowledgments is too large, the slow return of the ACKs directly impacts performance. Since ACKs are generally smaller than data segments, TCP can tolerate some asymmetry, but as a general rule, designers of subnets should be aware that subnets with significant asymmetry can result in reduced performance, unless issues are taken to mitigate this (RFC3449).

Several strategies have been identified for reducing the impact of asymmetry of the network path between two TCP end hosts, e.g. (RFC3449). These techniques attempt to reduce the number of ACKs transmitted over the return path (low bandwidth channel) by changes at the end host(s), and/or by modification of subnet packet forwarding. While these solutions may mitigate the performance issues caused by asymmetric subnets, they do have associated cost and may have other implications. A fuller discussion of strategies and their implications is provided in (RFC3449).

7 QoS considerations

Some of the general requirements for QoS should include:

- End-to-end QoS.
- Wide range of QoS-enabled services.
- End-to-end QoS should be supported within the local domain and across different network domains as well.
- Appropriate levels of QoS should be maintained even when Internet features, such as multicasting is applied.

The QoS network architecture should be flexible enough to support different QoS control mechanisms, which are defined in different access technology environments. The network should be able to provide mechanisms which could perform traffic and congestion control. Traffic control would include functions such as network resource management, packet marking, traffic shaping, and packet scheduling. Congestion control refers to functions which include packet discarding or explicit congestion notification. The network should also provide methods in which to negotiate QoS at both transport and service layers, and allow dynamic alterations of the QoS parameters. The network should allow operators to implement QoS policy control, where the policy-based management would be extended across multiple domains to ensure QoS.

It is generally recognized that specific service guarantees are needed to support real-time multimedia, toll-quality telephony, and other performance-critical applications¹. For some important services with strict end-to-end QoS requirements, such as conversational speech or streaming video, the QoS should be assured in case of integrated networking with different IP network domains or backbone networks. Most likely this would be ensured on a per service basis of specific flows of IP packets having been identified by the service.

There are at least two architectural approaches to providing mechanisms for QoS support in the Internet. The IP integrated services (Intserv) (RFC1633) is a working group formed to standardize a new resource allocation architecture and new service models for the Internet. The Intserv model includes two main components: the traffic control and the ReSerVation Protocol. Flows are identified by a flow specification (flowspec), which creates a stateful association between individual packets by matching fields in the packet header. Capacity is reserved for the flow, and appropriate traffic conditioning and scheduling is installed in routers along the path. The ReSerVation Protocol (RSVP) (RFC2205) (RFC2210) is usually, but need not necessarily be, used to install the flow QoS state. RSVP is a control signalling protocol which requires the introduction of states for specific information flows, although reservation states are “soft” in that they are regularly renewed by messages sent from the initiator of the reservation request. If not renewed, the reservations are timed-out.

Intserv defines two services, in addition to the Default (best effort) service:

- 1 *Guaranteed Service (GS)* (RFC2212) offers hard upper bounds on delay to flows that conform to a traffic specification (TSpec). It uses a fluid-flow model to relate the TSpec and reserved bandwidth (RSpec) to variable delay. Non-conforming packets are forwarded on a best-effort basis.
- 2 *Controlled Load Service (CLS)* (RFC2211) offers delay and packet loss equivalent to that of an unloaded network to flows that conform to a TSpec, but no hard bounds. Non-conforming packets are forwarded on a best-effort basis.

¹ Recommendation ITU-R M.1079 – Performance and quality of service requirements for IMT-2000 access networks.

The other architectural approach is called the IP differentiated services (Diffserv) (RFC2475). This is an alternative resource allocation scheme, which provides service differentiation by dividing the traffic into different classes at the edge of a network by nodes classified as boundary nodes. A boundary node classifies each incoming packet into a particular traffic class. Diffserv provides a scalable approach, but it in itself does not provide guaranteed QoS. Diffserv does not make any per-flow reservations. Rather, it provides QoS for aggregates of flow. Resources are assured by the prioritization for traffic classes.

Mobile systems face inherent tradeoffs between delay, throughput, reliability, and cost. Some subnets have parameters that manage bandwidth, internal connection state, and the like. Therefore, the following subnet capabilities may be desirable, although some might be trivial or moot if the subnet is a dedicated point-to-point link.

- 1 The subnet should have the ability to reserve bandwidth for a connection or flow and schedule packets accordingly.
- 2 Bandwidth reservations should be based on a one- or two-token bucket model, depending on whether the service is intended to support constant-rate or bursty traffic.
- 3 If a connection or flow does not use its reserved bandwidth at a given time, the unused bandwidth should be available for other flows.
- 4 Packets in excess of a connection or flow's agreed rate should be forwarded as best-effort or discarded, depending on the service offered by the subnet to the IP layer.
- 5 If a subnet contains error control mechanisms (retransmission and/or FEC), it should be possible for the IP layer to influence the inherent tradeoffs between uncorrected errors, packet losses, and delay. These capabilities at the subnet/IP layer service boundary correspond to selection of more or less error control and/or to selection of particular error control mechanisms within the subnet.
- 6 The subnet layer should know, and be able to inform the IP layer, how much fixed delay and delay jitter it offers for a flow or connection. If the Intserv model is used, the delay jitter component may be best expressed in terms of the TSpec/RSPEC model described in (RFC2212).
- 7 Support of the Diffserv class selectors (RFC2474) suggests that the subnet might consider mechanisms that support priorities.

8 End to end requirements for AIPN with access system

8.1 Introduction

The AIPN with access system should be capable of fulfilling certain end-to-end characteristics requirements. These requirements, if met by the system, will enable good end-user performance for a variety of end-user services including, but not limited to:

- Real-time, interactive applications, e.g. voice, video and real-time gaming applications.
- Non-real time, interactive applications, e.g. Web browsing, remote login, chat.
- Media streaming applications.
- Conversational services.

The present requirement applies to the entire end-to-end path from the user terminal to the other end-host or server including radio network, core network, and backbone network processing, buffering and propagation delays. Since the delay depends on the location of the end-host, the requirements should be considered as preferred values when the user is within the same continent.

8.2 Delay characteristics

The TCP sender bases its retransmission timeout (RTO) on measurements of the round trip delay experienced by previous packets. This allows TCP to adapt automatically to the very wide range of delays found on the Internet. If the path delay variance is high, TCP sets an RTO that is much larger than the mean of the measured delays. If the packet loss rate is low, the large RTO is of little consequence, as timeouts occur only rarely. Conversely, if the path delay variance is low, then TCP recovers quickly from lost packets; again, the algorithm works well. However, when delay variance and the packet loss rate are both high, these algorithms perform poorly, especially when the mean delay is also high.

Because TCP uses returning acknowledgments as a “clock” to time the transmission of additional data, excessively high delays (even if the delay variance is low) also affect TCP’s ability to fully utilize a high-speed transmission pipe. It also slows the recovery of lost packets, even when delay variance is small.

Mobile systems should therefore minimize all three parameters (delay, delay variance, and packet loss) as much as possible. Often these parameters are inherently in conflict. For example, on a mobile radio channel, retransmission (ARQ) and/or forward error correction (FEC) can be used to trade off delay, delay variance, and packet loss in an effort to improve TCP performance. While ARQ increases delay variance, FEC does not. However, FEC (especially when combined with interleaving) often increases mean delay, even on good channels where ARQ retransmissions are not needed and ARQ would not increase either the delay or the delay variance.

The tradeoffs among these error control mechanisms and their interactions with TCP can be quite complex, and are the subject of much ongoing research. It is therefore recommend that mobile systems provide as much flexibility as possible in the implementation of these mechanisms, and provide access to them as discussed above in the section on QoS.

8.3 Round-trip delay requirement

The round-trip delay requirement for small IP packets (0 payload) should be defined such that it will enable real-time applications at good quality when both end-hosts reside within the same continent.

Acceptable value is 50-70 ms average round-trip delay.

8.4 Packet loss requirement

The packet loss at the IP layer requirement should be defined such that it will enable a certain maximum download data rate for TCP based applications when both end-hosts reside within the same continent and the radio conditions are preferable.

Acceptable value is 0.001% in good radio conditions but maximum 0.1% packet loss ratio.

NOTE 1 – The maximum value for packet loss may also be sufficient for UDP applications.

8.5 Delay jitter requirement

The delay jitter requirement should be defined such that it will not harm real-time applications and will not cause significant TCP degradation due to spurious timeouts.

Acceptable value for delay jitter is 25 ms (for real-time gaming), for TCP based download applications occasional bursts of up to 50 ms are still tolerated.

9 Supporting mobility

9.1 Introduction

Internet users are increasingly mobile. Not only are many Internet nodes laptop computers, but pocket organizers and mobile embedded systems are also becoming nodes on the Internet. These nodes may connect to many different access points on the Internet over time, and they expect this to be largely transparent to their activities. Except when they are not connected to the Internet at all, and for performance differences when they are connected, they expect that everything will “just work” regardless of their current Internet attachment point or local subnet technology.

Changing a host’s Internet attachment point involves one or more of the following steps.

First, if use of the local subnet is restricted, the user’s credentials must be verified and access granted. There are many ways to do this. A trivial example would be an “Internet café” that grants physical access to the subnet for a fee. Subnets may implement technical access controls of their own. It is common practice for both cellular telephone and Internet service providers (ISPs) to agree to serve one another’s users; RADIUS (RFC2865) is the standard method for ISPs to exchange authorization information.

Second, the host may have to be reconfigured with IP parameters appropriate for the local subnet. This usually includes setting an IP address, default router, and domain name system (DNS) servers.

On multiple-access networks, the Dynamic Host Configuration Protocol (DHCP) (RFC2131) is almost universally used for this purpose. On PPP links, these functions are performed by the IP Control Protocol (IPCP) (RFC1332).

Third, traffic destined for the mobile host must be routed to its current location. This roaming function is the most common meaning of the term “Internet mobility”.

Internet mobility can be provided at any of several layers in the Internet protocol stack, and there is ongoing debate as to which is the most appropriate and efficient. Mobility is already a feature of certain application layer protocols; the Post Office Protocol (POP) (RFC1939) and the Internet Message Access Protocol (IMAP) (RFC3501) were created specifically to provide mobility in the receipt of electronic mail.

Mobility can also be provided at the IP layer (RFC3344). This mechanism provides greater transparency, viz., IP addresses that remain fixed as the nodes move, but at the cost of potentially significant network overhead and increased delay because of the sub-optimal network routing and tunneling involved.

Some subnets may provide internal mobility, transparent to IP, as a feature of their own internal routing mechanisms. To the extent that these simplify routing at the IP layer, reduce the need for mechanisms like Mobile IP, or exploit mechanisms unique to the subnet, this is generally desirable. This is especially true when the subnet covers a relatively small geographic area and the users move rapidly between the attachment points within that area. Examples of internal mobility schemes include Ethernet switching and intra-system handover in cellular telephony.

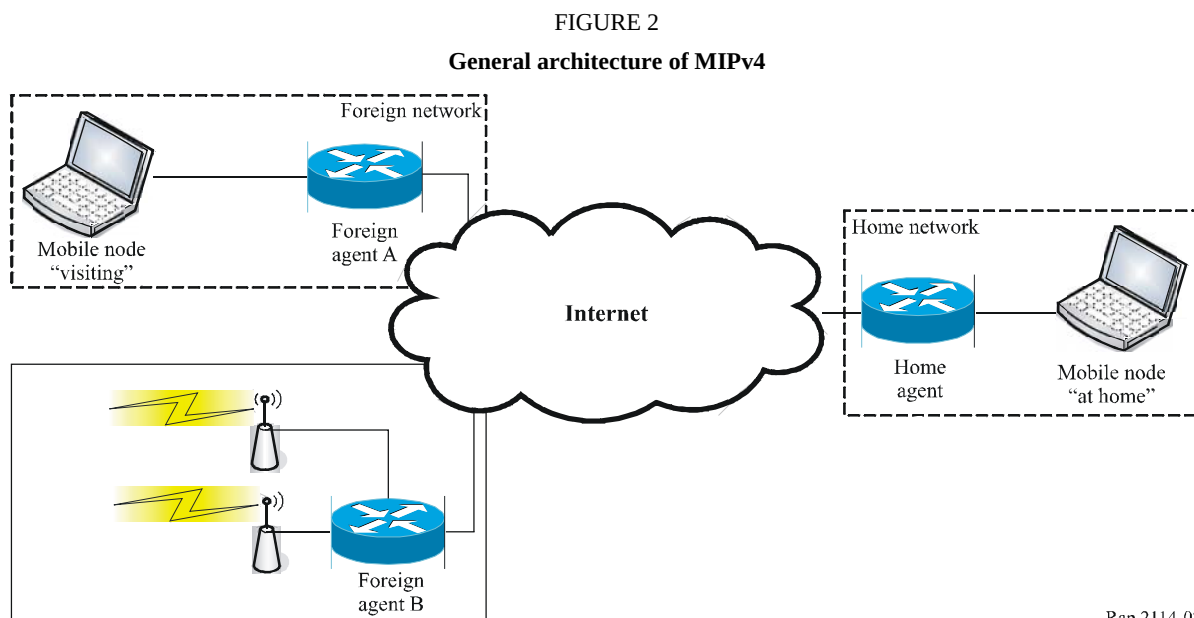
However, if the subnet is physically large and connects to other parts of the Internet at multiple geographic points, care should be taken to optimize the wide-area routing of packets between nodes on the external Internet and nodes on the subnet. This is generally done with “nearest exit” routing strategies. Because a given subnet may be unaware of the actual physical location of a destination on another subnet, it simply routes packets bound for the other subnet to the nearest router between the two. This implies some awareness of IP addressing and routing within the subnet. The subnet may wish to use IP routing internally for wide area routing and restrict subnet-specific routing to constrained geographic areas where the effects of suboptimal routing are minimized.

Key AIPN capabilities are characterized as follows:

- Able to accommodate fixed access systems and to inter-work with fixed networks in order to provide seamless services over fixed/mobile converged networks, e.g. charging and providing supplementary services.
- Able to support end-user, terminal and session mobility.
- Able to provide seamless terminal mobility within and across access systems. The user should experience no disruption in the service due to terminal mobility.
- Able to maintain a service during a change in access system, with no perceivable interruption from a user perspective.
- Able to support adaptation of services to the capabilities provided by the access systems during terminal mobility.
- Able to support terminal mobility based on criteria including radio conditions, service requirements, user preferences and operator policies.
- Able to notify end users of the degradation in cases when there is a degradation in service quality due to terminal mobility.
- Able to provide appropriate mechanisms to enable users to connect to the AIPN through multiple access systems.
- Able to support handover between CS voice services and AIPN equivalent services (e.g. Voice over IP).

9.2 Mobile IP

The IETF have developed mobility support for IP, generically referred to as Mobile IP. In this Report we refer to MIPv4 and MIPv6 when appropriate to distinguish between mobility support for IPv4 and IPv6. Figure 2 depicts the general architecture of MIPv4.



Rap 2114-02

A brief description of an example implementation of MIPv4 is as follows:

Mobile node: A node that can change its point-of-attachment from one link to another by, for example, moving between geographic locations, while maintaining ongoing communications. The mobile node uses only its “permanent” IP address – that address assigned by its home network.

Home agent: Resides in the home network; maintains a security relationship with the foreign agent in the visited network; intercepts packets destined to the mobile node’s home address and “tunnels” them to mobile node via the foreign agent.

Foreign agent: Resides in the visited network; maintains a security relationship with the home agent in the home network; acts as the default router for the mobile node while visiting; de-tunnels packets relayed from the home agent.

Further details may be found in RFC 2002.

When IP applications are supported over a mobile system, the mobile node may change its point of attachment by moving between geographic locations, causing the radio sub-system to handover to another cell site. This may or may not cause a handover to a different foreign agent.

9.3 Mobility and handover

The system should support mobility across the cellular network and should be optimized for low mobile speed from 0 to 15 km/h. Higher mobile speed between 15 and 120 km/h should be supported with high performance. Mobility across the cellular network should be maintained at speeds from 120 km/h to 350 km/h (or even up to 500 km/h depending on the frequency band). Voice and other real-time services supported in the CS domain in previous systems should be supported via the PS domain with at least equal quality as supported by previous systems (e.g. in terms of guaranteed bit rate) over the whole of the speed range.

The impact of intra system handovers on quality (e.g. interruption time) should be less than or equal to that provided by CS domain handovers in previous systems.

Mobile speed above 250 km/h, such as a high-speed train environment, represents a special case. In such case a special scenario applies for issues such as mobility solutions and channel models. For the physical layer parameterization the system should be able to maintain the connection up to 350 km/h, or even up to 500 km/h depending on the frequency band.

The system should also support techniques and mechanisms to optimize delay and packet loss during intra system handover.

The wireless network system for future mobile communication may vary, according to radio network topology, it can be divided into cellular network, distributed network, etc, on the other hand, according to the working function, the mobile user should have the ability to identify multiple radio access system connected IP network, and fast handover freely in different radio access system actively/passively.

10 Security requirements

10.1 Introduction

There are several security issues related to supporting IP applications whether the applications are run over a mobile system or not. Charging for an application often requires the user to be authenticated to ensure the user is allowed to access the application. In addition, the user may also need to authenticate the application to ensure that personal details, e.g. user’s current location or

bank details, are not revealed to the wrong application. Many applications will also have the need to encrypt and protect the integrity of the data that is exchanged between the mobile and application. This ensures that the data sent between the mobile and application can not be read by a man-in-the-middle or modified without it being detected by the receiving party.

The following general objectives should be considered in IP applications over mobile systems:

- Confidentiality
- Integrity
- Accountability
- Availability
- Non-repudiation
- Privacy

10.2 Authentication

During the process of handover between foreign agents, it should be avoided that new foreign agents which have not yet been authorized receive data sent from previous FA to the mobile node. For some agent handover algorithms, it is possible that an attacker be provided with an opportunity to act as a pseudo new foreign agent to intercept and capture the data which should normally be transmitted to the mobile node, if the new foreign agent send a “binding update” message to the previous foreign agent before getting the authentication from the mobile node; so it is required that new foreign agents should get the authentication from mobile node first to enhance the security performance of handovers between agents. Mobile nodes could send a “registration request” message to the new foreign agent and simultaneously send a “binding update” message to the previous foreign agent and inform it to cache the data. Subsequently the mobile node performs the authentication procedure with the new foreign agent; and the cached data would be transferred from the previous FA to the new one if the authentication is valid so that a smooth and safe handover mechanism is achieved.

Also, security should be considered along with route optimization. Mobile IP finds a solution to the problem of triangular routing; a mobile node could inform its correspondent node of the care-of address so as to communicate with it without passing through the home agent. But under this circumstance interactive authentication is required for both the mobile node and its correspondent node; in other words, a pair of secret keys are needed. On the other hand, allocating a pair of secret keys for every mobile node and its correspondent node would not be feasible when there are a great amount of mobile nodes and correspondent nodes; but it is practicable to assign a pair of secret keys for every mobile node and its HA.

10.3 Privacy

Privacy is a fundamental component of security. Privacy can be viewed as the right of individuals to protect the collection and storage of information related to them and to control the disclosure of that information. By extension, privacy may also be associated with certain technical means (e.g. cryptography) to safeguard from unauthorized disclosure to anyone other than the intended parties, so that only the explicitly authorized parties can interpret the content exchanged among them.

Privacy and confidentiality are terms often used synonymously. However, it should be noted that ITU-T Recommendation X.805 differentiates privacy and data confidentiality. Privacy relates to the protection of the association of the identity of users and the activities performed by them (i.e. online purchase habits, Internet sites visited, etc), whereas, data confidentiality specifically relates to the protection against unauthorized access to data contents (i.e. encryption, access control lists, and file permissions are methods often used to provide data confidentiality).

The term Privacy is referenced in several ITU-T Recommendations, including F.115, H.235, J.160, Q.1531, X.800, and X.805.

Location privacy and anonymity of the mobile node (MN), i.e. preventing the tracing of mobile user's point of attachment to some network, and identity concealment have gotten a little attention. This service has a great significance especially in wireless network which is more vulnerable to eavesdropping attack than wired network. The disclosure of the MN's location and identity allows unauthorized entities to track down its moving history, which can be a serious violation of privacy.

Fulfilling the security requirement is not the only condition for the successful deployment of mobile IP. Considering the usage of foreign network resources by a MN, it can be easily expected that a foreign network will require the MN to pay for network use. Hence, it is another issue to have a incontestable payment protocol. The payment protocol must make sure that it doesn't reveal location history of a MN as normal payments reveal the payer (HA) and payee (FA). Otherwise, the HA can know the foreign networks that the MN visited.

While providing location privacy and anonymity, the protocol must support revocable privacy rather than perfect privacy. In case of serious crime on communication or dispute on payment, the location history and identity must be revealed.

A rather different threat associated with the mobility management schemes is that of location privacy. For example, route optimization and per-host micromobility protocols might allow a correspondent to tell from a user's care-of address that they are away. It is very important that any mobility protocol is considered from a security point of view, to make sure that it does not have any obvious weaknesses that can be exploited and also that it does not open up a new security hole in another protocol.

Voice privacy can be provided in mobile systems by means of a private code or other means, and it is provided on the traffic channels only. All calls are initiated using the public code but the mobile station user may request voice privacy during or after call setup.

Information privacy is assured by using encryption mechanisms. In an effort to enhance the authentication process and to protect sensitive subscriber information, the method is provided to encrypt. If absolute location privacy is required, the mobile node can create a tunnel to its home agent. All datagrams destined for correspondent nodes will appear to emanate from the home network, and it will make it difficult for hackers to pinpoint the location of the mobile node.

Implementing IPsec for mobile IP is to protect the redirected packets sent from or to a mobile node against active/passive attack. In addition, this mechanism also helps packets sent by mobile nodes to traverse the firewall of the visiting or home network.

10.4 Technical aspects concerning security

10.4.1 Tunnelling mechanism

Employ tunnelling mechanism instead of source routing. Source routing has the disadvantage of insecurity other than complicated realization and large overhead. An attacker could use a fake care-of address as a media destination address so as to prevent the mobile node from obtaining the useful information. But mobile IPv6 could use routing header to realize source routing safely.

10.4.2 Employing IPsec

IPsec (Internet Protocol Security) was developed by the IETF. It is an integral part of IPv6 and has also been "backwards fitted" to IPv4. IPsec is made up of two components, which are IP Authentication Header (AH) and the Encapsulating Security Payload (ESP); they could carry out the functions of authentication of the IP header and the payload, integrity checking, non-repudiation and encryption of IP header and payload. These mechanisms are designed to be algorithm-

independent. This modularity permits selection of different sets of algorithms without affecting other parts of the implementation. IPSec is a kind of standardized security associations which could provide security guarantee on link layer and transport layer; service providers could implement security strategy according to the service level agreement (SLA) established with users so as to make the management of security associations easier. Moreover, there is only one security association required between the foreign network in which the mobile node is located and the home network; therefore the expansibility of security management is greatly improved.

10.4.3 Configuration of firewalls

Configuration of firewalls should be based upon specific requirements. Firewalls could protect the information in private networks against unauthorized visits from illegal nodes. Packet filtering firewall has the property of high speed, independency on applications and low-price. But the configuration of this kind of firewall is complicated. And the leakage of mainframe's IP address in the private network could not be avoided; also, suspicious behaviours could not be recorded. Application-level firewalls normally use a mainframe connected to a router as a relay mainframe; which can support more sophisticated security strategy and has a simple configuration. Private networks could be shielded by the firewall, but would suffer lower speed and the possibility of unable to provide connection with legal mobile node. Firewall with secure tunneling could be used to set up VPNs on public networks to enhance the Internet level security. Secure tunnel (built up with IPSec) which could pass through the Internet, combined with previous firewalls, could make up a strengthened VPN security system. On the original end of the secure tunnel, the transmitting node has IP data packets encrypted and encapsulated to be IPSec packets which would subsequently be transmitted through the tunnel to the receiving node, which would decrypt the received IPSec packets and resume them into the original IP packets.

10.4.4 Balance between security and complexity

Complexity of protocol realization should also be considered along with the improvement of security. It is required by the diversity of services that various security mechanisms be adopted in order to distribute various security level guarantee to different users; but too many security strategies may cause the problem that the management of the system becomes greatly complicated. To simplify the security management, it is feasible to limit mobile nodes from foreign links to visit specific resources. Security could also be strengthened by means of developing FA's intelligentized level, for instance, add access control lists (ACL) which are commonly used to filter IP packets to foreign agents (FA) and demand mobile nodes to hold certificates to communicate with FAs.

11 Service architecture

The service architecture provides a secure, extensible framework under network operator control. It supports efficient, flexible deployment of end-to-end IP applications. This architecture includes infrastructure elements for service processing, subscriber databases, media gateways and servers, and nodes for policy and charging control, as well as the terminals themselves. Real-time multimedia IP services may be provided as well as less challenging services like streaming and interactive services like messaging.

12 Interoperability and interworking

12.1 Introduction

The AIPN should provide a high level of basic system performance including low communication delay, low connection set-up time and high communication quality.

The AIPN should provide efficient usage of system resources. In particular, the scarcity of the radio resource should be respected within the AIPN by ensuring that radio resources are utilized as efficiently as possible. The AIPN should also support effective usage of power resources within mobile terminals by minimizing the impact on mobile terminal battery life (standby and active).

The AIPN should be able to accommodate a vast number of terminals and users as well as be able to support a wide variety of diverse devices. Examples of terminals that should be supported by the AIPN include terminals which main purpose is to include a sensor/RF tag, household appliances/media players with a wireless communication module, as well as traditional mobile terminals. Identification, addressing, and routing schemes within the AIPN should be provided to support this communication environment; in particular the AIPN should support naming and addressing schemes for a given user/session.

The AIPN should be able to efficiently support a variety of traffic models e.g. user-to-user, user-to-multicast and traffic models generated by ubiquitous services.

The AIPN should provide functionality as appropriate to support international roaming with other AIPNs.

The AIPN should provide appropriate mechanisms to support independent operation of services, AIPN, and/or access systems.

12.2 Legacy systems

Constant technological change often weakens the communication services value of legacy systems, which have been developed over the years through huge investments. Despite the availability of more cost-effective technology, legacy systems continue to provide a competitive advantage through supporting special communication services and containing invaluable knowledge and historical data.

In contrast to modern distributed and layered architectures; legacy systems execute communication services policies and decisions. They are characterized by rigid, predefined process flows, which make integration with user relationship management software and Internet-based business applications torturous and sometimes impossible.

Legacy modernization is crucial for organizations spending too much to maintain the communication services value of their outdated information systems. Also driving the need for change is the industry's movement toward new mobile communication platforms. The new system paradigms leverage a component-based, distributed computing model capable of automating communication services internally or with partners via mobile IP.

Several options exist for modernizing legacy systems, defined as any monolithic communication system that's too difficult and expensive to modify to meet new and constantly changing users requirements. Which technique ranging from software update and legacy wrapping to permanent, solutions such as automated migration or replacing the system with the instead of hardware board. which is also more complex.

On the other hand, a variety of options exist, including replacing the system with a packaged application or non-intrusive measures. Each of the approaches makes sense under certain circumstances. The latter methods provide quick and inexpensive access to legacy functionality, while the former can eliminate legacy applications in which the code quality is too poor to migrate.

The solution to the legacy system problem can be provided by a cooperative network, and by systems built over generic IP networking technologies. The key points are that the system should be layered on demand, encourage reuse of independent modularized functional blocks, support multiple services and service creation, ensure consistent end-to-end connectivity across different

access technologies, and cooperate in terms of network control, operations, and maintenance. The architecture shall include the endpoints of communications as part of the communications system, and should provide a secure and trusted environment in which network functions are performed; the network should self-organize dynamically.

13 Transmission efficiency

The cost and limited availability of spectrum requires that the wireless technology transporting the IP traffic be as efficient as possible.

At the physical layer the measure of spectral efficiency is bits per second/Hz of spectrum used. Optimizing this requires a careful design of the system to minimize overhead and exploit the different QoS requirements of IP applications, the variability of the wireless channel, and the mobility of multiple users in the network.

Above the physical layer, the spectral efficiency for IP applications is maintained by using protocols with low overhead such as compression protocols. In particular, IP header compression protocols for IP multimedia such as RObust Header Compression, specified in IETF RFC 3095 and Enhanced Compressed RTP, specified in IETF RFC 3545 offer a crucial role in efficiency.

14 Example of multi-media applications

The wide range of multimedia applications that run over IP networks requires that the wireless system be designed to support all the particular requirements of each application.

14.1 Voice-over-IP (VoIP) over mobile systems

Following are the advantages of carrying voice over IP versus conventional circuit-switched voice services:

- 1 The packetization of voice allows the voice codec and network to jointly exploit variations in the amount of information in the speech signal to reduce the data rate that is needed while not sacrificing voice quality.
- 2 The flexibility of the IP transport and intelligent-client architecture of IP networks allows the timely introduction of rich and new supplementary services that can be used with the voice services.

In the future, as the use of VoIP applications becomes more and more widespread, the enablement of voice services based on wireless IP will be essential.

VoIP services over mobile systems will have to overcome the following challenges:

- Good QoS support over several different network technologies.
- Real-time service insists strict delay and packet loss requirements over packet switched networks.
- Factors influencing speech quality:
 - Impairment of audio quality due to audio processing.
 - Delay, Jitter, and Loss of packets.
 - Wireless environment are characterized by fluctuations in transmission quality.
 - Physical layer leads to highly error-prone environment.

- Wireless links generally offer lower bandwidth compared to wireline links.
- Packet losses due to bit errors.
- Mean opinion score (MOS).

14.2 Web surfing

The web-surfing experience is no longer limited to static webpages described by markup languages. With the increasing use of downloaded executable applets and scripts, the web surfing multimedia experience has expanded beyond still images to dynamic animation. The richness of these downloaded applications requires downloading of larger clips for execution in the user's web browser.

14.3 Streaming audio and video

Streaming audio applications generally require medium amounts of bandwidth with a bound on the amount of total jitter experienced at the receiver. Streaming video requires higher bandwidth with similar bounds on the jitter.

Since streaming applications are generally not real time, the transport of streaming traffic typically does not require immediate delivery. However, to limit the size of the de-jitter buffer required in mobile terminals and reduce tune-in delay for broadcast streaming services, the jitter must be bound by the network.

14.4 Digital video telephony

Video telephony is one of the most challenging IP applications to support since it requires high bandwidth, low latency, and low jitter.

15 Summary

Mobile communication systems should support diverse and complex services with different traffic characteristics and QoS levels. IP applications over mobile systems is one of promising technologies to provide affordable and compatible services as used in wired communication systems.

This Report has outlined key technical and operational characteristics to support IP applications over mobile systems from various aspects, in particular: it has

- introduced the key capabilities of the All IP Network (AIPN) which is an IP-based network providing common capabilities that are independent to the relevant type of mobile service being provided and the access system being used;
- identified key characteristics to support IP applications over mobile networks and provided general characteristics for diverse QoS support;
- further described in detail, amongst those characteristics, end-to-end characteristics for AIPN with access system, including delay characteristics, round-trip delay, packet loss and the jitter of the delay;
- addressed mobility support for IP and handover since supporting mobility is becoming more important due to the wide spread of laptop computers and other mobile embedded devices;
- identified the general characteristics and discussed related technologies to ensure security, as the demands are increasing for IP applications over a mobile system that equip security functions;

- discussed the issues on interoperability and interworking as well as transmission efficiency since the efficient use of system resources becomes more important, particularly scarce radio resource;
- finally provided some examples of multi-media IP applications that run over a mobile system and discussed their characteristics.

16 Appendices

The following 9 appendices illustrate specific applications and provide valuable information in terms of validation of the characteristics listed in the main body of this Report.

Appendix 1

Description of some specific IP applications over mobile systems

1 Introduction

Section 2 contains information on features currently available in 3GPP², whereas § 3 deals with future developments currently under study.

2 Features currently available

2.1 IP multimedia subsystem (IMS)

In Release 5, IMS architecture has been developed in order to provide multimedia services based on IETF developed protocols SIP/SDP and provide service enablers in order to provide services developed within and outside of 3GPP (e.g. enable 3rd party service integration with IMS). The architecture for IMS enables delivery of services like speech, video, combinational services, presence, IMS messaging etc. IMS also provides the infrastructure for easy service development; for example, Push to Talk over Cellular (PoC) is one example of the usage of IMS in order to deliver a specific service and ability to charge for these services independent of PS domain charging.

Additional functions available via IMS are necessary enablers to support PoC, IMS messaging, and Presence services. By combining the support of messaging with other IMS service capabilities, such as Presence, new rich and enhanced messaging services for the end users can be created such as Instant Messaging, Chat, Store and Forward Messaging with rich multimedia components. The concept of presence, whereby users make themselves “visible” or not “visible” to other parties of their choice, a services such as group and private “chats” to take place. Presence is an attribute related to mobility information, and provides a different capability to be exploited by other services. The concept of presence enables other multimedia services to exploit this key enabler to support other advanced multimedia services and communications. See also TS 23.228 – “IP Multimedia Subsystem (IMS); Stage 2”.

² More detailed information is provided at <http://www.3gpp.org>.

2.2 End to end QoS procedures and service based local policy control

In Release 5, end to end QoS architecture for UMTS networks and interworking with external network procedures have been developed over the PS domain, where interworking may be achieved by:

- signalling along the flow path (e.g. RSVP, LDP);
- packet marking or labelling along the flow path (e.g. DiffServ, MPLS);
- interaction between Policy Control and/or Resource Management elements;
- Service Level Agreements enforced by the border routers between networks.

Service based local policy provides mechanism for binding media, policy control over the media (and UMTS bearer) based on the session information from IMS signalling, charging correlation and gating function using a token based approach that requires terminal support.

2.3 Supporting mobility

In 3GPP specifications 23.060 and 29.060, the procedures and protocol details for GPRS Tunnelling protocol (GTP) operation are specified. GTP provides mobility within a GPRS network including one or several access types, e.g. WCDMA RAN, GSM/Edge RAN, and GAN.

The Mobile Station (MS) set up a connection towards an external IP network through activating a PDP context. At activation of a PDP context an MS is able to use either a static IP Address or a dynamically assigned IP Address belonging to its GPRS domain in the MS's home or visited IMT-2000 WCDMA network. The MS selects the IP PoP at PDP Context activation through a configured or default Access Point Name (APN) regardless of whether the MS has a static or dynamic IP Address. The MS keeps its PDP context (IP PoP) and maintains the IP Address persistent throughout the packet data session even when handing off between radio networks connected to separate SGSNs.

Fast handover mechanisms in the context of both intra-SGSN and inter-SGSN handover are supported. At inter-SGSN handover, pre-establishment of tunnels to the new serving SGSN are set up prior to the handover and forwarding of packets from the old serving SGSN to the new serving SGSN is supported. The mechanisms are referred to as SRNS Relocation for inter-SGSN mobility within a WCDMA RAN and PS Handover for inter-SGSN mobility between a WCDMA RAN and a GSM/Edge RAN. The fast handover minimizes data loss when the MS is handing off between radio network controllers connected to separate SGSNs.

2.4 Support for IPv6

Cooperation with IETF has resulted in achieving improved support of IPv6 PDP contexts in the PS domain.

2.5 IP Transport in UTRAN

In Release 99 and Release 4, only ATM can be used at the transport layer in the various interfaces. This Work Item introduces the possibility to use IP at the transport layer in the Iub, Iur, Iu-Ps and Iu-Cs interfaces, as an alternative to ATM. However, the use of ATM at the link layer under IP is not precluded.

The introduction of IP as a transport protocol in the radio network does not imply an end to end IP network; the UE may be given an IP address by the higher layers, but it will not be part of the UTRAN IP network (which is private), and packets will be encapsulated in the corresponding User Plane protocol.

The work item has made a choice for the protocols to transport the radio and signalling bearers over IP. Different solutions are adopted: UDP is used in the user plane in the three interfaces, and SCTP with additional protocols is used for the signalling bearers. With respect to the IP version, IPv6 is mandatory and IPv4 is optional, although a dual stack is recommended.

Additionally, the Work Item resulted in decisions on QoS and interworking with ATM transport networks:

- Diffserv is the mechanism to provide different service levels, and several alternatives are allowed for the traffic flow classification. It is allowed also that the QoS differentiation can be provided either on a hop-by-hop basis or on a edge-to-edge basis.
- Interworking with Release 99/Release 4 and Release 5 ATM nodes is required, and it can be accomplished via a dual stack, a logical interworking function or a separate interworking unit.

2.6 ROHC Robust Header Compression functionality (ROHC)

Under the Release 4, the ROHC was introduced. Its benefit is an important reduction in header overhead, simply because the fields of the headers of IP packets are either constant or changing in a known pattern. Hence it is possible to send only information regarding the nature of the changing fields of those headers. This leads to a reduction in the size of header, from 40 octets into only few octets while the payload is around 30 octets for some applications (e.g. IP based voice applications) and with IP version 4, and from about 60 octets to only few octets with IP version 6. This translates directly into bandwidth efficiency. The ROHC scheme is claimed to be more suited to cellular environment and changing links than the previous compression schemes. RFC3095 “RObust Header Compression (ROHC)” is the IETF proposal for IP header compression specially designed for real time IP services over wireless links. ROHC was included in the Release 4 of UTRAN as one of the compression schemes to be provided by the PDCP (Packet Data Convergence Protocol) sublayer in the RNC. As ROHC is part of the PDCP layer, there is a compressor and decompressor pair in the RNC and a corresponding pair in the UE. During SRNS relocation the source RNC gives the role of the serving RNC (SRNC) to the target RNC, therefore compressor/decompressor have to be relocated as well. The straightforward solution in place in Release 4 was to initialise the header compression in both peers after relocation, which results in problems like high probability of lost speech frames. This could be avoided by not initialising compression but continuing it in the target SRNC from the place in which the compression ended in the source SRNC. The required changes to perform RFC3095 context relocation in the SRNS context relocation were introduced in Release 5. In order to perform the ROHC relocation, RANAP messages that carry RAB contexts during SRNS relocation are updated to carry also the ROHC/RFC3095 contexts for each RAB. The ROHC context IE to be transferred is defined in the RRC protocol specification. “RFC3095 Context Info” container to RANAP information elements “Forward SRNS Context” and “RANAP Relocation Information” were added to RANAP. Further details can be found in TR 25.844 (“Radio Access Bearer Support Enhancements”) and in TR 25.860 (“Radio Access Bearer Support Enhancements”) in TR 25.844 – “Radio Access Bearer Support Enhancements” and TR 25.860 – “Radio Access Bearer Support Enhancements” respectively.

2.7 IMS signalling flag

IMS signalling traffic is carried by an interactive PS RAB. In order to enable UTRAN to apply special handling of “signalling RABs” compared to other interactive PS RABs, a signalling “flag” was introduced as an additional level of QoS for Interactive traffic RABs.