Bringing two worlds together:
AAL2 over IP for Radio Access Networks

Winthir Brunnbauer * **
winthir@brunnbauer.org
* Infineon Technologies NA Corp.
System Architecture Group Communications
San Jose, CA 95112, USA

Gordon Cichon * **
gordon@cichon.com
** Technische Universität München
Lehrstuhl für Integrierte Schaltungen
D-80290 München

***Abstract*—With the approaching 3G wireless standard, the development of an efficient technique for transporting a mix of voice and data over limited bandwidth links is crucial. State of the art GSM systems use the connection oriented ATM protocol, while data networking is based on the best-effort IP protocol. ATM protocols have been developed for efficient universal networking with mixed voice and data traffic, and they have matured for more than a decade. This paper presents a new method for an efficient transport of voice and data over shared links by leveraging the high level part of the ATM protocol, the AAL2 adaption layer, on top a IP based transport layer. It turns out that this solution combines the advantages of both approaches, and it is superior to either one of the isolated solutions.**

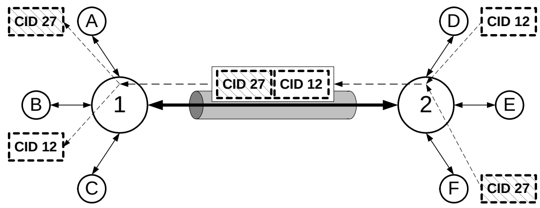

Fig. 1. AAL2 Multiplexing

I. Introduction

Cost reduction requirements force wireless service providers to use their available communications bandwidth as efficiently as possible in order to achieve maximum utilization of their scarce spectrum resources. The upcoming 3G wireless networks will be the first large-scale implementations of the ancient idea of an integrated “universal network” where latency sensitive voice traffic and high volume data traffic coexist on a shared transmission medium due to limited resources on wireless links. Efficient and mature network protocols play a key role in this effort.

ATM (Asynchronous Transfer Mode) was designed more than a decade ago in order to provide quality of service (QoS) for data and voice in an integrated backbone network. ATM Adaptation Layers (AAL) leverage the high efficiency of ATM to different applications. In the case of voice traffic, AAL Type 2 (AAL2) is optimized to transport packets with small payloads at low latencies. Fig.1 shows the AAL2 multiplexing of two independent voice streams over a shared virtual connection. ATM has a fixed length of cells with 5 octets of header and 48 octets of payload, of which AAL2 take another 3 octets per voice packet.

The most important application for data networking, the Internet, is based on a connectionless protocol, the Internet Protocol (IP) Version 4 (IPv4). The protocol defines a header of 20 octets and a variable size payload. For the the most common choices of link layers, such as PPP and Ethernet, the size of the payload is limited to 1500 octets. IP provides an unreliable best-effort service which does not guarantee packet delivery to the destination. In IP, each router delivers packets with best effort. If the resources are exhausted, packets can simply be discarded. In contrast, ATM facilitates reservations in the switches along the route for an ensured transport. The installation, maintenance and modification of these networks are quite easy and noncritical.

Today, the most prevalent protocol to transmit voice over IP (VoIP) is the Real-Time Protocol (RTP). RTP consumes an additional 12 to 76 octets per packet for header information. The rest of this paper contrasts RTP with a new approach to lay AAL2 on top of IP. This approach combines the efficient method of multiplexing in AAL2 with the easy scalablity and the low degree of maintenance in IP. Additionally, the User Datagram Protocol (UDP) may be used if more voice channels than the limited number of voice channels that AAL2 provides are required .

Regular AAL2/ATM and RTP/IP will be compared to the proposed AAL2 over IP. For conveying compressed voice packets of 17 octets, the performance characteristics of both protocols will examined using a system simulation.

II. Established Protocols

Before going into details about the specifics of AAL2 over IP, here is a brief introduction into ATM and IP, respectively.

A. ATM

The high-speed backbone network of longhaul service providers is primarily operated using the ATM protocol. It is connection oriented and the basic transmission unit consists of a fixed lendth cell with a size of 53 octets.

The connection-oriented nature of the protocol is designed to prevent congestion situations. A connection can only be established if there are sufficient resources along the route. This implies a reliable connection and a fast transfer between two endpoints. The small cell size allows low latency by multiplexing different streams. Smaller packets with high priority can easily overtake larger packets with low priority. The traffic management of ATM comprises fifteen years of experience.

4	8	16	3	1	8	384 (48 bytes)
GFC	VPI	VCI	Type	CLP	HEC	Payload

Fig. 2. ATM header

For some applications, the fixed 5+48 byte format of ATM (Fig.2) often doesn’t quite fit, especially when 49 or few more octets are to be transferred. So five AAL (ATM Adaptation Layers)[1] were introduced (see Table I) in which AAL Type 5,

[1] AAL3/4 has been replaced by AAL5.

or AAL5 is the most notably for data transfer. It supports connection-oriented, reliable variable bit rate data services. AAL Type 2, or AAL2, provides connection-oriented services for variable bit-rate applications like some audio and video schemes. Enhanced Full Rate Speech Codecs produce approximately 17 bytes every 20 milliseconds for compressed voice applications [1]. Other codecs produce between 4.5 and 30 octets per packet. These packets are actually too small to fill an entire ATM cell. AAL2 is designated to multiplex these small packets with low overhead to avoid too much padding of the remaining unused bits of the transmitted ATM cell.

	real-time	bandwidth	connectionless
AAL1	yes	fixed	no
AAL2	yes	variable	no
AAL3/4	no	variable	yes/no
AAL5	no	variable	yes/no

TABLE I
ATM ADAPTATION LAYERS

B. *Internet Protocol*

In the ages of the cold war, a data network, the ARPA net, was developed to maintain communication after a nuclear strike in which large parts of the infrastructure would have potentially been destroyed. For this reason, the network was designed in a decentralized and redundant fashion. Realtime communication was not an issue at this time. The protocol does not even depend on reliable packet transmission. These features make the infrastructure very scalable. The Internet emerged as a civil application of this technology.

The IP protocol has serious setbacks for efficient voice communication over the long haul: in times of heavy congestion, packets can be discarded by intermediate routers without further feedback. Additionally, since the internet protocol does not make any assumptions about the intermediate network topology, there is no guarantee of packet transmission latency.

Traffic management protocols like RSVP [3], DiffServ [4] and MPLS [5] expand the unreliable service to support a reliable network environment. For instance, RSVP (ReSerVation Protocol) reserves resources along the packet route. If a new packet flow tries to allocate unavailable resources, the allocation will be denied. This assures that the packet won't be discarded along the route. DiffServ (Differentiated Services) defines queues with different priorities, e.g., a small high priority control packet is released earlier from the queue than a huge low priority data transfer packet, even if the data packet arrived earlier. MPLS (Multiprotocol Label Switching) is an effort to reduce latency by encapsulating the whole packet and adding a label. This label can be read by a MPLS-Router which switches the packet to a egress port without looking into the packet.

A big advantage of IPv4 scalability is the decentralized structure. This makes the installation, the setup and the enhancement of new and established structures noncritical. Also, the heterogeneity enables IP4 to be used on different network platforms.

Since the address space of the internet protocol version 4 is limited to 32 bits, the next generation wireless network will be based on IP version 6, which provides a 128 address bits. In the consumer electronics network, each consumer is expected to own hundreds of appliances connected to the internet. Assigning individual addresses to each of them makes an expansion of the address base inevitable. However, the size of the transmission header is extended from 20 to 40 octets. Without loss of generality, we assume IPv4 in the following reasoning. For IPv6 our arguments become even stronger.

III. RTP OVER UDP/IP

RTP was developed to convey compressed voice and video over a network and synchronize the serval packets at the receiver. The header of RTP contains among other things a sequence number, a time stamp, a synchronization source and from zero to sixteen contribution sources [6]. The sequence number is used to detect packet loss and restore packet sequence. The timestamp reflects the sampling instant of the first octect in the RTP data packet. The synchronization source (SSRC) reflects the last instance which generates or mixes the multimedia data. The contributions source (CSRC) lists up to fifteen sources which contributes in this mixed multimedia stream.

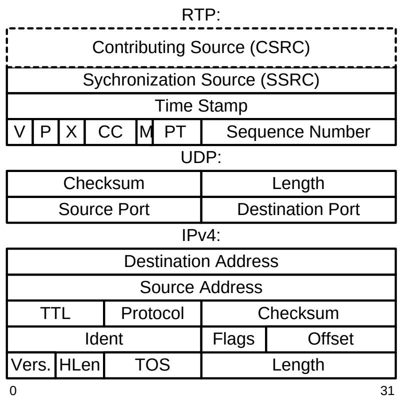

Fig. 3. Protocol Stack of RTP/UDP/IPv4

RTP is usually used in combination with UDP and IPv4. In Fig.3 the protocol stack of RTP/UDP/IP is depicted. UDP is an unreliable transport layer which enhances IP by ports, which helps to assign arriving packets to specific applications. If a packet arrives at the receiver, the port number helps to deliver the packet to the proper application. The better known counterpart for reliable transmission is TCP. Here an unreliable transmission protocol is more useful, because there is no need for retransmission in a transient environment. If a voice packet is delayed for longer than the buffer time space of the encoder, the packet will be discarded anyway, even if it is received error-free.

IV. WHY AAL2?

RTP/UDP/IP consumes between 40 and 104 octets for the header subject to the number of CSRCs, in contrast to the size of the compressed voice packet which is only about 17 octets. So the ratio between the header and the payload is unbalanced.

Instead of using one header for each voice packet, it is much more effective to pile several voice packets onto one header. AAL2 provides such functionality. It packs up several voice packets with a small AAL2 header into one transport packet (s. Fig.4). The advantage of this method is a higher link utilization. Utilization is defined here as the ratio of the size of the payload compared to the size of the whole packet.

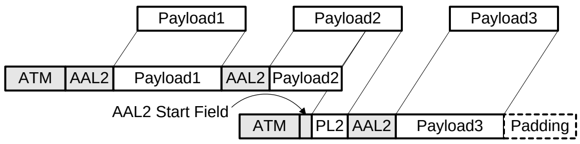

Fig. 4. Principle of ATM/AAL2 Multiplexing

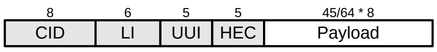

Fig. 5. AAL2 header fields

The AAL2 header consists of a Channel Identifier field, a Length Indicator field, a User-to-User-Indication field and a Header Error Correction field (s. Fig.5).

In a scenario where a lot of small voice packets are ready to be transmitted over the link, the notion is just to replace the ATM header with the IP header. With the IPv4 header, the limitation of 48 octets payload is removed. Now the header has 20 octets instead of 5 octets, but it must be transmitted much less frequently. Also, the pointer for locating the first AAL2 packet in the succeeding IP packet needs less time in IP than in ATM (s. Fig.6). It must be sent in the first position of each payload to identify the begining of the next AAL2 packet.

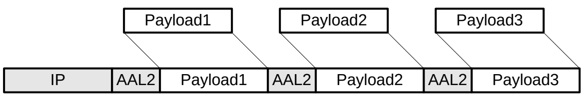

Fig. 6. Principle of IP/AAL2 Multiplexing

Fig.7 depicts the utilization of AAL2 over ATM and AAL2 over IP proportional to the number of voice packets to convey. So except for the cases with 2 and 6 voice payloads, the utilization of IP is better than ATM.

For instance, a compressed voice payload in a plain ATM packet has a utilizaion of 17 octets (AAL2-Payload) / 53 octets (ATM) = 32.1%. With AAL2, the header will be enlarged by three bytes with every AAL2 voice payload, but after that these packets can be merged together in a single ATM cell. Due to the nature of ATM, all cells have a fixed size of 53 octets. With one 17-octet-packet, the ATM cell comprises 5 octets ATM header, 3 octets AAL2 header, 17 octets voice payload and 28 padding octets, which simply fill up the remainder of the ATM cell. With two 17-octet-packets, the ATM header occupies 5 octets in the ATM cell, the AAL2 header 3 octets twice, the voice payload 17 octets twice and the padding 8 octets. The utilization rises linear. But when a third voice packet has to be conveyed, the utilization drops because a second ATM cell must be used to transport the second half of the third voice payload with 17 octets. This is because it does not fit in place of the 12 padding octets.

In Fig.8 voice packets with uniformly distributed sizes between 12 and 22 octets are used. It is apparent that there is not much difference except for the lobes of ATM line which are flattened. AAL2 over IP has a higher utilization of around 10 percentage points than AAL2 over ATM. This is because the large header of IP is only used once.

The usage of IP avoids the restriction of the size of the payload to 48 octets. Also, the combination of AAL2 and IP leverages the traffic management knowhow of ATM and enables the widespread and inexpensive network layer of IP. In scenarios where a bigger part of the traffic mix is voice and ATM was used before (such as in radio access network (RAN)), existing hardware solutions of AAL2 can be reused. This makes it possible to archive a better interoperability with the legacy of RANs. In a self-contained RAN-IP network, e.g., DiffServ [4] and RSVP [3] provide the opportunity to avoid traffic congestion and assure the packet flow.

The drawback of the usage of AAL2 is lack of information such as timestamp and identifier of the origin of the voice stream (SSRC and CSRC). But current RAN environments do not provide mixing of voice streams anyway. All the packets are sent point to point. Therefore, the timestamp which is used for mixing several audio or video streams together can also be waived. [8]

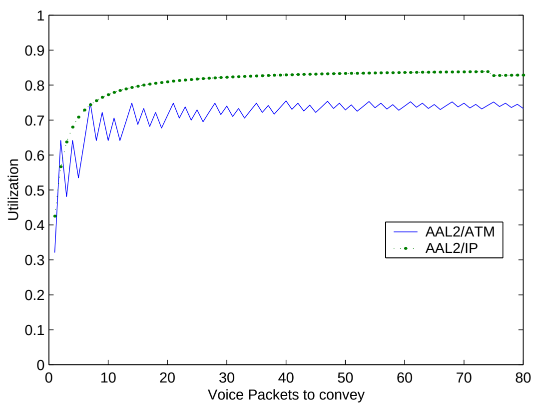

Fig. 7. Comparison of AAL2/ATM and AAL2/IP with a constant voice packet length of 17 octets

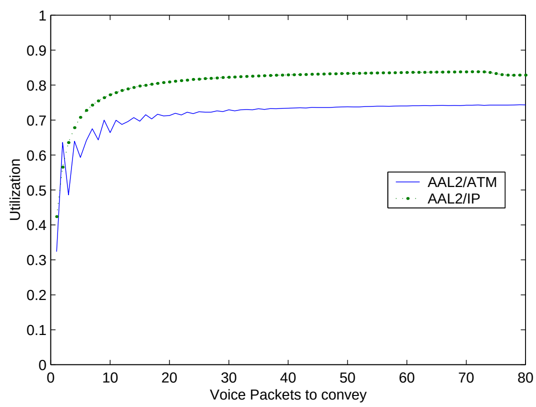

Fig. 8. Comparison of AAL2/ATM and AAL2/IP with a uniform distributed voice packet length between 12 and 22 octets

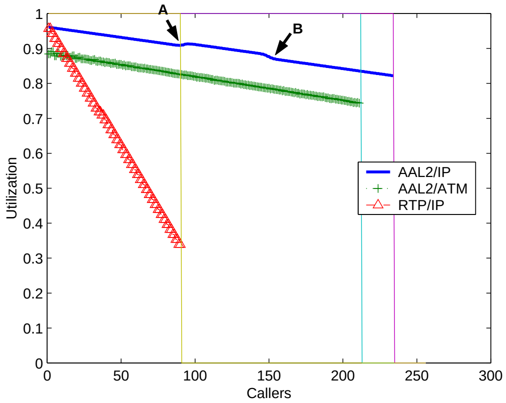

Fig. 9. Utilization

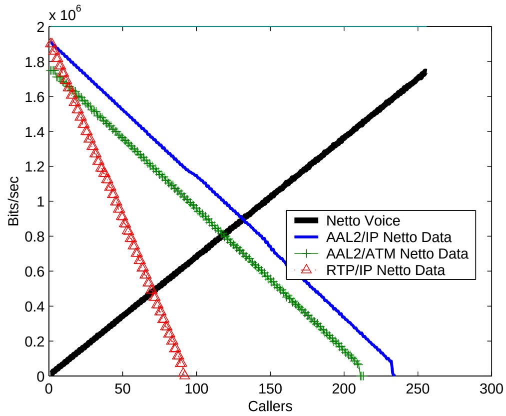

Fig. 10. Used Payload bandwidth

V. UDP or not UDP?

The context identifier (CID) of AAL2 consists of only 8 bits which allows 248 different contexts and 8 reserved codepoints for control issues. Each Context contains information about the caller, the callee and the path upon which the packet should proceed. The result is up to 248 different calls are possible for each connection. In order to increase the maximum number of channels, an additional transport layer has the ability to multiplex the CID by using different port numbers.

VI. Simulation and Results

The following simulation scenario describes a link within a Radio Access Network, which might connect a radio network controller to a Node-B (basestation). This link is in our example an E1 line with a data rate of 2 Mbit/s, which is a common set-up. Each caller produces, at his source, between 12 and 22 octets, uniformally distributed, every 20ms. Since these packets have high latency constraints to meet the required quality of service, they are bundled up in special high priority packets and sent in the first place without delay. The rest of the bandwidth is filled up with data packets. These data packets are part of a low priority data transfer and only sent when there is sufficient bandwidth available.

The voice packets are encapsulated by AAL2 with 3 octets of header. These packets are multiplexed into one huge packet which is then processed by ATM or IP. ATM segments the payload to one 48-octet followed by several 47-octet chunks. With the 47-octet chunks, one octet is appended as a start field. This field is used as a pointer which shows the position of the succeeding AAL2 packet (s. Fig.4). IP takes the full voice payload and attaches it to the UDP/IP header. In the case that the payload is larger than 1500 octects, the payload is split up in pieces of 1500 octets and the same method with the start field is used.

The remaining bandwidth is filled up with 1500 octet data pieces and corresponding headers. ATM uses the AAL5 to ensure reliable transmission over the link. AAL5 adds 8 octets and enough padding octets to fill up to a multiple of 48 octets. IP just takes the partial data packets and transmits them over the link as they are.

The Fig.9 depicts the utilization in relation to the number of callers sending voice packets. The two common methods, AAL2 over ATM and RTP over UDP/IP, and our proposal, AAL2 over UDP/IP, are shown. ATM with AAL2 has a utilization of 89.2% with no voice packets and only data packets. The utilization drops down to 75.1% with only voice packets and no data packets. Over this E1 link up to 210 callers can be supported. With RTP/UDP/IP the utilization with no callers and only data packets is with 96.5% higher than AAL2/ATM. However, the decrease in utilization is much higher with the raising number of voice packets. It can only support up to 90 users with a utilization of only 34.2%.

AAL2 over UDP/IP, delivers the best solution. With only data, it delivers a high utilization of 96.5% as RTP/UDP/IP. It also provides services for up to 230 callers and the utilization remains as high as 83.1%.

Arrow 'A' and 'B' in in Fig.9 shows a phenomenon which is caused by the granularity of the packet sizes. At this point, a more optimal distribution of the payload is archived. Instead of distributing the information over one voice and two data IPv4 packets, only one voice and one data packets are needed. This raises the utilization about 2%. Beyond Arrow 'B', the number of voice packets exceeds the maximum number of possible octets within a IPv4 payload. A second IPv4 packet is needed which lowers the utilization back to values with only few callers using the link.

Fig.10 depicts the used bandwidths of the different kinds of payload in relation to the number of callers. The rising curve shows the usable bandwidth of the voice packets without any headers with the increasing number of users. This figure is the same for all three methods. But the three methods have different bandwidth consumptions for the encapsulated voice packets. As additional bandwidth is needed for voice, less bandwidth is remaining for data. AAL2/ATM and AAL2/IP uses AAL2, which needs only 3 octets for each voice packet. In addition, ATM needs 5 octest of header fields every 48 octets of payload. RTP however asks for a full RTP/UDP/IP header each voice packet for 36 octets.

AAL2/IP provides the best solution since it combines the advantages of only three octets of header with AAL2 and a low header repetition with IP. While the RTP/UDP/IP bandwidth is consumed with only 90 callers, AAL2/IP has remaining bandwidth for 1.2 Mbits/s or another 140 calls available. When comparing AAL2/ATM to AAL2/IP, IP has an additional 200 kbit/s for data and bandwidth for 20 more callers.

While RTP/UDP/IP uses all bandwidth of the link for 90 callers, AAL2/IP has more than half of the available bandwidth left to support voice and data packets.

VII. Conclusion

AAL2 instead of RTP on top of the IP protocol gives a significant advantage in terms of both performance and bandwidth utilization. It is possible to accomodate more than twice as many telephone callers with AAL2 as with RTP, which gives the further an significant edge over the latter. The benefit of the shown protocol combination is further reinforced by the fact that there are efficient hardware implementations for both IP and AAL2 available. There is an efficient hardware implemenation of AAL2, which is used by the upcoming IWORX radio network processor from Infineon([7]). As IP packet processing is becoming a commodity, there are a host of providers of more or less efficient networking processors to facilitate this type of processing ([10][9][2]).

The know-how and expertise of the ATM and IP worlds can be successfully combined in order to create an efficient universal network in the wireless environment. Since the requirements and expectations of well established stationary voice and data services vastly differ, these networks will tend to coexist for still quite some time before they may or may not converge into a universal network. However, in the mobile world, there is only a single transmission medium. Building a universal network with QoS is a straightforward development. This development is further driven by aggressive investments and reinforced by technological agility of the European players who back the UMTS platform.

Appendix

Thanks to Charles Bry, Gunnar Hagen and Ingolf Ruge

References

[1] 3rd Generation Partnership Project 3GPP TS 26.071 V3.0.1: Technical Speicification Group Services and System Aspects; Mandatory Speech Codec speech processing functions AMR Speech Codec; General Description; http://www.3gpp.org. September 1998.

[2] Marco Heddes: IBM Power Network Processor Architecture. In *HotChips*, number 12, 2000.

[3] Braden et al. RFC 2205, RSVP: Resource ReSerVation Protocol, IETF Network Working Group; http://www.ietf.org. September 1997.

[4] Nichols et al. RFC 2474, DiffServ: Definition of the Differentiated Service Field (DS Field) in the IPv4 and IPv6 Headers, IETF Network Working Group; http://www.ietf.org. December 1998.

[5] Rosen et al. RFC 3032: MPLS Label Stack Encoding, IETF Network Working Group; http://www.ietf.org. January 2001.

[6] Schulzrinne et al. RFC 1889, RTP: A Transport Protocol for Real-Time Applications, IETF Network Working Group; http://www.ietf.org. January 1996.

[7] Infineon. ATM Interworking Controller, IWORX - PXB 4225, Datasheet, http://www.infineon.com/. January 2001.

[8] Terry G. Lyons Kotikalapudi Sriram, Yung-Terng Wang: Anomalies Due to Delay, Loss in AAL2 Packet Voice Systems: Performance Models, and Methods of Mitigation. In *IEEE Journal on Selected Areas in COmmunication, Vol.17, No.1*, January 1999.

[9] Dave Husak: The C-5 Digital Communications Processor. In *HotChips*, number 12, 2000.

[10] Steve Sheafor, Vitesse Network Processors: Optimizing Architecture for Bandwidth Cindy Lindsay, and Flexibility. In *HotChips*, number 12, 2000.