Abstract - IP telephony will be more and more important during the next years to come. The terminals will be portable cordless devices, with images. The products of today do not fulfill all these requirements. Non-image IP phones of early 2001 are Ethernet-based devices, and the processing is based on general-purpose CPUs and DSPs. It is concluded that to achieve full functionality in a low-power and still flexible solution, specialized hardware architectures must be used.

INTRODUCTION

In a couple of years telephony will be transported on IP-networks. To be able to transmit image as well as voice will be natural and the terminal will be cordless. Those statements are well accepted in industry today and companies start to release IP-based products for the huge telephony market.

All these demands on functionality and easy to use interfaces makes the IP telephony terminal a complex device. This becomes a problem when considering the low-power demand on cordless, battery-driven appliances.

Clearly new and specialized hardware architectures are necessary to satisfy the specifications. Still flexibility must not be omitted since protocol standards keep changing and the products must be easily upgradable with new functionality on a pure software basis.

The aim of this paper is to investigate which hardware architectures are used in products today and to suggest improvements to those.

FUNCTIONALITY OF AN IP TELEPHONY TERMINAL

Two standards exist for IP Telephony, H.323 and SIP. H.323 is a ITU standard and presently version 2 is being used. SIP is a standard from IETF. Although in details very different, the two standards try to fulfill the same purpose, that of how to make phone calls over IP-based networks. H.323 however is a more extensive standard than SIP and specifies for example a large number of call cases. From now on only H.323 is considered in this paper. Many of the complicated issues, such as bridging to ISDN and POTS, we do not have to consider when looking at the terminal. This is taken care of by gateways and gatekeepers [1].

The IP telephony terminal has to be able to perform certain tasks which are listed in table 1.

<table>
<thead>
<tr>
<th>Task</th>
<th>MIPS consumption</th>
<th>Complexity level</th>
<th>Program size</th>
<th>Data buffer size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initiate a call</td>
<td>VL</td>
<td>M</td>
<td>L</td>
<td>L</td>
</tr>
<tr>
<td>Register at a gateway</td>
<td>VL</td>
<td>M</td>
<td>L</td>
<td>L</td>
</tr>
<tr>
<td>Communicate with a gatekeeper</td>
<td>VL</td>
<td>M</td>
<td>L</td>
<td>L</td>
</tr>
<tr>
<td>Answer an incoming call</td>
<td>VL</td>
<td>L</td>
<td>L</td>
<td>L</td>
</tr>
<tr>
<td>Demultiplex packet flow</td>
<td>L</td>
<td>H</td>
<td>H</td>
<td>H</td>
</tr>
<tr>
<td>Handle encryption and decryption</td>
<td>M</td>
<td>M</td>
<td>H</td>
<td>M</td>
</tr>
<tr>
<td>Handle voice codecs</td>
<td>H</td>
<td>VH</td>
<td>H</td>
<td>H</td>
</tr>
<tr>
<td>Perform echo cancellation</td>
<td>H</td>
<td>M</td>
<td>M</td>
<td>H</td>
</tr>
<tr>
<td>Handle image codecs</td>
<td>VH</td>
<td>VH</td>
<td>H</td>
<td>VH</td>
</tr>
</tbody>
</table>

Table 1 explanations: VL=Very Low, L=Low, M=Medium, H=High, VH=Very high.

PROCESSING DEMANDS

The first five tasks listed in table 1 do not require very much processing power counted in multiplications and additions. Non the less they have to be considered specifically since they require heavy protocol processing. This protocol processing is not performed well by an ordinary CPU or DSP and therefore special architectures must be considered to be able to keep flexibility and low-power consumption. The protocol stack that
is used can be seen in figure 1. It is hard to say exactly how
MIPS consuming the protocol processing is, but it takes a not
negligible share of a 100 MIPS 32-bit RISC according to [6].
The encryption and decryption handling are very tricky issues
for an IP telephony terminal. First of all signalling channels
must be encrypted so no one can interfere with call connec-
tions and payment. Then the media channels must be secured
as well. Different algorithms are used for these purposes. The
computation intensive part is the encryption and decryption of
media channels. Each terminal can signal encryption capabili-
ties and if no common level of encryption can be found a call
can be denied due to security denial. An IP telephony terminal
definitely should have enough processing power to handle
encryption. Encryption is used at many layers, 802.11 has
WEP (Wired Equivalent Privacy) specified. On the IP level,
IPSec supports encryption and on the application layer the
encryption standard is called H.235. Encryption, and thus
decryption, can be mixed into the codec, or used on the coded
data.

There are several voice codecs available. For H.323 narrow-
bond communications G.723.1 is used. Voice codecs can be
executed efficiently on normal DSPs. The G.723.1 codec
requires 15-20 MIPS on a fixed-point DSP and compresses the
voice signal to 5.3 kbit/s, 6.4 kbit/s, or 8 kbit/s.
The sound transmission from loudspeaker to microphone on a
mobile terminal, leading to an echo at the other party, is nor-
mally too big to be neglected. Therefore room echo cancelling
must be performed on the voice before it is coded and trans-
mitted. Room echo cancellation requires 40 MIPS on a fixed-
point DSP.

Among image codecs there also exist plenty to chose from,
H.263 and MPEG-2 are two of the most popular. MPEG is
more flexible, but H.263 is preferable in low bit rate communi-
cations. H.263 is also very good for video with very little
movement, for example video conferencing where normally
only lips and eyes are moving on a steady background. Com-
pressed bit rates vary from 20 kbit/s to several Mbits/s,
depending on frame rate, picture resolution, and desired qual-
ity. All modern video codecs include motion estimation, DCT,
and some other coding technique. This implies that special
architectures must be used for the video codec.

The last thing to consider is the wireless communication link.
The now mostly used standard for new installations of wireless
LANs is 802.11b. Other standards are available, such as Hiper-
lan/2, but here only 802.11b is considered. The layer 2 part of
the 802.11b standard is protocol processing and can be han-
dled by the same hardware as the rest of the protocol process-
ing. The physical layer must be handled by special circuitry.
The 802.11b standard supplies 11Mbit/s shared bandwidth. This is enough for several IP phones and some other equip-
ment.

EXISTING PRODUCTS

IP phones, that can be bought in the beginning of year 2001 are
not portable solutions. Instead most of them have two Ethernet
connections, with some switching capability. It is intended that
they are connected to the switch and to the desktop computer,
this saves new cables, since the one physical connection at the
desk can be used for the IP phone as well as for the desktop
computer.

Apart from the lack of portability existing IP phones are not
supporting images.

In the rest of this section some existing IP phones and IP
phone chips are presented.

The APLIO IP Phone

The APLIO IP Phone is based on a single chip solution. It does
not support images or wireless communication. Voice codecs,
telephony and VoIP is taken care of by one ARM7 core and
two OakDspCores [2].

Broadcom BM1100

Broadcom BM1100 has similar functionality as the APLIO
chip, but also includes Ethernet communication on the same
chip. The chip consists of one RISC core and one DSP [3].

Cisco VIP 30

The Cisco VIP 30 has very similar functionality as the Broad-
com product. Although Cisco does not specify the internal
architecture, it does not seem to be optimized, since it
demands 100-230 VAC and weights 1.5 kg [4].

ADTech IP Phone

Also the ADTech IP Phone has similar technical specification
as the Broadcom product, but it is based on a single DSP from
Texas Instruments [5].

PREFERRED ARCHITECTURE

As can be seen in the previous section, several IP phones are
available and all are based on traditional architectures, where
specialized computing demands are solved by using general-
purpose processors and DSPs. Figure 2, shows a better archi-
tecture. This architecture includes wireless LAN connection
and supports images.
The processing is split into two parts. One part is performed at link layer speed, and the other at application speed. Here reception is explained, but sending has similar behavior. First, 802.11b physical and MAC must be performed at link layer speed. Also IP/TCP/UDP is performed at link layer speed in the domain-specific protocol processor. Thereby extra memory buffering is avoided, and so power is saved. After demultiplexing the RTP packet flow, and in some cases decryption, the coded data is stored in a buffer for decoding at application speed.

The micro controller coordinates operation and controls the accelerators. It is heavily interacting with the protocol processor on connection establishment and tear-down. The micro controller also handles the user interface.

The big advantage with using 802.11b connection is that it supports mobility within the office and that the network infrastructure will soon be installed in many offices anyway. This leads to cost sharing and easier maintenance. The voice quality has been proved to be good on such a network [7].

![Architecture overview of an IP image phone for 802.11b WLAN connection](image)

**FIGURE 2.** Architecture overview of an IP image phone for 802.11b WLAN connection

**CONCLUSION**

The IP image phone terminal is a complex device, with many different computing demands. To achieve a low-power and still flexible solution, specialized hardware architectures must be used.

**ACKNOWLEDGMENTS**

This study was supported by the INTELECT program of SSF (Swedish Foundation for Strategic Research) and supervised by associate professor Dake Liu.

**REFERENCES**


