SCTP based Handover Mechanism for VoIP over IEEE 802.11b Wireless LAN with Heterogeneous Transmission Rates

John Fitzpatrick, Seán Murphy and John Murphy

Performance Engineering Lab,
School of Computer Science and Informatics,
University College Dublin,
Ireland
E-mail: john.fitzpatrick@ucd.ie

Abstract— In this paper a transport layer handover mechanism for Voice over Internet Protocol (VoIP) in 802.11b using the Stream Control Transmission Protocol (SCTP) is proposed. The multi-homing feature of SCTP is used to allow connections to several 802.11b Access Points (AP's). Probing packets that model VoIP encoded data at various transmission rates are used to obtain quality metrics from each of the available networks. Handover decisions are made based on the Mean Opinion Score (MOS) calculated from the ITU-T E-Model for voice quality assessment using the obtained measurements.

The handover mechanism is shown to operate in 802.11b networks with heterogeneous transmission rates using multiple VoIP codecs. The results show a high correlation between the MOS predicted by the proposed mechanism and the MOS experienced by a VoIP call present in the network. Results verify the accurate operation of the scheme using multiple VoIP codecs at various transmission rates. A simulation showing an SCTP endpoint handover between heterogeneous transmission rate AP's is presented.

I. INTRODUCTION

The common view among researchers of Fourth Generation (4G) mobile communication is that it will be a heterogeneous network environment. It will offer ubiquitous access to services across multiple wireless access technologies such as WLAN, 3G and Wi-Max. There are still many challenges for services such as VoIP to operate across multiple wireless networks. One of these challenges is seamless handover.

There is still a lot of debate within the research community about the optimum layer to perform handovers. In [1] each of the layers was analyzed as a candidate for implementing a mobility solution. In this work it was concluded that implementing mobility solutions at the transport layer was most suitable.

One transport layer protocol that seems to be a good candidate for handover mechanisms is Stream Control Transmission Protocol (SCTP). SCTP is the third transport layer protocol to be ratified by the Internet Engineering Task Force (IETF) [2]. SCTP provides reliable, connection oriented communication to endpoints that may have multiple IP addresses and interfaces.

SCTP offers an effective method for attaching to multiple IP networks with a feature known as multi-homing. Our work uses the multi-homing feature to allow seamless handovers across wireless IP networks. Similar work was proposed by researchers at the University of Oklahoma, with a mechanism called SIGMA, a seamless handover scheme for data networks [3]. Their work used the multi-homing feature of SCTP to enable seamless handovers across base stations in different IP domains.

In [4] a delay centric handover mechanism utilizing the multi-homing feature of SCTP was proposed. In [5] a transport layer handoff solution entitled cellular SCTP (cSCTP) was proposed. This solution utilized the dynamic address reconfiguration extension [6] for SCTP and used the mobility management functionality of SIP to enable seamless handover. In [7], [8] and [9] studies analyzing the VoIP capacity of 802.11b networks were carried out. The results obtained in our work agree with those previous studies.

Our work differs from those discussed by focusing on making handover decisions based on the VoIP Quality of Service (QoS) obtained from the available networks. The proposed handover scheme can operate in WLAN networks where each node can communicate with the Access Point (AP) at different transmission rates. In this paper we call this network a heterogeneous transmission rate network. The proposed scheme probes each available network using simulated VoIP data operating at different transmission data rates using different VoIP codecs. The probing mechanism allows network performance metrics to be obtained for each network and a handover decision to be made. The network performance metrics are mapped to a well known voice quality metric using the ITU-T recommended E-Model [13].

It is envisaged that the proposed mechanism be used in heterogeneous network environments. However, initial design and validation of the work as presented here is done in a homogeneous network context using 802.11b WLAN APs.

This paper is structured as follows. Section II gives an overview of the features of SCTP and the E-Model, pertinent to the proposed handover mechanism. In Section III, the design and implementation of the handover scheme is discussed. Section IV describes the simulation setup and is followed by a discussion of the results obtained in section V. This paper is then concluded in section VI.
II. SCTP & The E-Model

In this section, the details of SCTP and the E-Model relevant to the proposed handover mechanism are described.

A. SCTP

Stream Control Transmission Protocol (SCTP) is a reliable transport layer protocol, which inherited many of the core features of TCP such as congestion control and retransmission. SCTP also includes enhancements over TCP.

A central concept in SCTP is the definition of an association. An association in SCTP is analogous to a connection in TCP.

An important feature of SCTP is Multi-Streaming. Multi-Streaming allows independent streams of data to be transmitted across a single association with no reliance on the delivery order of packets in other streams. Multi-streaming is used in SCTP to solve the TCP problem of Head of Line (HOL) blocking that arises from TCP’s strict byte ordered delivery.

A key difference between TCP and SCTP is that in SCTP an association can span multiple IP addresses. Each of the IP addresses can be bound to separate IP interfaces. This feature is known as multi-homing. As this feature is critical to the operation of the proposed mechanism, it is discussed in greater detail below.

1) Multi-Homing

An example of SCTP multi-homing is shown in Figure 1. In this example each endpoint has two physical links and two network layer interfaces bound to a single association.

A heartbeat is sent periodically every Retransmission Timeout (RTO), RTO being a user/application definable time. When a HEARTBEAT packet is received by an endpoint, the packet is processed and a HEARTBEAT ACK packet is sent back. Each HEARTBEAT packet contains a timestamp of when it was sent. The time difference between transmission and receipt of the ACK packet can be used to estimate the Round Trip Time (RTT) between the endpoints.

B. The ITU-T G.107 E-Model

1) E-Model Overview

The E-model is a computational model for estimating the subjective quality of a VoIP call. It is standardized by the ITU-T (International Telecommunications Union Technical standards) as G.107 [13]. The E-Model combines loss and delay impairments based on the concept that perceived quality impairments are additive. The primary use of the E-model is in the design of codecs and transmission networks. The output of the E-model algorithm is a scalar rating of call quality called the R value.

Our approach is to use the E-model in real time to obtain voice quality values for each network and make handover decisions based upon this metric.

The R value is calculated from:

\[ R = R_o - I_s - I_d - I_e + A \]

where:
- \( R_o \) : Basic signal-to-noise ratio
- \( I_s \) : Impairments simultaneous to voice encoding
- \( I_d \) : Impairments due to network transmission
- \( I_e \) : Effects of equipment (e.g. low bit rate & loss)
- \( A \) : Advantage factor

\( R_o \) and \( I_s \) are parameters associated with the voice signal and are not affected by transmission over the network. \( I_d \) and \( I_e \) are the only variable parameters affected by network transmission. The advantage factor \( A \) is used to offset the reduced quality users may be willing to accept in certain circumstances such as in a mobile environment. In this work the advantage factor is set to 0, so that the results are directly comparable to wired VoIP and PSTN (Public Switched Telephone Network) systems.

The E-model output R can be converted into the more commonly known metric MOS (mean opinion score). The MOS system is more commonly used to measure how a user rates call quality. Table 1 shows how the E-model R rating maps to the mean opinion score.

<table>
<thead>
<tr>
<th>R value (lower limit)</th>
<th>MOS (lower limit)</th>
<th>User Perception</th>
</tr>
</thead>
<tbody>
<tr>
<td>90</td>
<td>4.34</td>
<td>Very Satisfied</td>
</tr>
<tr>
<td>80</td>
<td>4.03</td>
<td>Satisfied</td>
</tr>
<tr>
<td>70</td>
<td>3.6</td>
<td>Some users dissatisfied</td>
</tr>
<tr>
<td>60</td>
<td>3.1</td>
<td>Many users dissatisfied</td>
</tr>
<tr>
<td>50</td>
<td>2.58</td>
<td>Nearly all users dissatisfied</td>
</tr>
</tbody>
</table>

Table 1 - G.107 VoIP call ratings
2) Coding and Loss Impairments

In VoIP over WLAN packet loss is an important consideration. Packet loss is accounted for within the E-model by the equipment impairment factor \( I_e \) which is loss and codec dependent. The distortion introduced by a certain amount of loss varies for different codecs.

Each codec is assigned an equipment impairment factor value depending on how much it degrades the perceptual voice quality. In general the lower the codec bit rate the greater the equipment impairment factor. Low bit rate codecs suffer greater in the presence of loss due to inter packet dependencies introduced by their encoding mechanisms. Encoding laws such as G.729 obtain lower bit rates using higher compression at the cost of their maximum attainable quality. For example in non-error conditions the \( I_e \) value for G.711 is 0 and 11 for G.729.

In the ITU-T recommendation G.113 [10], the values for \( I_e \) under varying loss conditions are given for each codec. In this paper the codecs G.711 with PLC (Packet Loss Concealment) and G.729 are considered.

III. HANDOVER DESIGN

This section details the design and implementation of the proposed handover mechanism.

A. Heartbeat modifications

The original heartbeat mechanism of SCTP was modified to send multiple heartbeat packets with particular time spacing between consecutive packets. Packet size, the number of packets and the time spacing between them can be modified such that the probing mechanism can model any VoIP encoding scheme.

To correctly estimate the MOS of a call on a WLAN, the heartbeat mechanism needs to mimic the behaviour of the voice over IP codec being used. For example a VoIP call using the G.711 codec can be approximated as transmitting 80 byte packets every 10ms. Therefore each heartbeat packet is 80 bytes long with a time spacing of 10ms between each consecutive packet, this means that the SCTP heartbeat probe very closely simulates G.711 encoded VoIP data. Through similar modifications to the heartbeat probe many VoIP codecs can be modelled. Note that the heartbeat probe packets consist of null information, as SCTP heartbeat probes are not delivered to the application layer. Further development of this work will replace the null information with real VoIP data.

B. Implementing the E-model

While the E-model was primarily designed as a network planning tool, some work has been done on applying the E-Model to real time environments. In [11] specific values were chosen for parameters within the E-Model. This simplified variant can then be used in a real time context. Using their work the E-model calculation can be reduced to:

\[
R = 93.34 - Id - Ie
\]

\( Id \) takes into account network delay and jitter parameters. The equipment impairment factor \( I_e \) is loss and codec dependent.

C. Network Parameter Calculations

Network measurements are used to calculate the MOS that can be obtained by using each of the available networks. The measurements are calculated using the RTT values obtained from heartbeat packets. A train of heartbeat packets is sent to each endpoint in the association every \( T \) seconds. The number of heartbeat packets in each probe train is important for the accuracy of the Handover mechanism. Multiple experiments comparing the SCTP mechanism’s calculation of MOS and the MOS experienced by a VoIP call present in the network were carried out. From these it was found that on average 25 heartbeat packets is needed to accurately calculate the MOS. The accuracy of this mechanism is shown in the results section.

\( Id \) is dependent on delay and jitter. These are determined by the RTT values obtained from the heartbeat probe trains. As \( Id \) is dependent on the one way delay this must be determined from the RTT values. Using the assumption that the network link is symmetric, the one way delay is calculated as half the RTT. Jitter is calculated using the E-Model recommended RTP jitter algorithm from RFC1889.

If packets are excessively delayed they can be considered lost. For this work it was necessary to choose a delay threshold beyond which a packet is considered lost. The threshold chosen was twice the maximum one way delay with the addition of the encoding and decoding delays. One way delay values below 150ms do not have an affect on the interactivity or quality of a VoIP call [10]. Therefore depending on the codec used the time constraint value was between 300 ms and 350 ms.

For each probe train a percentage loss calculation is made. The \( I_e \) parameter can be obtained based on the loss calculation and the codec from the values given for each codec in [13].

D. Handover Decision

Every \( T \) seconds the E-model algorithm calculates a MOS for each network link. These values can then be used to make a handover decision. The MOS values are passed to another algorithm to decide which network link to use. Essentially the handover mechanism will select the network link that offers the best quality of service (QoS). The QoS metric for call quality is the MOS and so this is used to make the handover decision. A small amount of hysteresis is used to prevent rapid handovers due to small network fluctuations. The hysteresis value chosen for the presented results was 0.1 MOS.

The handover is done using a standard feature of SCTP that allows a new primary address to be selected from the list of alternate addresses. This feature allows the primary interface, through which all the data is flowing to be changed. The change does not cause any delay or interruption to the flow of data to the endpoint. This makes the handover seamless and transparent to the user, as can be seen in [14].

IV. SIMULATION SETUP

Implementation and simulations were carried out using the Network Simulator 2 (NS2). An SCTP implementation for
NS2 from the University of Delaware [15] was used as the basis for the handover scheme.

To simulate wireless background traffic, constant bit rate (CBR) sources are used. Each CBR source provides a stream of packets that simulates simplex voice data. There is no inherent support in NS for full duplex application traffic. Full duplex traffic was modelled using two simplex traffic sources.

The motivation behind simulating voice calls as full duplex was based on the most commonly used VoIP program Skype, which uses full duplex (i.e. no silence suppression is used). Skype uses full duplex for two reasons. Transmitting silent packets maintains UDP bindings at NAT (Network Address Translation). Also, if data is being transmitted over TCP the silent period packets prevent a reduction in the congestion size during the silent period [16]. In a heterogeneous environment it can be expected that a VoIP application would require this functionality to overcome these unknown network conditions.

Figure 2 shows one of the simulated scenarios. This scenario consists of a wireless component from the SCTP endpoint to the WLAN AP and a wired component from the AP to the correspondent endpoints. Other simulations used the same topology but differed in the number of nodes. The wireless SCTP node is in the overlapping region of both WLANs and has a connection to both networks. The SCTP endpoint is where the VoIP call from the wireless SCTP node would be terminated, and so it is the quality of the link to this endpoint that is being measured by the wireless node.

Each CBR source was set up with particular application layer packet sizes and data rates. The packet size and data rate in each case depends upon the codec being modelled.

In the standard distribution of NS it is not possible to individually configure each wireless node to use a different transmission rate. Rather, all wireless nodes must use the same transmission rate. In this work it was required to simulate multiple wireless nodes using different transmission rates. This was done by modifying the packet payload and application layer data rate of each CBR source. The packet size and data rate for each node was increased so that it applies the same load to the network as if it was transmitted at a lower rate. In tables 2 and 3 the data rates and packet sizes for G.711 and G.729 to simulate VoIP at different transmission rates are shown. The simulations were configured such that all wireless LAN transmissions were at a rate of 11Mbps.

<table>
<thead>
<tr>
<th>Table 2 - Packet Sizes for G.711</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transmission Rate</td>
</tr>
<tr>
<td>(Mbps)</td>
</tr>
<tr>
<td>11</td>
</tr>
<tr>
<td>5.5</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 3 - Packet Sizes for G.729</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transmission Rate</td>
</tr>
<tr>
<td>(Mbps)</td>
</tr>
<tr>
<td>11</td>
</tr>
<tr>
<td>5.5</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>1</td>
</tr>
</tbody>
</table>

The accuracy of the above approach was verified through simulation. Simulations were performed in which all nodes communicated with the AP at each of the standard wireless LAN transmission rates. Simulations were also done using the packet size modification to model the use of the same data rates. The results of these simulations concurred, thus verifying the validity of the approach.

Each simulation was run 5 times for 1000 seconds. Results obtained were averaged over these 5 simulations.

V. RESULTS

Simulations were performed to verify the accuracy of the SCTP probing mechanism for each codec over a varying number of VoIP calls. The probing mechanism predicts what the quality of a VoIP call will be if added to the WLAN network. As can be seen in Figures 3 and 4 there is high correlation between the actual MOS experienced by a CBR source present in the network and the MOS estimated by the proposed mechanism.

In [7], [8] and [9] it was found that an 802.11b WLAN can support 6 VoIP calls using the G.711 codec and 7 calls using the G.729 codec, both with a frame size of 10ms. The capacity...
of the 802.11b WLAN predicted by the proposed mechanism agrees with the previously published work.

The accuracy of the proposed probing mechanism operating at transmission rates lower than 11Mbps was also investigated. The previously discussed modifications to the payload packet size were used to model the varying data rates of 802.11b.

Figures 5 and 6 show the MOS predicted by the proposed mechanism and the MOS experienced by a VoIP call in the network at different data rates. This demonstrates that the proposed mechanism operates successfully at multiple data rates. As can be seen, using low data transmission rates greatly reduces the capacity of the WLAN.

A simulation scenario was chosen to verify the operation of mechanism in heterogeneous transmission rate environments. Four calls of various transmission rates were placed on an 802.11b WLAN AP, 2 calls at 11Mbps, 1 call at 5.5Mbps and 1 call at 2Mbps. The effect on performance of adding a further call from each transmission rate was then determined for each codec. Figures 7 and 8 show the results of those simulations.

As can be seen when using G.711 any call added at a transmission rate below 11Mbps has a detrimental effect on the quality of all calls in the network. However, when the G.729 codec is used the quality remains high for all call additions except a call with a 1Mbps transmission rate.

This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE ICC 2006 proceedings.
A simulation using the network topology shown in Figure 2 is now presented. The results of which are shown in Figure 9. The same call scenario was used as described previously. Both networks had the same 4 call scenario placed on it from time 0, giving a mean MOS value of 4.38 in each. Initially, WLAN1 was set as the primary network. The time interval between handover decisions as previously discussed was set at 2 seconds. The simulation had duration of 1000 seconds. At time 500 seconds a G.711 VoIP call operating at a transmission rate of 1Mbps was added to WLAN1. As can be seen in Figure 9, at 500 seconds when the 1Mbps call is added the MOS of WLAN 1 drops down to a low MOS value of 2.31. At time 502.59 seconds the proposed mechanism performs a handover to WLAN 2 in which the call quality is still high, thus demonstrating a handover.

VI. CONCLUSION

In this paper a handover mechanism for VoIP over heterogeneous transmission rate IEEE 802.11b networks has been proposed. Modifications were made to the E-Model and SCTP to allow real time quality assessment of VoIP calls over multiple data networks.

The proposed mechanism predicts what the quality of a VoIP call will be if added to the network. Results were presented showing a high correlation between the estimated MOS and the MOS experienced by a VoIP node added to the network. The VoIP capacity of 802.11b predicted by the scheme agrees with previously published work. Simulations verifying the operation of the mechanism under various load conditions at different transmission rates were also presented.

A simulation scenario was chosen to demonstrate the accuracy of the mechanism operating in a heterogeneous transmission rate environment over 802.11b. Finally, a simulation showing a handover based on the MOS of the available networks was presented.

ACKNOWLEDGMENTS

The support of the Irish Research Council for Science, Engineering and Technology (IRCSET) and the Informatics Research Initiative of Enterprise Ireland is gratefully acknowledged.

REFERENCES