Performance Study on Audio Codec and Session Transfer of Open Source VoIP applications

Cheng-Suan Lee, Khong Neng Choong, So Gean Koh, Chee Onn Chow, and Mazlan Abbas

***Abstract*—**Voice over Internet Protocol (VoIP) application or commonly known as softphone has been developing an increasingly large market in today's telecommunication world and the trend is expected to continue with the enhancement of additional features. This includes leveraging on the existing presence services, location and contextual information to enable more ubiquitous and seamless communications. In this paper, we discuss the concept of seamless session transfer for real-time application such as VoIP and IPTV, and our prototype implementation of such concept on a selected open source VoIP application. The first part of this paper is about conducting performance evaluation and assessments across some commonly found open source VoIP applications that are Ekiga, Kphone, Linphone and Twinkle so as to identify one of them for implementing our design of seamless session transfer. Subjective testing has been carried out to evaluate the audio performance on these VoIP applications and rank them according to their Mean Opinion Score (MOS) results. The second part of this paper is to discuss on the performance evaluations of our prototype implementation of session transfer using Linphone.

***Keywords*—**audio codec, softphone, session transfer.

I. INTRODUCTION

WITH the increasing demand for multimedia contents and with the increased usage of devices such as smart phones, PDAs, laptops, desktop etc., a user in a home or enterprise network has a number of devices at hand which most probably have the capability of offering similar multimedia services but with varying capabilities. Hence, it is logical that user may choose one device at one time for using a particular multimedia service, and change to another device at another time. It is also possible for the user to start a multimedia session on one device and later transfer the same session over to another device without interruption to the session. In the context of voice call, this is commonly known as call forwarding or transferring. Such concept is generalized as session transfer in this paper.

The common implementation of session transfer requires user to know the IP address of the target device, and to enter the IP address during the session as to perform the session transfer. Such restriction degrades user experience as far as seamless mobility and ubiquitous computing is concerned because it interrupts the ongoing application session. The situation gets worst if the user is unaware that he/she has entered the wrong IP address or SIP URI of the target device to be transferred to, this is particularly true to the elder user group. A proposed solution in this paper is to simplify and automate the process of session transfer across devices by addressing the above limitations. The proposed system is to make use of the location server to track the location of the active user device and those surrounding devices such that whenever the session transfer request is triggered, the current session on the active user device could be seamlessly transferred to a selected target with only a single button press, which the nearest device to the active device. Figure 1 shows the SIP interactions of location-assisted session transfer between the various SIP entities in a network. The original session was between the mobile device and the corresponding node (CN). To trigger a transfer as in one of the use cases, the mobile user will go near to the target device and press a designated button on the mobile device; this will trigger a series of message exchange to enable the transfer. This includes requesting the SIP URI of the nearest device (i.e. the target device) followed by re-establishing a new session from the CN to the target device.

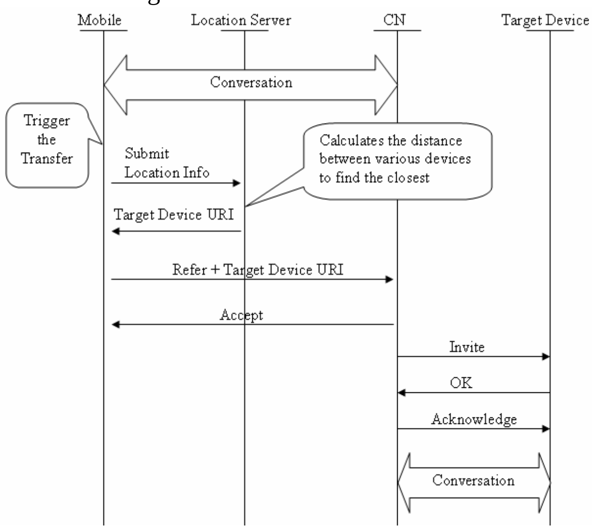

Fig. 1 Location-assisted Session Transfer from Mobile Device to the Target Device [14].

A prototype implementation of the location-assisted session

Cheng-Suan Lee[1], Khong Neng Choong[2] and Mazlan Abbas[3] are with MIMOS Berhad, Technology Park Malaysia, 57000 Kuala Lumpur, Malaysia (phone: +603-8995 5000; fax: +603 86579674; e-mail: csuan.lee@mimos.my[1], kn.choong@mimos.my[2], mazlan.abbas@mimos.my[3]).

So Gean Koh was with University of Malaya as undergraduate student (e-mail: gph_koh@hotmail.com).

Chee Onn Chow is with the Electrical Engineering Department, University of Malaya, 50603 Kuala Lumpur, Malaysia (phone: +603 796774457; e-mail: cochow@um.edu.my).

transfer using the open source Linphone, and a customized location server serving static location information have been developed at the initial phase. The objective of this paper is to describe the experience gained throughout the development cycle, starting from the selection of the right open source VoIP to work with, to the design and implementation, as well as to the stage of conducting performance evaluation.

The remaining of this paper is organized as follows. In Section 2, selection criteria on finding an open source softphone that meet the requirements is described and the evaluation on MOS comparison is emphasized as one of the important parameters. This includes description on the testbed setup, the test methodology and results. In Section 3, further discussion about the performance evaluation and results of the prototype implementation. Related works will be highlighted in Section 4 followed by Section 5 to summarise the paper.

II. Performance evaluation of softphones

A. Selection Requirements

Three requirements on the selection of an open source softphone have been imposed to implement the idea of location-assisted session transfer. The three requirements are as follows:

- Must deliver good VoIP audio quality
- Must support both video and audio streams (video conferencing capability)
- Ability to support both IPv4 and IPv6 (optional) addressing

For a VoIP call, the audio quality during conversation is the main concern of all users. Thus, it is very important to conduct an audio performance evaluation as to determine the Quality of Service (QoS) or the Mean Opinion Score (MOS) value for VoIP. The four open source softphones that have been evaluated are Ekiga [1], Kphone [2], Linphone [3] and Twinkle [4]. Different methods and techniques have been used to process the audio signal in the audio codec in these four different softphones. In short, there is a tradeoff to be made here such that applying a low compression codec will scarify higher bandwidth but incur less processing time. On the contrary, adopting a high compression codec will conserve network bandwidth but impose high processing requirement.

IPv6 is a consideration here because the proposed solution is wanted to possibly leverage on the advantages brought by this next generation IP addressing protocol. Among others is the increased of address space, support of QoS, security enhancements etc. Most important, with IPv6 each of the devices could be directly addressed and connected, and hence getting closer to realize the ubiquitous computing society. The softphone is needed to be able to support both video and audio streams for two simple reasons. First, video imposes higher bandwidth requirement as to challenge the implementation and second, softphone with video support looks more compelling.

B. Test Methodology

Out of the 3 selection criteria, audio performance is the only one that requires testing and performance comparison. From the literature, there are 2 common ways to conduct such evaluation, either in a subjective manner where human users are asked to judge the quality or performance of different audio streams, or in the objective manner where automation tool is used to measure the packet flow, delay, jitter etc so as to determine the audio quality.

Some objective tests using VQManager [5] have been conducted. However, there were some unexpected problems running VQManager with two of the softphones. Hence, the subjective testing method in this paper is approached until the issue is properly resolved.

Under this subjective testing, a group of 40 undergraduates in the age group of 20-24 years old have taken part as the respondents to rate the audio quality. They were divided into 6 test groups randomly where the subjective tests were carried out in 6 different time slots for 5 days, so as to guarantee the reliability of the test result. The audio performance of the softphones is a qualitative parameter and requires human interaction; therefore this subjective testing is necessary to determine the audio performance. Upon the completion of this subjective testing, the audio performance ranking of these four softphones can be determined. From there, the selection of one softphone for the implementation of location-assisted session transfer is then justified.

Listening-opinion test method as stated in ITU-T recommendation P.800 [6] has been chosen. In the listening-opinion test, a group of listeners will listen to some voice samples and later rate their audio quality according to MOS scale in a questionnaire prepared as shown in Appendix I. By averaging the summation of the total individual scores rated by them, the MOS value for that sound sample can be obtained.

C. Testbed Setup

The tests were conducted using both the campus-wide and private lab networks in order to capture the impact of different network's traffic on the softphone's audio performance. The campus-wide network and private lab network configuration are shown in Figure 2 and 3.

Three PCs (one server and two softphone clients) with pre-installed Ubuntu 7.10 operating system are used in the testbed. The server PC shall serve as the SIP server running SER [7] while the two client PCs run the four softphones - Ekiga 2.0.11, Kphone 4.2, Linphone 1.7.1 and Twinkle 1.0.1 in sequence. Codec PCMU is set to be the default codec in all these softphones. Each of the client PCs is equipped with an input microphone and an output speaker.

To start the test, all softphones running on the client PC must first register to the SIP server before any VoIP session is started. Once a call is established, the sound samples will then be played at one of the client PC and user at the other end will have to listen and judge the audio quality. The test procedure will be repeated for all four softphones to complete the performance evaluation and comparison.

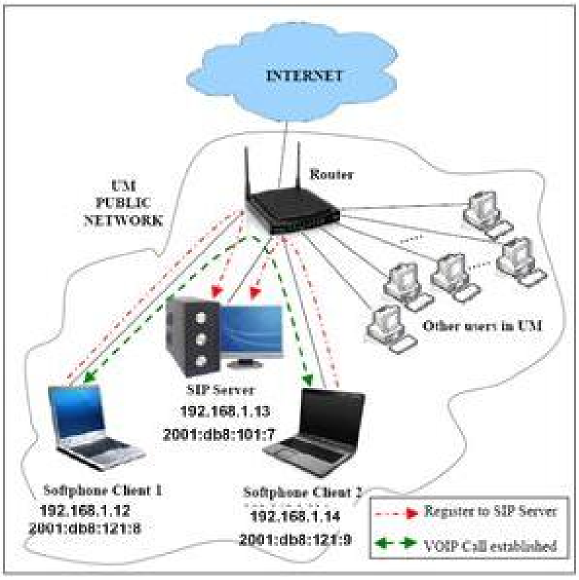

Fig. 2 Campus-wide network setup for MOS evaluation [11]

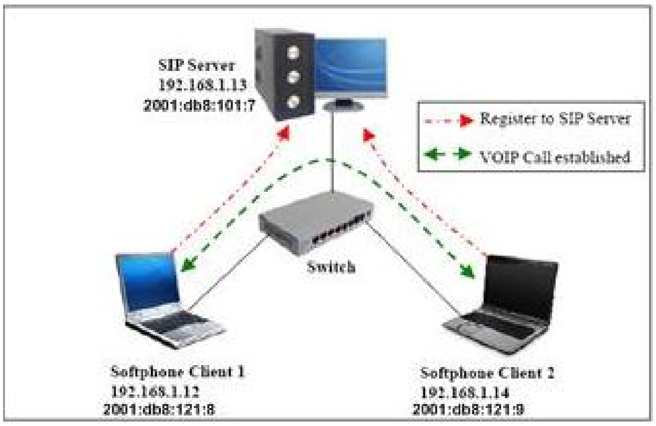

Fig. 3 Private network setup for MOS evaluation [11]

There are 3 sound samples to be evaluated. The details are as follows:

- Sound sample I: A computer-generated male voice (through Festival speech synthesis system) reading some English-language phrases as suggested by ITU-T recommendation P.800:
  - You will have to be very quiet
  - There was nothing to be seen
  - They worshipped wooden idols
  - I want a minute with the inspector
  - Did he need any money?
- Sound sample II: A 40 seconds-clipped Slow Beat song with the title of "My love will get you home"
- Sound sample III: A 40 seconds-clipped Fast Beat song with the title of "Where is the love"

D. Result Discussions

The collected results following the test procedure outlined in the previous section are shown and discussed here. Table 1 gives a summary of the overall test results for all 6 groups across four softphones tested on both campus-wide and private networks.

Overall, as a sanity check, it is clearly shown that those tests which were performed over the private network (Group 2, 3 and 4) gained higher audio performance rating as compared to the test group over campus-wide network (Group 1, 5 and 6). This shows that the softphones do get affected by the traffic of different network sizes.

It is also interesting to find out that Ekiga showed the greatest variation in MOS value across all six test groups. It was given high MOS value in Group 2, 3 and 5 (mainly above 4.0), but low MOS value in Group 1, 4 and 6 where 4 is conducted in a private network. Ekiga demonstrated inconsistency in its audio performance as compared to others. Its audio quality varied from rate 4.9 (Excellence) to 1.9 (Bad). The reasons of such variance and poor performance are unable to be identified. However, Ekiga often face difficulties in initializing calls and encountered segmentation fault from observation. Hence Ekiga is concluded unstable at the time of the test.

TABLE 1 MOS VALUE RATED FOR ALL TEST GROUPS

Group	Network	Sound sample	MOS for softphone			
			Ekiga	Kphone	Linphone	Twinkle
1	Public	I & II	2.89	2.94	4.03	4.77
2	Private	I & II	4.68	3.8	4.48	4.84
3	Private	I & II	4.93	4.27	4.67	4.6
4	Private	I & III	1.91	3.44	4.75	4.64
5	Public	I & III	4.36	3.16	4.12	3.88
6	Public	I & III	3.96	3.4	4.28	4.36
Total			22.73	21.01	26.33	27.09
Overall MOS (average)			3.79	3.50	4.39	4.52

Kphone showed overall the lowest MOS (average of 3.50) among all four softphones. This is due to the background noise during the VoIP call. It was observed that whenever a call is established, slight background noise was heard. This is due to the absence of the good noise cancellation module in Kphone. However, in some cases, Kphone outperformed Ekiga irrespective of the presence of background noise as its audio quality is better as compared to Ekiga. However, some users found it annoying and hence not favourable to Kphone. As conclusion, Kphone has acceptable audio performance with certain level of tolerance.

Twinkle and Linphone showed the highest MOS with 4.52 and 4.39 in average. Between these two softphones, Linphone had lower MOS than Twinkle as observed in test group 1, 2 and 6, but higher MOS than Twinkle for the rest. In fact, many users felt that there was no difference between these two softphones in terms of their audio performance.

E. Selection Comparison

To meet the requirements as stated earlier, it is also important to consider other parameters besides the audio quality. Table 2 tabulates the overall MOS achievements of all four softphones and their respective ranking.

The primary consideration in selecting the desired softphone would be the audio performance which was evaluated and represented by the overall MOS and the ranking since softphone is meant for VoIP calling. Kphone at the bottom of the ranking with background noise and Ekiga at the third place which was inconsistent and unstable as pointed out in the discussion above were therefore filtered out from the selection choice. Since Linphone and Twinkle had comparable audio performance, other features provided were looked into. Linphone is equipped with IPv6 support and video conferencing capability in addition than Twinkle. As a conclusion, Linphone was selected to be further enhanced with the implementation of location-assisted session transfer.

TABLE II SELECTION CRITERIA ON SOFTPHONES

Softphone	Ekiga	Kphone	Linphone	Twinkle
Overall MOS	3.79	3.50	4.39	4.52
Rank	3	4	2	1
IPv6 support	No	Yes	Yes	No
Video conferencing	Yes	With external application	Yes	No

III. PERFORMANCE EVALUATION OF SESSION TRANSFER

A. Testing Methodology

Some modifications have been done to the original Linphone with additional of session transfer capability according to the flow shown in Figure 1 [14]. Performance evaluation has been conducted using both the testbeds shown in Figure 2 and 3. The collected performance metrics are the transfer time and packet loss. Both of these metrics could be collected using the Wireshark application [8].

Transfer time could be divided into 2 types namely the estimated transfer time and exact transfer time. The estimated transfer time is defined as the time duration when a SIP REFER request had been sent and when the new conversation between the CN and the target device is established (i.e. when the first Real Time Protocol (RTP) packet is sent or received). The exact transfer time is the time between the terminations of initial call acknowledged by the last RTP packet received/sent and the establishment of the new RTP session where the first RTP packet is sent/received. By monitoring the packets captured by Wireshark, the session transfer time between the devices can be determined.

Packet loss is determined for three different phases, i.e. before, during and after the transfer so as to check for the audio performance of in the integrated Linphone. Since the session transfer is performed by terminating the initial call and establishing the new call, the packet loss before and after transfer would mean for the loss of packets during the initial call and transferred call.

By using the RTP stream analysis feature provided in Wireshark, the packet loss before and after the transfer can be easily determined. The challenge is on determining the packet loss during the transfer because Wireshark is unable to capture any packet during the transfer period as no packet gets transmitted. Hence, the packet loss during transfer was computed manually by multiplying the transfer time with the average packet transmitting rate.

Figure 4 and 5 show the two common ways to determine the transfer time from Wireshark application. Firstly, the time when Wireshark captured the desired packets (i.e. the last RTP packet in the initial call and the first RTP packet in the transferred call) are determined. The time difference between these two packets is the exact transfer time as shown in Figure 4. The estimated transfer time is computed as the difference between the time stamps when the REFER request is sent out and the first RTP packet in the transferred call is received as shown in Figure 5. Generally, the estimated transfer time is longer than exact transfer time. The exact transfer time considers only the time taken for the establishment of the transferred call after the initial call has been terminated; whereas the estimated transfer time comprises the exact transfer time and also the time taken for other SIP requests (NOTIFY and BYE) to take place before terminating the initial call.

B. Result Discussions

20 rounds of test cycles have been conducted to tackle issues of deviation, and report the average value as tabulated in Table 3.

TABLE III AVERAGE TRANSFER TIME IN CAMPUS AND PRIVATE NETWORK

Network	Estimated transfer time (s)	Exact transfer time (s)
Private	0.409	0.307
Campus	0.439	0.340

In general, the transfer time obtained for campus network is slightly longer than in the private network. The estimated and exact transfer time required in campus network is 0.439 and 0.340 seconds respectively. This is not a favourable result as a slight gap will be detected during the conversation. Hence, it is identified that further on reducing this gap to below 0.3 seconds will be investigated.

There is no packet lost before and after the session transfer. As for calculating the packet lost during the transfer, the average exact transfer time is multiplied with the average RTP packet transmitting rate as shown below.

Average RTP packets / sec = 79.432

From Table 3, the average exact transfer time on the campus network is 0.340s. Hence, the packet loss during transfer here will be,

N = 0.340 x 79.432

N = 27 packets

IV. RELATED WORKS

There have been some researches discussing the performance study on VoIP application such as [12] and [13]. In [12], Performance study on two well-known applications, i.e.: Skype and MSN Messenger have been done. However, the test is only conducted by 1 person as contrast to ours which involve 40 persons. In [13], new subjective test methodology for VoIP has been introduced. However, the test is conducted by listening the speech on web which not in the real time basis.

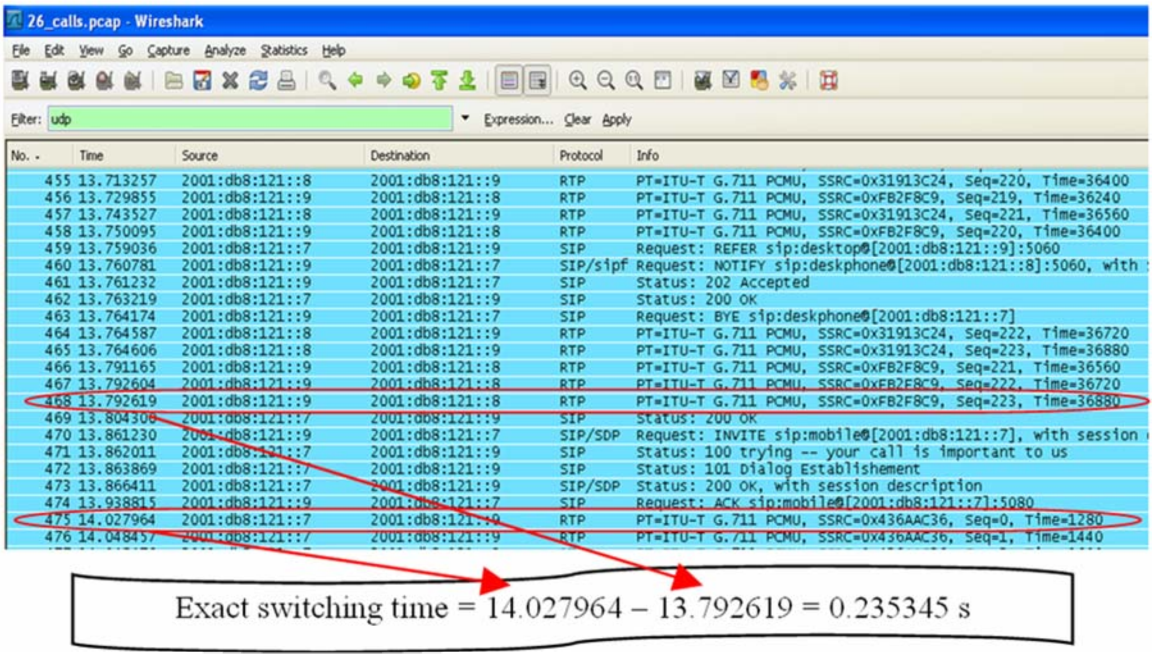

Fig. 4 To determine exact transfer time

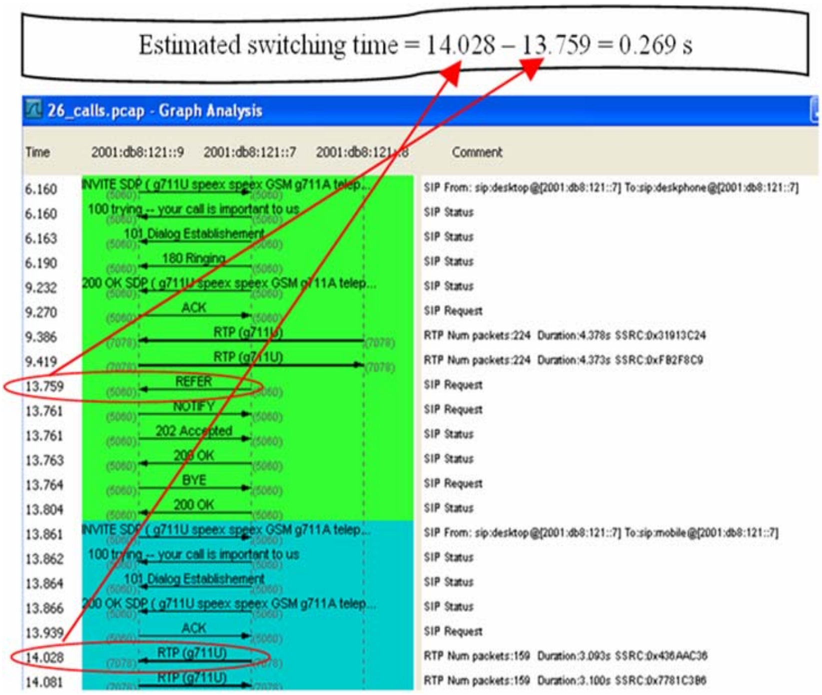

Fig. 5 To determine estimated transfer time

There are different research groups working on session transfer in different ways generally categorized into IP-based, socket-based and SIP-based approaches. Among these 3 approaches, IP-based solution is not favourable approach because it lacks of application-level information for managing session transfer. Example of such solution is Migrate [9]. Socket-based approach requires installation of additional middleware that utilizes the concept of virtual socket [10]. SIP-based approach is considered the cleaner approach because it stays at the application-level and understands the application requirements, and most importantly it also possesses the mechanism to manage session mobility with its readily available SIP messages such as the SIP REFER and SIP INVITE messages.

V. Conclusion

The concept of location-assisted session transfer has been described and the detailed selection process of finding the suitable open source VoIP to incorporate the session transfer is explained. Detailed performance evaluations have also been conducted on the prototype implementation and results were discussed with possible performance weaknesses identified to be further work on.

The next step is to work on the weaknesses found in the prototype implementation, and also incorporates the use of a fully functional location server. This will of cause include another round of performance evaluation to justify the concept of location-assisted session transfer.

REFERENCES

[1] (2009) Ekiga [Online]. Available: http://www.ekiga.org
[2] (2009) Kphone [Online]. Available: http://www.wirlab.net/kphone
[3] (2009) Linphone [Online]. Available: http://www.linphone.org
[4] (2009) Twinkle – SIP softphone for Linux [Online]. Available: http://www.xs4all.nl/~mfnboer/twinkle/index.html
[5] (2009) VQManager – VoIP quality monitoring software [Online]. Available: http://www.manageengine.com/products/vqmanager/
[6] (2009) ITU-T Recommendation P.800 – Method for Subjective Determination of Transmission Quality [Online]. Available: http://www.itu.int/rec/T-REC-P.800-199608-I/en
[7] (2009) SER – SIP Express Router [Online]. Available: http://www.iptel.org/ser/
[8] (2009) Wireshark [Online]. Available: http://www.wireshark.org
[9] A. Snoeren and H. Balakrishnan. *An End-to-end Approach to Host Mobility*. In 6th ACM MOBICOM, Boston, MA, August 2000
[10] K. Ohta, T. Yoshikawa, T. Nakagawa, Y. Isoda and S. Kurakake. *Adaptive Terminal Middleware or Session Mobility*. Ditributed Computing Systems Workshop, Proceedings, May 2003.
[11] S.G. Koh, *Evaluation and Enhancement on Softphones*, Thesis submitted to the Faculty of Engineering, Univerity of Malaya, 2009.
[12] Gao Lisha, Luo Junzhou. "*Performance Analysis of a P2P-Based VOIP software*" Advanced International Conference on Telecommunications and International Conferece on Internet and Web Applications and Services (AICT/ICIW 2006), February 2006
[13] L Sun and E Ifeachor. "*New Method for Voice Quality Evaluation for IP Networks*" Proceeding of the 18th International Teletraffic Congress, Berlin, Germany, September 2003
[14] MIMOS Patent Pending (K.N. Choong). Application number PI20084462, 2008, A Method for Session Transfer between Communication Devices

APPENDIX I

Name: ______________________________ Date: ____________________

Age: _______________________________ Time: ____________________

	Softphone 1 (Ekiga)			Softphone 2 (Kphone)		
1. Clear conversation (can differentiate every tone)	Clear ☐	Moderate ☐	Not clear ☐	Clear ☐	Moderate ☐	Not clear ☐
2. Distortion (compared to reference)	Slightly ☐	Moderate ☐	Heavily ☐	Slightly ☐	Moderate ☐	Heavily ☐
3. Background noise (affect conversation)	Slightly ☐	Moderate ☐	Noisy ☐	Slightly ☐	Moderate ☐	Noisy ☐
4. Word drops (discontinuous in conversation)	No drops ☐	Moderate ☐	Heavily ☐	No drops ☐	Moderate ☐	Heavily ☐
5. Satisfaction (willing to use)	Satisfy ☐	Moderate ☐	Not satisfy ☐	Satisfy ☐	Moderate ☐	Not satisfy ☐
6. Overall comment and suggestions:	____________			____________		

	Softphone 3 (Linphone)			Softphone 4 (Twinkle)		
1. Clear conversation (can differentiate every tone)	Clear ☐	Moderate ☐	Not clear ☐	Clear ☐	Moderate ☐	Not clear ☐
2. Distortion (compared to reference)	Slightly ☐	Moderate ☐	Heavily ☐	Slightly ☐	Moderate ☐	Heavily ☐
3. Background noise (affect conversation)	Slightly ☐	Moderate ☐	Noisy ☐	Slightly ☐	Moderate ☐	Noisy ☐
4. Word drops (discontinuous in conversation)	No drops ☐	Moderate ☐	Heavily ☐	No drops ☐	Moderate ☐	Heavily ☐
5. Satisfaction (willing to use)	Satisfy ☐	Moderate ☐	Not satisfy ☐	Satisfy ☐	Moderate ☐	Not satisfy ☐
6. Overall comment and suggestions:	____________			____________		

Rank these four softphone according to their performance (1- Best, 4- Worst)

Ekiga ________ , Kphone ________ , Linphone ________ , Twinkle ________

Thank you!