國立暨南國際大學資訊工程學系

碩士論文

以 RTP 代理伺服器為輔助的網路電話品質量測

Voice Quality Monitoring Assisted by RTP Proxy Servers

指導教授：吳坤熹博士

研究生：李霓雅

中華民國九十九年 六月

國立暨南國際大學碩（博）士論文考試審定書

資訊工程 . 學系（研究所）

研究生 李霓雅 所提之論文

以 RTP 代理伺服器為輔助的網路電話品質量測

Voice Quality Monitoring Assisted by RTP Proxy Servers

經本委員會審查，符合碩（博）士學位論文標準。

學位考試委員會

[illegible] 委員兼召集人

吳坤熹 委員

黃一泓 委員

中華民國 九十九 年 七 月 七 日

博碩士論文電子檔案上網授權書

（提供授權人裝訂於紙本論文書名頁之次頁用）

本授權書所授權之論文為授權人在 暨南國際大學 資訊工程學系 _______組 98 學年度第 二 學期取得 碩士 學位之論文。

論文題目： 以 RTP 代理伺服器為輔助的網路電話品質量測
指導教授： 吳坤熹

茲同意將授權人擁有著作權之上列論文全文（含摘要），非專屬、無償授權國家圖書館及本人畢業學校圖書館，不限地域、時間與次數，以微縮、光碟或其他各種數位化方式將上列論文重製，並得將數位化之上列論文及論文電子檔以上載網路方式，提供讀者基於個人非營利性質之線上檢索、閱覽、下載或列印。

- 讀者基非營利性質之線上檢索、閱覽、下載或列印上列論文，應依著作權法相關規定辦理。

授權人：李霓雅
簽　名：李霓雅　　　　中華民國 99 年 08 月 26 日

致謝

兩年的時間，要感謝的人、事、物太多太多。

感謝老天爺爺讓我的生命中總是出現這麼多美好的人，讓我覺得自己也未免幸運得太過頭了。

感謝吳董在兩年前相信我是團還能夠扶得上牆的爛泥。沒有當初的鼓勵，沒有今天的論文。

感謝阿爸跟媽咪，總是包容著我的任性，支持我做任何自己想做的事，我愛你們喔！

感謝 409 的大家，這兩年因為有你們，快樂加倍了，哀傷減半了。

感謝出現在我生命中的人們，希望你們一切都好。

論文名稱：以 RTP 代理伺服器為輔助的網路電話品質量測

校院系：國立暨南國際大學科技學院資訊工程學系　　頁數：33

畢業時間：99 年 6 月　　學位別：碩士

研究生：李霓雅　　指導教授:吳坤熹博士

中文摘要

自從網路電話(Voice over Internet Protocol, 簡稱 VoIP)問世以來，聲音品質一直都是網路電話服務面臨的最大挑戰。隨著網路電話的使用者越來越多，其聲音品質也就越受到關注。以往聲音品質的量測方式，都是藉由控制影響聲音品質的變數並在使用者端進行測量。然而，在真實網路上，網路電話的管理者想要在使用者端進行聲音品質量測是非常困難的一件事。而目前市面上有多家軟硬體網路電話廠商，要實做一套能夠在各家產品上都適用的聲音品質量測系統幾乎是不可能的。

在此同時，由於近年來網路位址不足的原因，Network Address Translator (NAT，俗稱網路分享器)這個解決方法廣為佈建在大大小小的各種網路中。然而 NAT 的存在造成了某些點對點的網路應用服務受到影響，網路電話服務就是其中之一。為了讓網路電話能順利穿越 NAT，目前的網路電話系統中都會建置一台 RTP Proxy Server。

在本文中，我們將提出一套利用上述的 RTP Proxy Server 為輔助，在網路上測量聲音品質的量測系統。因為此量測系統並沒有在使用者端進行任何的修改，所以並不侷限於特定的軟硬體網路電話；此系統將量測後所得到的數據記錄下來，以圖形的方式呈現，希望能夠在使用者對於網路電話的聲音品質有疑慮的時候，提供給網路電話服務的管理者一份客觀的除錯依據。

我們並以實作驗證，利用此簡易量測法所獲得的數值，相較於利用使用者端設備進行量測的正規方式，誤差不到百分之一。可說是以較低的成本，所獲得一項相當令人滿意的量測成果。

關鍵詞： 網路電話、聲音品質量測、RTP 代理伺服器

Title of Thesis：Voice Quality Monitoring Assisted by RTP Proxy Servers

Name of Institute：Department of Computer Science and Information Engineering, College of Science and Technology, National Chi Nan University   Pages：33

Graduation Time：06/2010   Degree Conferred：Master

Student Name：Ni-Ya Li   Advisor Name：Dr. Quincy Wu

English Abstract

Since the Voice over Internet Protocol (VoIP) application was introduced, voice quality has always been a big issue. As more and more people use VoIP applications, the quality issue now becomes critical. Traditionally, the measurement of voice quality has to perform the test on both client-sides. However, in a real network, it is not always possible for VoIP service providers to control the IP phone directly and measure the voice quality on client-sides. Because there are many VoIP products made from different manufacturers, right now, it is almost impossible to find a measurement system which is applicable to all VoIP products.

Meanwhile, in recent years, because of the exhaustion of IP addresses, Network Address Translator (NAT) was introduced to mitigate the shortage of IP addresses. Nevertheless, NAT causes serious problems for many peer-to-peer Internet applications, such as VoIP. Thus, VoIP applications need solutions for NAT traversal. For the past years, there are lots of NAT traversal mechanisms suggested, such as static assignment, Virtual Private Network (VPN), and relay-based proxy servers. RTP Proxy Server is a relay-based proxy server, which is the most popular one among these NAT traversal mechanisms. Nowadays, in most VoIP systems there exists a RTP Proxy Server to relay RTP packets and solve the problem of NAT traversal.

In this thesis, we design a monitoring system, named RTP-M which works with RTP Proxy Server to measure the VoIP quality. Because this system is independent with client-sides, it can be applied to any VoIP end devices. Moreover, RTP-M depicts the measured voice quality in graphical forms which are more intuitive for human beings. We hope that our RTP-M can provide VoIP administrators with the troubleshooting information when users have any complaint about voice quality.

Our implementation shows that, the voice quality measured by RTP-M, has only some negligible error compared to voice quality measured by the formal way on the client-sides. Considering the convenience and low lost, the precise is fairly satisfactory.

Keywords: Voice over Internet Protocol (VoIP), Voice Quality Monitoring, RTP Proxy Server

Table of Contents

Figure Index

Table Index

1. Motivation

As Internet technology evolves, bandwidth becomes larger and servers become powerful; many services which are unimaginable in the past become very popular now. Voice over Internet Protocol (VoIP) is one of services which were impossible 15 years ago but nowadays is widely implemented on Internet. Different from the traditional public switched telephone network (PSTN), which adopted the circuit switching for voice transmission, VoIP is an application implemented on packet-switched networks. The voice quality of PSTN is assured because it establishes a channel between two clients during communication, while others cannot share this channel. On the contrary, as its name implies, VoIP applications transmit voice over Internet which is a packet-switched network. In VoIP applications, voice were digitized and divided into small packets delivered over a shared channel; thus, the quality of VoIP application is not assured as in PSTN. Since the population of VoIP users grows rapidly as shown in Figure 1, the quality issue of VoIP application becomes more critical.

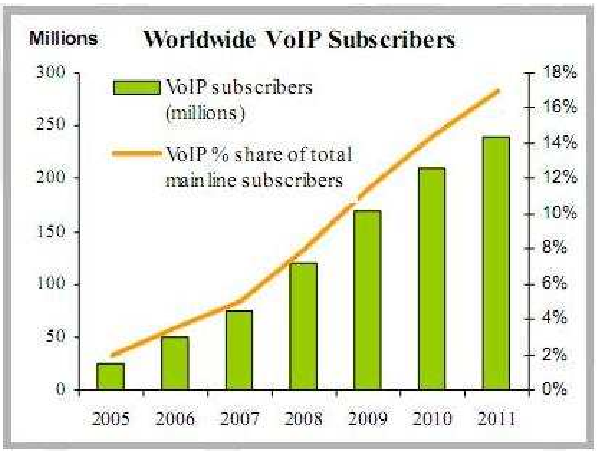

Image source:
http://www.newsterindia.com/wp-content/uploads/2008/08/voip-worldwide.jpeg

Figure 1: Population Growth of VoIP Users

Hence, we want to propose a monitoring system to measure the voice quality over Internet, and hope that this system can provide VoIP administrators with useful troubleshooting information when users have complaints with voice quality. We named our system as RTP-M. In the following, we will study a few common measurement mechanisms of voice quality in Chapter 2, and make a comprehensive survey of current VoIP architecture in Chapter 3. The illustration of RTP-M is given in Chapter 4, with experimental results shown in Chapter 5. We then conclude this thesis and describe some possible future work in Chapter 6.

2. Voice Quality Measurements

Many methods of voice quality measurement are developed for years. These methods can be roughly separated into subjectivity and objectivity. For subjective measurements, the voice quality is based on human perception, and the quality is scored by human beings. On the contrary, the objective measurements compute the voice quality by equipments with specific algorithms or mechanisms, and the objective measurement can be intrusive or non-intrusive. In this chapter we will introduce four methods of measurements including subjective measurement, and objective measurement contains intrusive method and non-intrusive method.

2.1 Mean Opinion Score (MOS)

In voice communications, the mean opinion score (MOS) which is specified in International Telecommunication Union Tele-communications Standardization Sector (ITU-T) Rec. P.800 [1] provides a numerical indication of voice quality. MOS is ranged from 1 to 5 and the higher score means the better quality. The mapping between score and voice quality is shown in Table 1.

Table 1: Mean Opinion Score (MOS)

Quality of Voice	Mean Opinion Score
Excellent	5
Good	4
Fair	3
Poor	2
Bad	1

MOS test requires a certain number of people to hear a voice, and each one of them gives a rating within 1 to 5 for the voice what they are listening to. Then an arithmetic mean is calculated, and the mean is the value of MOS. There are many restrictions while conducting MOS test, like the volume of the test room, the environmental noise, and so on.

Because MOS is a subjective measurement, it is usually time-consuming and expensive as hiring people to make estimations; besides, it cannot really reflect the impairment caused by transmitting voice over Internet, such as delay, jitter, packet loss, etc.

2.2 Signal to Noise Ratio (SNR)

SNR is an intrusive method of quality measurement. This method typically use two signals to estimate distortion and further obtain the voice quality. One of two signals is the original signal, and the other is the distorted signal which is generated by a distortion system or by delivering over Internet.

The value of SNR represents how much a signal has been corrupted by noise, and it is usually expressed by logarithmic decibel scale.

The formula of calculating SNR is originally defined to be:

$$SNR = \frac{P_{signal}}{P_{noise}}$$

Where, P_{signal} is the signal strength, and P_{noise} is the noise level. Mostly SNR defined in decibel scale is written as:

$$SNR_{db} = 10\log_{10}(\frac{P_{signal}}{P_{noise}}) = P_{signal,dB} - P_{noise,dB}$$

The advantage of SNR is easy to implement, and the disadvantage is that it is not suitable for real-time transmission. As the formula we see above, SNR is calculated by comparing two signals; however, in a real-time communication, like VoIP, it is not always possible to separately acquire the original signal and the signal transmitted over Internet.

2.3 Perceptual Evaluation of Speech Quality (PESQ)

As shown in Figure 2, PESQ, which is defined in ITU-T Rec. P.862 [2], is also an intrusive method as SNR.

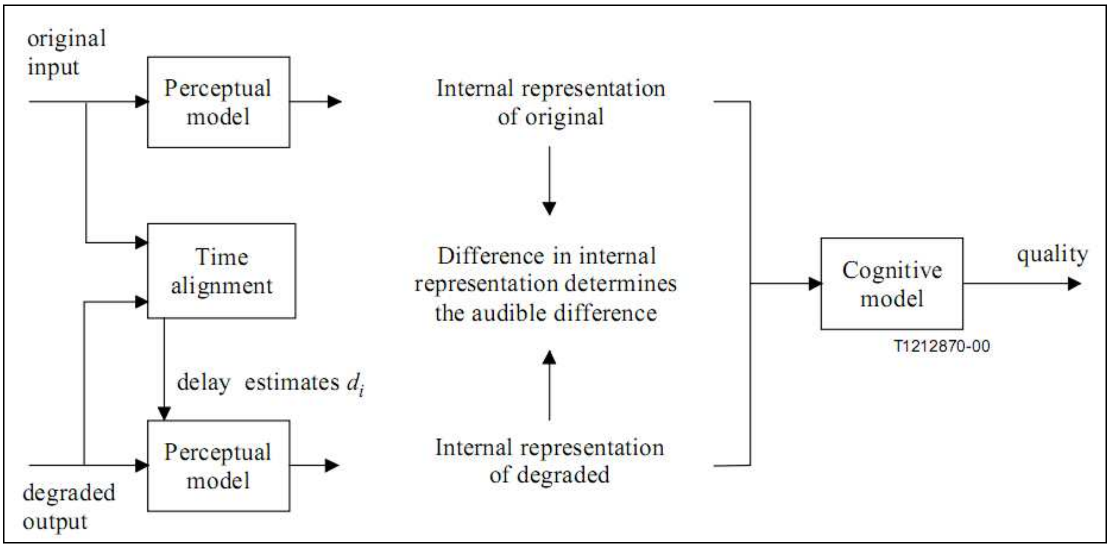

Figure 2: PESQ Model [2]

PESQ obtains a quality score by comparing an original input with a degraded output. The degraded output is generated by adding impairment factors to the original signal. These impairment factors can be loudness loss, delay, sidetone, or echo, etc. Because PESQ gives accurate predictions of subjective quality in various conditions, it is widely used by phone manufacturers, network equipment vendors and telecom operators.

2.4 E-Model

E-Model defined in ITU-T Rec. G.107 [3] is the most popular non-intrusive method of objective measurement. It considers the voice impairments caused by transmitting over Internet. The Figure 3 shows the reference of E-Model.

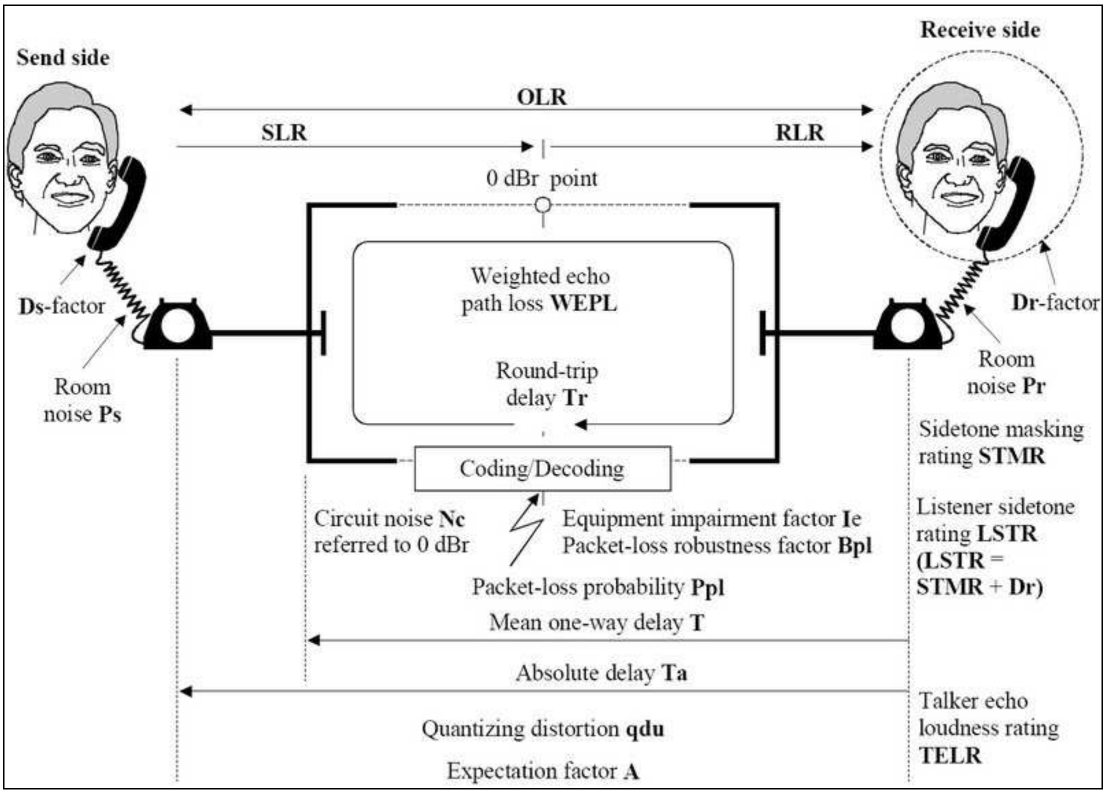

Figure 3: Reference of E-Model [3]

E-Model is computable, and the output is a scalar quality rating value *R*, which ranges from100 to 0; the 100 represents the quality is the best, and the 0 means the worst quality.

According to [4], in VoIP networks, the formula for calculating *R* is:

$$R = 94.2 - Ie - Id \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.. \qquad (1)$$

Here, *Ie* is associated with codec types and packet loss rate, and *Id* is associated with one-way delay. The formulae for *Ie* and *Id* are the following equations respectively:

$$Ie = \lambda_1 + \lambda_2 \ln(1 + \lambda_3 e) \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \qquad (2)$$

$$Id = 0.024 * d + 0.11 * (d - 177.3) * H(d - 177.3) \qquad \ldots\ldots \qquad (3)$$

In (2), λ_1, λ_2 and λ_3 are the impairment factors of different codec types as shown in Table 2, and *e* represents the packet loss rate.

In (3), *d* means the one-way delay (in milliseconds), and $H(x)$ is the step function:

$H(x) = 0$, if $x < 0$

$H(x) = 1$, if $x \geq 0$

From the above formulae, we know that we can calculate *R*, as long as the codec types, the value of packet loss rate and the value of one-way delay are known.

Table 2: Impairment Factors of G.711 and G.729

Codec	λ_1	λ_2	λ_3
G.711	0	30.00	15
G.729	10	47.82	18

Among the above four methods for assessing voice quality, a non-intrusive method is certainly the better choice, because we want to establish a monitoring system to measure the quality of voice in real-time. Therefore, we adopt E-Model as the measurement approach of our monitoring system.

3. VoIP Network Architecture

3.1 Session Initiation Protocol (SIP)

Session Initiation Protocol (SIP) is an application-layer signaling protocol, which handles the creating, modifying, and terminating of multimedia sessions. It was proposed by Internet Engineering Task Force (IETF), and defined in RFC 3261 [5]. SIP has the similar addressing form to an e-mail address, for instance sip:alice@example.com, which is called SIP Uniform Resource Identifier (SIP URI), and its messaging syntax is text-based.

SIP follows the client/server model and defines two basic network entities: user agents (UAs) and SIP servers. The UA consists of a user agent client (UAC) and a user agent server (UAS), and the difference between them is that the UAC creates and sends requests while the UAS is responsible for answering requests. On the other hand, SIP servers have logically three different types: Proxy Server, Redirect Server and Registrar, but actually the three types of SIP servers can be collocated on the same host. In the following, we assume that SIP Proxy Server and Registrar are installed in the same host, and we use Figure 4 to illustrate the flow of SIP messages in making a SIP call.

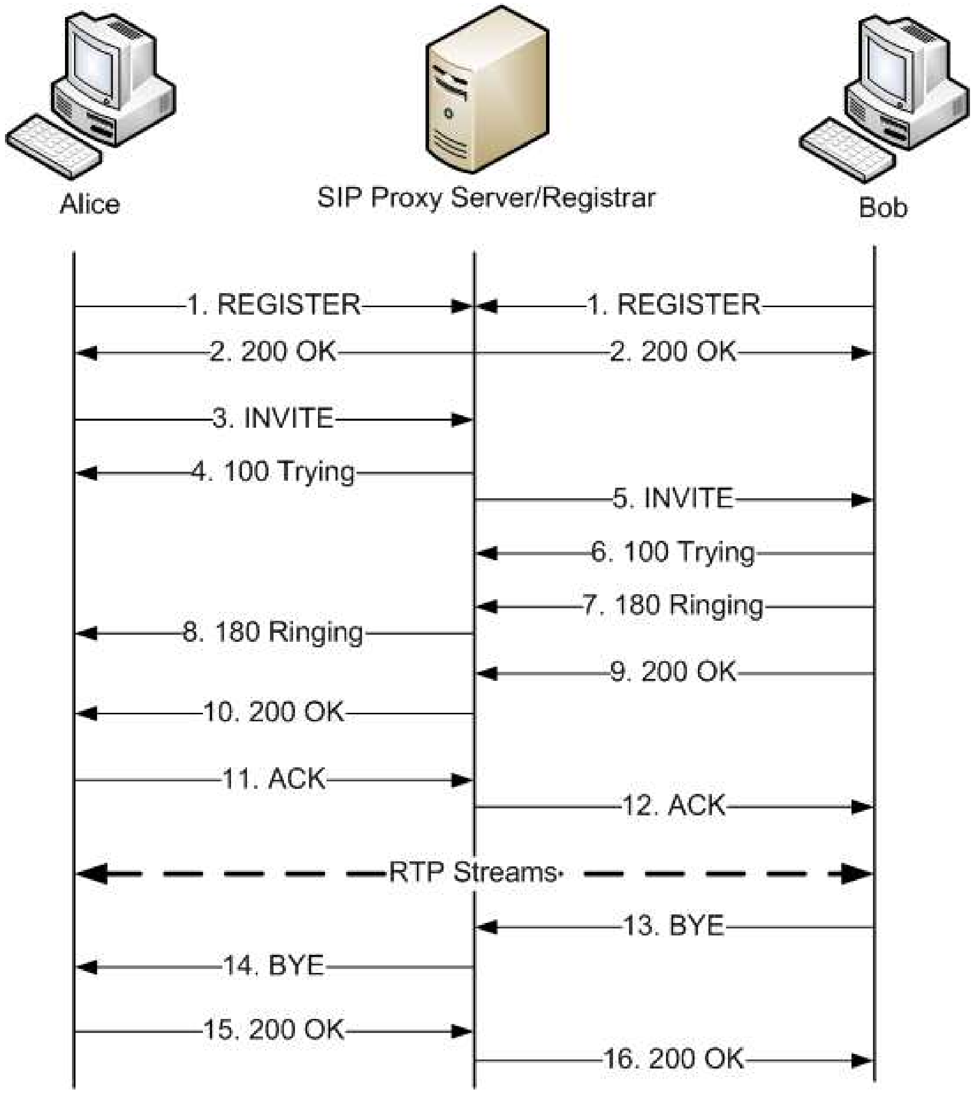

Figure 4: Example of SIP Message Flow

A VoIP conversation generally goes through the following steps:

1. First, both Alice and Bob have to send the RESIGTER requests to the Registrar.
2. When the Registrar receives the REGISTER requests, it responds 200 OK messages to Alice and Bob. Now, Alice and Bob register at the SIP Proxy Server.
3. Then, if Alice wants to invite Bob for communication, Alice sends an INVITE message to the SIP Proxy Server.
4. The SIP Proxy Server replies a 100 Trying message to Alice indicating that it has received an INVITE message and it is processing this invitation.
5. Meanwhile, the SIP Proxy Server forwards the INVITE message to Bob.

6. When receiving an INVITE message, Bob sends a 100 Trying to SIP Proxy Server to inform the SIP Proxy Server that this INVITE message is being handled.
7. Then Bob's phone rings, and it sends a 180 Ringing message to the SIP Proxy Server.
8. When the SIP Proxy Serve receives the 180 Ringing message from Bob, it forwards this message to Alice.
9. When Bob answers the call, it sends 200 OK message to the SIP Proxy Server to inform that this invitation is accepted.
10. As the SIP Proxy Server receives the 200 OK message from Bob, it forwards this message to Alice.
11. After Alice receives the 200 OK message, she knows that this invitation has been accepted by Bob; then an ACK message is sent to the SIP Proxy Server.
12. The SIP Proxy Serve receives and forwards the ACK message to Bob; at this moment, the call between Alice and Bob is established, and the RTP streams between Alice and Bob start.
13. When either party wants to terminate this call, say Bob, it sends a BYE message to the SIP Proxy Server.
14. The SIP Proxy then forwards the BYE message to Alice.
15. After Alice receives the BYE message from the SIP Proxy Server, it replies a 200 OK message to the SIP Proxy Server.
16. The SIP Proxy Server forwards the 200 OK message to Bob; now, this call is terminated.

3.2 Real-time Transport Protocol (RTP)

Real-time Transport Protocol (RTP) [6] is proposed to deal with the real-time multimedia data, like audio. The RTP header has a minimum size of 12 bytes, and the size can be extended by attaching optional header extensions. As shown in Figure 5 below, the RTP header contains fields such as version (V), payload type (PT), sequence number, timestamp, and so on.

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|V=2|P|X|  CC   |M|     PT      |       sequence number         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                           timestamp                           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|           synchronization source (SSRC) identifier            |
+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+
|            contributing source (CSRC) identifiers             |
|                             ....                              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 5: RTP Header Format [6]

The timestamp represents the sampling time of the first octet in the RTP data packet. The sequence number is randomly assigned when the first RTP packet is sent, and after that, the sequence number increments by one for each RTP packet sent. As mentioned before, we adopt E-Model as the measurement model of RTP-M, and the field of sequence number can be used to measure the packet loss rate. Another filed we used is the payload type (PT), this field has 7 bits, which identifies the codec type of RTP payload, and it uses a number (0 - 127) to indicate different codec types. A set of default mappings is specified in RFC 3551 [7]. We list several common codec types in Table 3.

Table 3: Mapping of RTP Payload Type and Codec

Payload Type	Codec
0	PCMU
3	GSM
8	PCMA
18	G.729

3.3 Session Description Protocol (SDP)

Session Description Protocol (SDP) is specified in RFC 4566 [8], it provides a format to describe the information of multimedia sessions. We can see the structure of SDP in Figure 6 below.

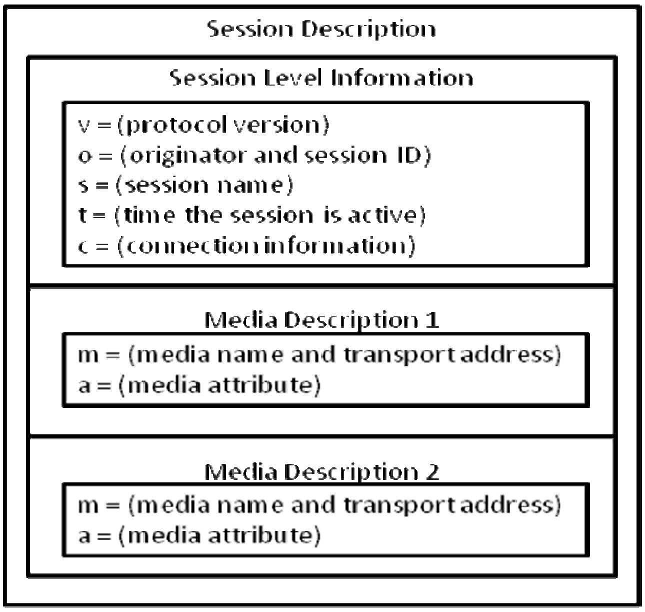

Figure 6: Structure of SDP

In Figure 6, there are two important fields in a SIP call. One is the c field which provides connection data including network type, address type, and connection address, for instance, c = IN IP4 163.22.21.194. The other is m field which contains the media type, the transport number, the transport protocol, and the media format description, for example, m

= audio 22222 RTP/AVP 0. According to the above two fields, when two SIP UAs want to communicate with each other, they know what IP address and port number to send the RTP streams, and also know the codec type of this RTP stream. To see how SDP and SIP work together to establish a multimedia session, readers may refer to RFC 3264 entitled "An Offer/Answer Model with SDP" [9].

3.4 RTP Proxy Server

Because of the foreseeable depletion of IP addresses, Network Address Translator [10] (NAT) is introduced to Internet. The advantage of NAT is making more than one device surf Internet with one public IP address; nevertheless, NAT takes away the end-to-end property of IP addresses, and fails some Internet services, like VoIP applications. Hence, the VoIP applications need a solution of NAT traversal. RTP Proxy Servers [11] are proposed to solve the problem of NAT traversal; an RTP Proxy Server is a software proxy server relaying RTP streams, and it can work together with a SIP Proxy Server which we mentioned in Section 3.1. The collaboration between a SIP Proxy Server and an RTP Proxy Server is as shown in Figure 7.

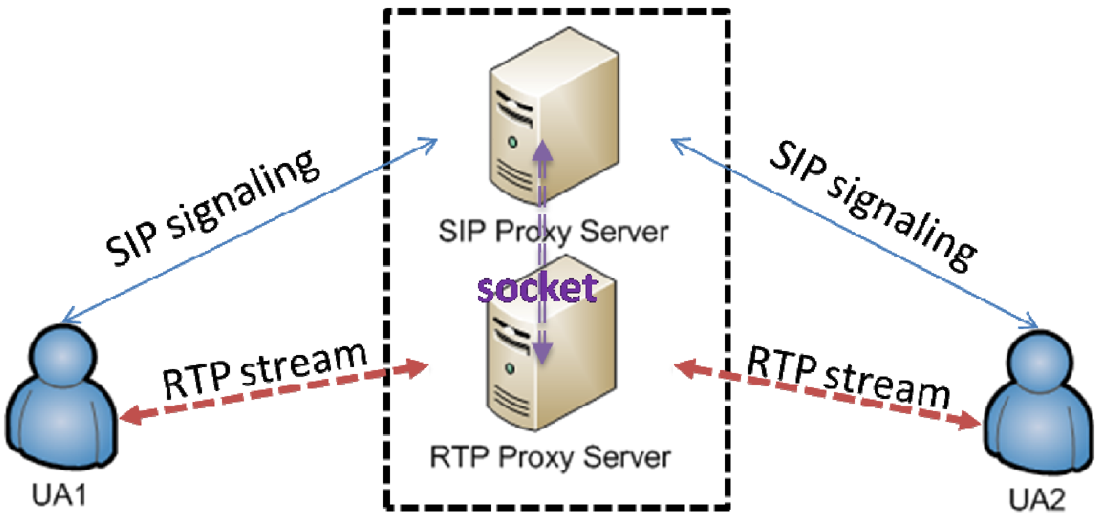

Figure 7: Collaboration of a SIP Proxy Server and an RTP Proxy Server

(In Figure 7, SIP Proxy Server and RTP Proxy Server can be installed on the same host or in different hosts, so we surrounded them with a dotted rectangle.) At first, when either UA1 or UA2 wants to communicate with each other, they followed the steps in Section 3.1 to make a SIP call. When the session is established, the RTP Proxy Server starts to relay RTP streams between UA1 and UA2.

Because the problem of NAT traversal usually exists, an RTP Proxy Server almost becomes the essential component of a VoIP system, and RTP streams will always flow through the RTP Proxy Server. Because packet loss rate and codec types can be derived from RTP packets, which is mentioned in Section 3.2, we build our voice quality monitoring system on an RTP Proxy Server. In this way, we can obtain the two parameters required to calculate the R value in E-model, and the remaining parameter is the one-way delay. In the next chapter, we will briefly illustrate the design of RTP-M and give assumptions in calculating the one-way delay; then experiments are conducted to verify these assumptions.

4. RTP-M

As described in the previous chapter, we want to build a voice quality monitoring system on an RTP Proxy Server. However, from Chapter 2, we know that the measurement of voice quality should be arranged in client-side to obtain the end-to-end delay. Since there are lots of IP telephones, which can be hardphones or softphones, produced by different manufacturers, such as X-Lite, Linphone, D-Link, ZyXEL, and Cisco. It is very difficult to require all of these IP telephones to support the software that measure the voice quality. Therefore, we turned to think about measuring the voice quality on other Internet node, i.e., we built RTP-M on an RTP Proxy Server as described in the previous chapter. Nevertheless, according to the formula to calculate the R value in E-Model, we also need to gain the value of the one-way delay. Now the problem is, RTP-M is located in the middle of two UAs, so we cannot get the actual value of one-way delay from one UA to the other UA. In the following subsection we will show our system architecture, and propose several possible assumptions of one-way delay; then demonstrate how we can use other measurement to approximate the one-way delay according to these assumptions.

4.1 System Architecture

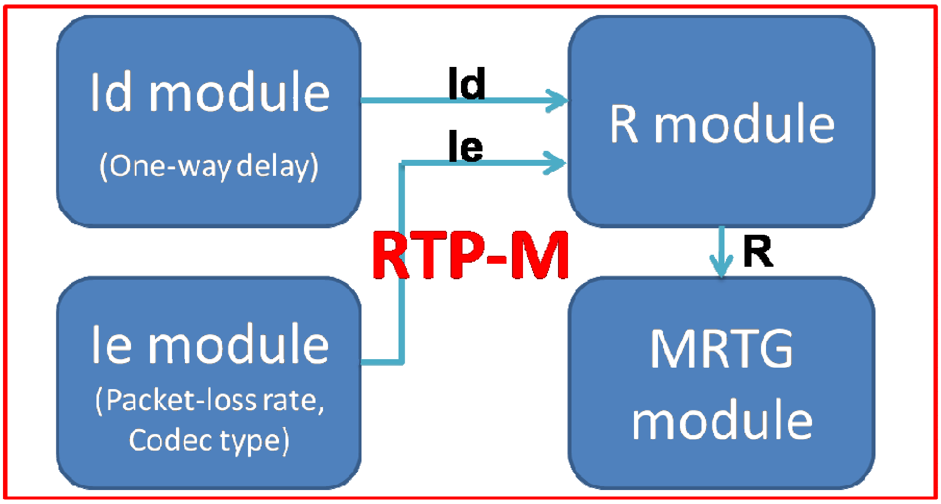

Figure 8: RTP-M Modules

As shown in Figure 8, our RTP-M has four modules. They are *Id* module, *Ie* module, *R* module, and MRTG [12] module. The first three modules, as described in Section 2.4, will calculate the *R* value of E-Model and the MRTG module will graph the trend of *R*. The following flowchart illustrates the detail process of RTP-M.

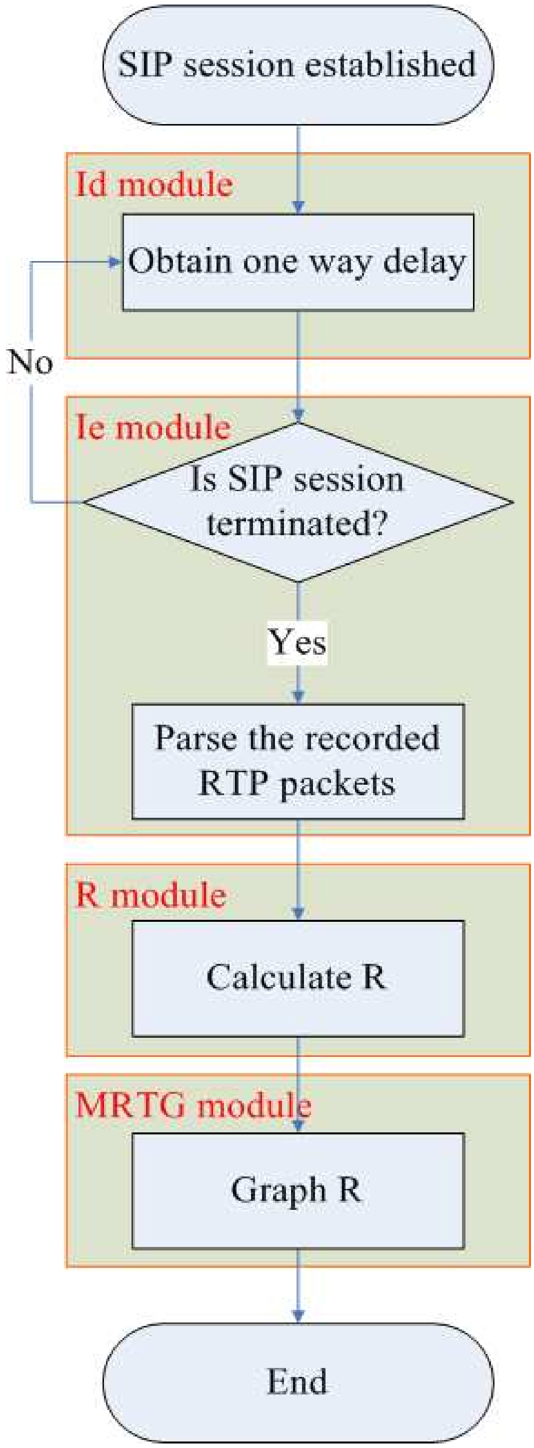

Figure 9: Flowchart of RTP-M

As shown in Figure 9:

1. When a SIP call is established, RTP-M tries to get the one-way delay. (In next section, we will discuss several solutions for obtaining the value of one-way delay.)

2. If one of two UAs wants to terminate this call, the RTP-M starts to parse the recorded

RTP packets relayed by the RTP Proxy Server, and gains the codec type and packet loss rate from the fields of payload type and sequence number in RTP headers.

3. After calculating the *Ie* and *Id* according to Formula (2) and Formula (3) in Section 2.4, *R* can also be estimated by Formula (1) in Section 2.4.
4. As the MRTG module receives the values of *R*, it graphs the trend of *R* as an image file to be shown in a webpage.

4.2 Assumptions of One-Way Delay

In this section, we will discuss three possible ways to obtain the value of one-way delay:

1. Ping Command:

   Ping command is usually used to probe whether a remote device is alive or down. When a local machine executes this command, Internet Control Message Protocol (ICMP) echo request packets will be sent to the remote device. As the remote device received an ICMP echo request, it will return an ICMP echo reply and the local machine will show the ping statistics containing the value of round-trip time (RTT), as shown in Figure 10. We can divide the value of RTT by 2 as the assumed one-way delay.

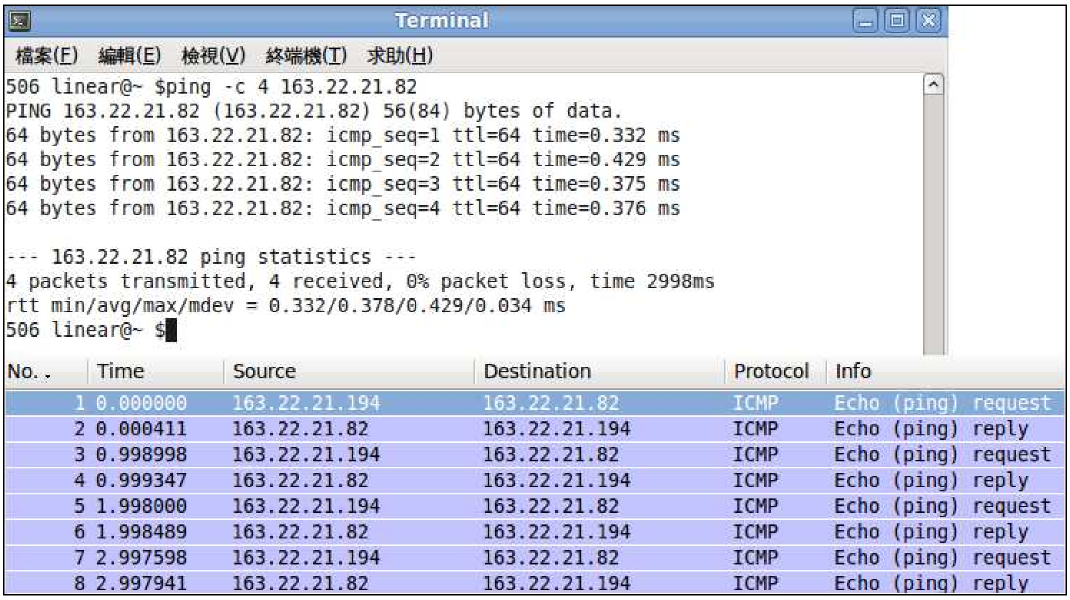

Figure 10: Ping Command

2. UDP ping command:

   There is one problem in the previous approach: when the remote device is located behind a NAT, the ICMP echo request can only arrive at the NAT instead of the remote device under consideration; hence, we may use the UDP ping command to solve this problem. The operation of this command is to send a UDP packet to the remote device. Even if the remote device is located behind a NAT, the UDP packet can also be relayed to the remote device by NAT. If an ICMP unreachable is returned from the remote device, we can also obtain the RTT. However, our experiments show that, lots of devices will not return the ICMP unreachable message, but quietly discard this UDP packet when receiving such a UDP packet. (See Figure 11)

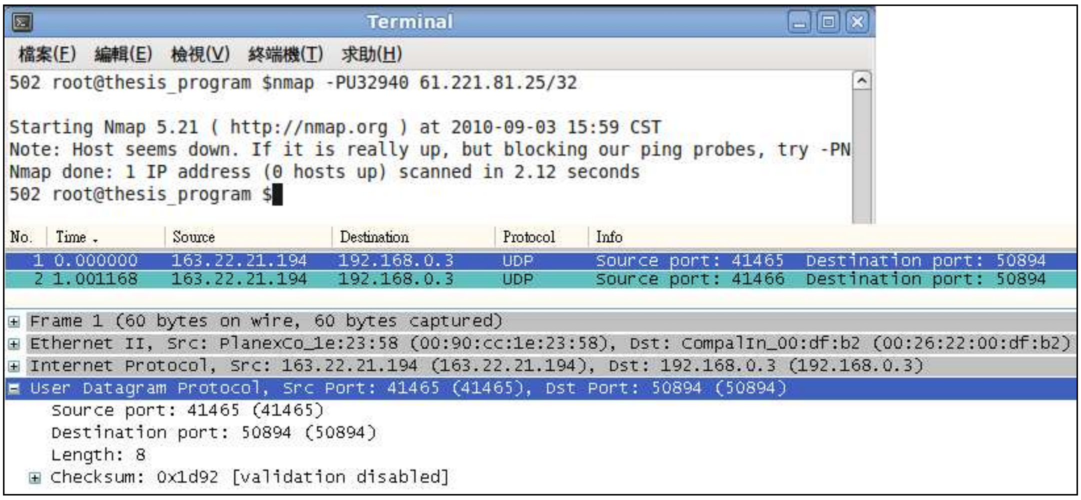

Figure 11: UDP Ping Command

3. SIP OPTIONS:

   SIP OPTIONS is a SIP method, which is used to query the ability of UAs. This approach solves the problem which the previous two approaches suffer: (See Figure 12)

   a. If a UA is behind the NAT, the SIP OPTIONS request still can reach the UA.

   b. According to [5] when a UA receives an OPTIONS request, the UA must return a response.

   After we obtain the round-trip time for the SIP OPTIONS request/response, we divide it by 2 as the assumed one-way delay.

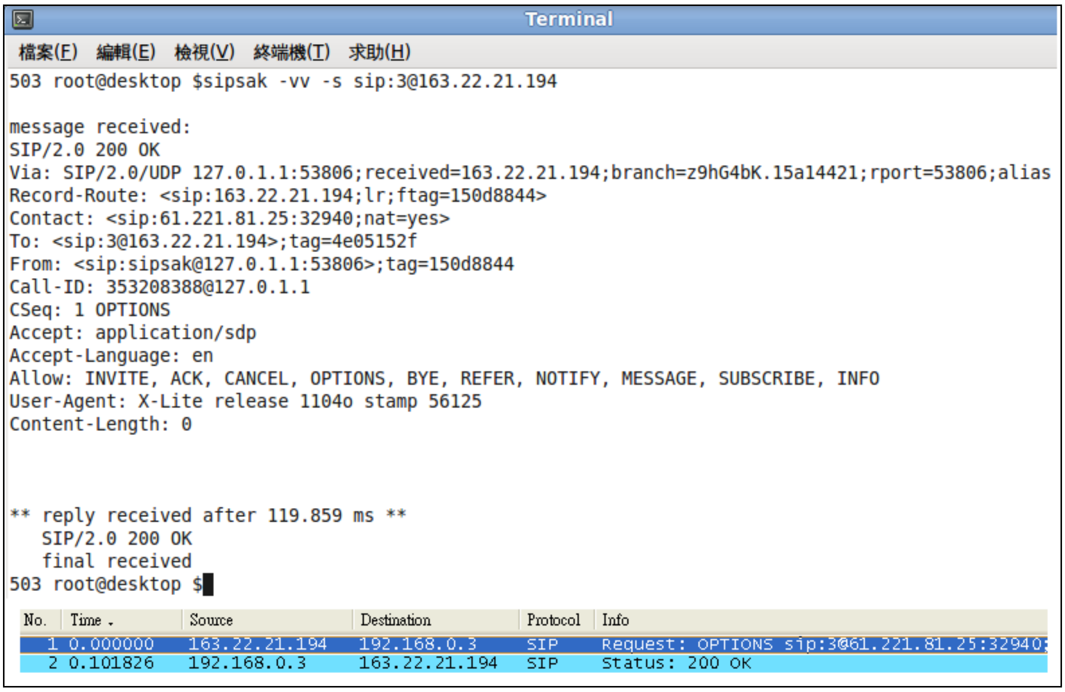

Figure 12: SIP OPTIONS Method

5. Verification and Implemented Result

As we described in Section 3.4, the RTP Proxy Server is located in the middle of two UAs, so the real value of one-way delay cannot be precisely obtained on the RTP Proxy Server. Hence, we make an assumption that using the RTT/2 from ping command or dividing the reply time by 2 obtained from SIP OPTIONS method as an approximation for the one-way delay. In this chapter we use a program called DIST-V to verify these assumed one-way delay. Besides, we will introduce a commercial VoIP monitoring software, and compare it with DIST-V and RTP-M; then we show the final implemented result of our RTP-M.

5.1 DIST-V

DIST-V [13] is a program that originally used to do VoIP stress testing. The architecture is shown in Figure 13 below.

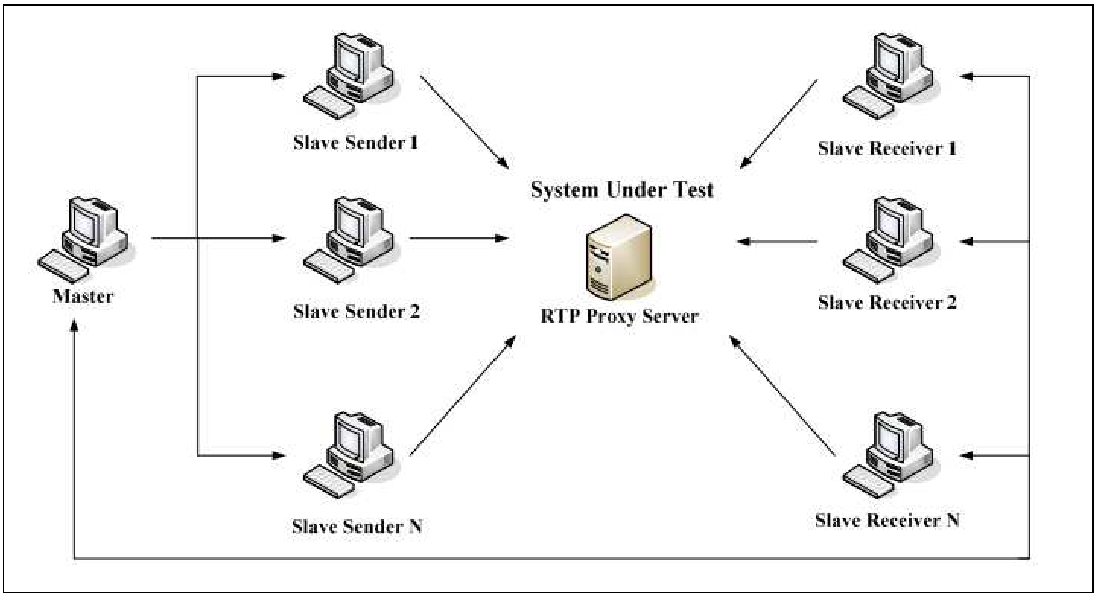

Figure 13: The Architecture of DIST-V [13]

The DIST-V uses a Master to give orders to a set of senders and receivers. Before the

Master orders the senders/receivers to establish SIP calls, the senders/receivers will synchronize their time with the same NTP [14] server. In this way, the system time of senders/receivers is synchronized; then the Master orders senders to send RTP packets with the assigned payload type, and these RTP packets were sent to receivers relayed by the RTP Proxy Server. The senders will report to the Master for the timestamp when RTP packets are sent, and the receivers also report to the Master for the timestamp when RTP packets are received, so the Master can get the value of one-way delay. Meanwhile, the Master control the number of RTP packets sent, and obtain the number of RTP packets received from receivers, therefore the Master knows the packet loss rate. Three parameters including one-way delay, codec type, and packet loss rate described in E-Model are all known by the Master, thus DIST-V can obtain the accurate value of R.

5.2 VQManager

VQManager [15] is paid software for monitoring VoIP voice quality developed by ZOHO Corporation. The cost is charged by the number of licensed phones; minimum should be 10 phones, and the cost is NTD 65,000. It can be installed on Windows or Unix-like systems.

Figure 14: Screenshot of VQManager

Figure 14 shows the screenshot of VQManager, where voice quality is represented by MOS and *R* Factor. The user interface is pretty, but there are some restrictions when using VQManager to monitor the voice quality. First, the communication time should be long enough. Second, two UAs cannot be located behind the same NAT. The third restriction of VQManager is that the way it calculates the value of one-way delay is based on the RTCP [16] packets. However, some hardphones such as Cisco 7960 IP Phones, do not correctly provide the timestamp in RTCP headers for calculating one-way delay, and D-Link DPH-150SE which we used to experiment does not even send the RTCP packets. In other words, if the two UAs use Cisco 7960 IP Phones or D-Link DPH-150SE to make a SIP call, then the VQManager will set the value of one-way delay as zero, and obtain an unreasonably high MOS value.

5.3 Comparisons

In this section, we arrange experiments that compare the *R* value of DIST-V, VQManager, RTP-M (using ping), and RTP-M (using OPTIONS).

First, we list our testing components

- User Agents:
  - DIST-V (Master/Receiver/Sender on the same device)
    1. Hardware:
       - CPU: Intel (R) Atom™ CPU N280 1.66GHz
       - Memory: 2GB
    2. Software:
       - Operating System: CentOS 5.5
       - DIST-V
  - IP Phone:
    1. D-Link DPH-150SE
- RTP-M/VQManager:
  - Hardware:
    1. CPU: Intel(R) Pentium(R) Dual CPU E2180 @ 2.00GHz
    2. Memory: 1GB
  - Software:
    1. Operating System: Ubuntu 9.04
    2. OpenSER: 1.2.1-notls
    3. RTP Proxy: 1.2.1
    4. RTP-M

Then we do the experiments in three cases:

A. SIP Proxy Server and UAs on the same subnet (See Figure 15)

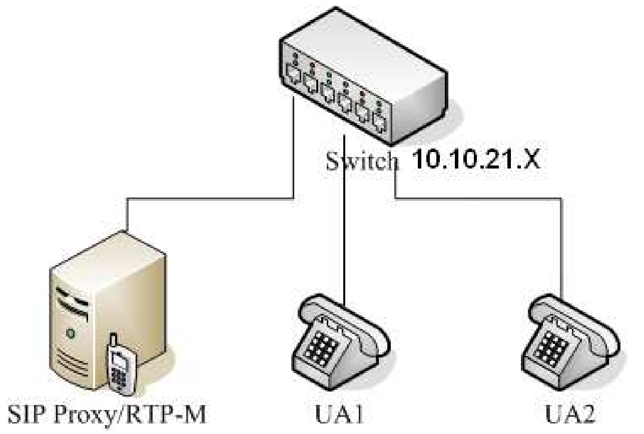

Figure 15: SIP Proxy Server and UAs on the Same Subnet

B. SIP Proxy Server and UAs on different subnets (See Figure 16)

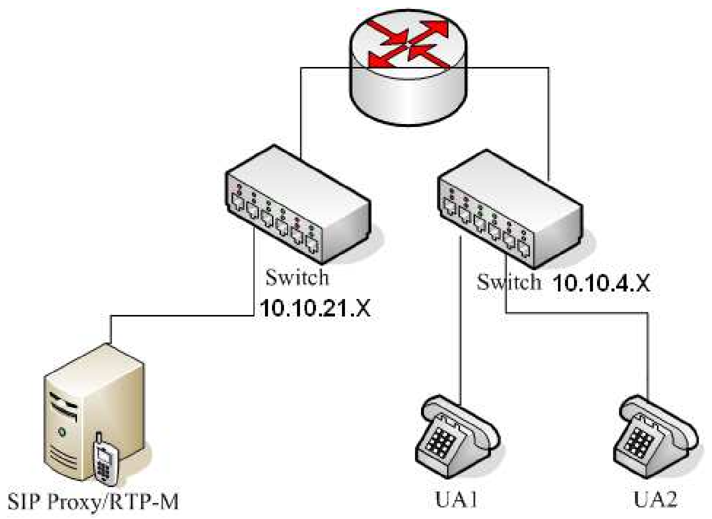

Figure 16: SIP Proxy Server and UAs on different Subnets

C. SIP Proxy Server and UAs on different WANs (See Figure 17)

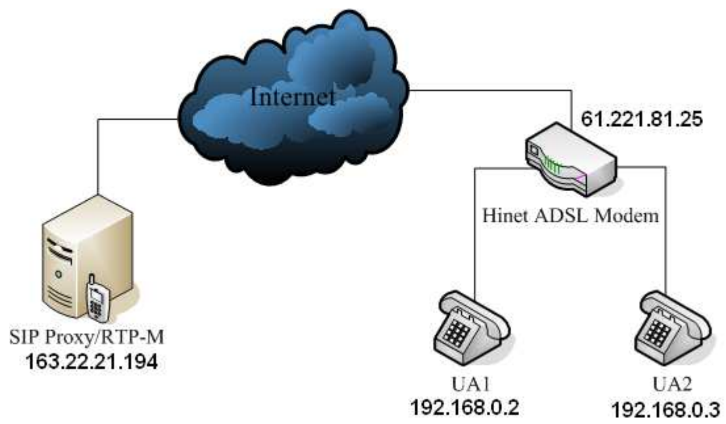

Figure 17: SIP Proxy Server and UAs on different WANs

In the above three cases, we generated continuously 30 SIP calls to calculate the average *R* of DIST-V, VQManager, RTP-M (using OPTIONS), and RTP-M(using ping). The duration of each call is 10 seconds. The experiment result is shown in Table 4 and the trend of R is shown in Figure 18.

Table 4: Experiment Result of R

	CaseA	CaseB	CaseC
RTP-M(using ping)	94.13171	94.10805	93.87571
VQMananger	93.83784	93.93333	93.88889
DIST-V	93.98203	93.98176	93.73007
RTP-M(using OPTIONS)	93.40520	93.36182	92.86979

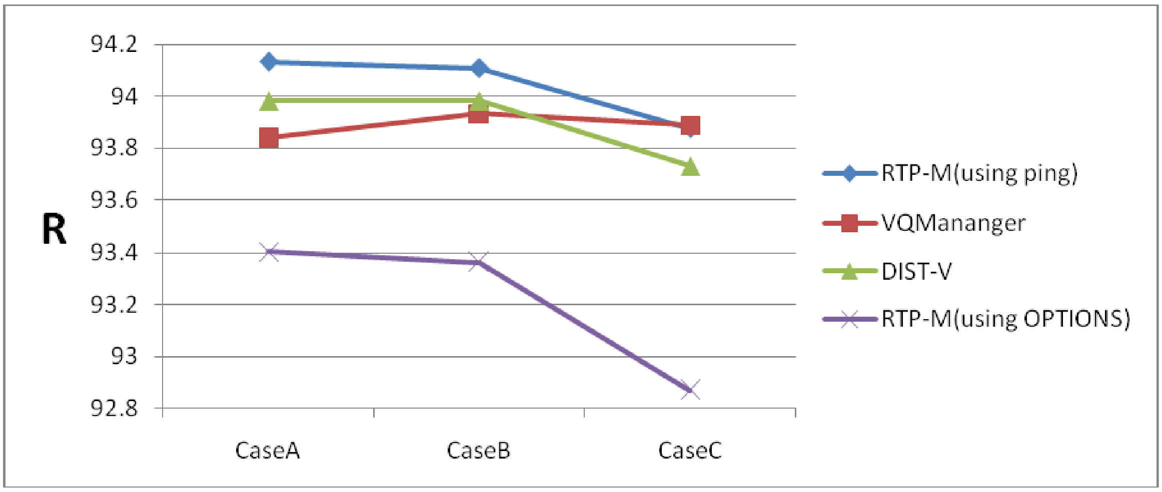

Figure 18: Trend of R

From Figure 18, we can observe regular pattern that the *R* value of DIST-V is between the *R* value of RTP-M (using ping) and the *R* value of RTP-M (using OPTIONS). It is because that when using ping command to get the value of one-way delay, the packet of ICMP echo request will not reach the UAs behind NAT, hence the one-way delay is shorter than DIST-V, and thus *R* is larger than DIST-V. On the other hand, when UAs received SIP OPTIONS requests, UAs have to parse the packets and compose SIP responses for return.

This process takes extra time, so the one-way delay of RTP-M (using OPTIONS) is larger than DIST-V, and consequently the R is smaller than DIST-V. Besides, the experiment result of VQManager has no obvious pattern. However in the process of experiment, we observed that when UAs do not send RTCP packets, VQManager will set the one-way delay as zero, and therefore the value of quality will always be higher than the actual one.

Table 5: Errors Compared with DIST-V

	CaseA	CaseB	CaseC
RTP-M(using ping)	0.16%	0.13%	0.16%
VQMananger	-0.15%	-0.05%	0.17%
RTP-M(using OPTIONS)	-0.61%	-0.66%	-0.92%

We assumed that the *R* value of DIST-V is correct, and calculated the errors between DIST-V and others. As shown in Table 5, we can see that the error of VQManager is irregular, and the error of RTP-M (using ping) is smaller than RTP-M (using OPTIONS). It is because that there is only one NAT in the middle of UAs and RTP-M in our experiment environment, so error is smaller than the error of RTP-M (using OPTIONS) caused by processing time of SIP OPTIONS messages. In addition, the ping command is very likely to encounter the problem of no response. Many firewalls will filter ICMP requests. For example, in Windows 7 the default firewall is to ignore the ICMP request. Hence, our system will adopt using SIP OPTIONS method as a more reliable approach to get the assumed one-way delay.

5.4 Implement result

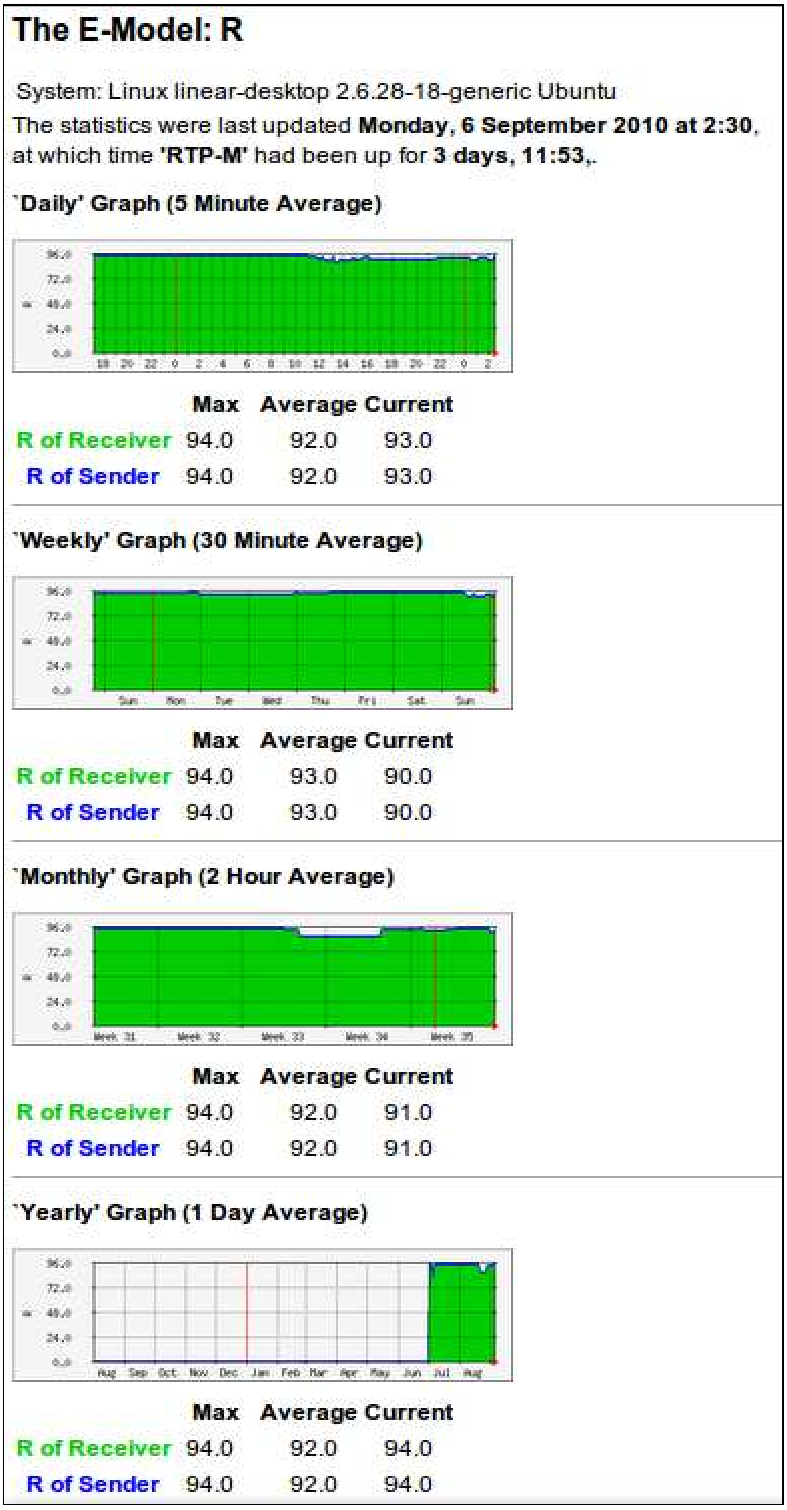

Figure 19: Implement Result

As shown in Figure 19, the graph generated by MRTG [12] has the x-axis of time and the y-axis of R. The graph will redraw every five minutes. In addition to daily graph, there are weekly graph, monthly graph, and yearly graph. Hence, we can observe the long-term variation of voice quality.

6. Conclusion and Future Work

In this thesis, we proposed a voice quality monitoring system RTP-M, and implemented it. RTP-M can monitor any VoIP device in Internet as long as it supports SIP OPTIONS message. Furthermore, in order to verify the feasibility of RTP-M, we use a program that controls the one-way delay, codec type, and packet loss rate of two UAs to make a comparison with RTP-M.

We also compared our RTP-M with VQManager which is a suit of paid software. This software provides a beautiful user interface, more functions, and it supports other voice transmission protocols, like H.323 [17] and Cisco Skinny Call Control Protocol (SCCP). However the quality measurement of this software has many restrictions, like the UAs have to enable the RTCP functionality, otherwise VQManager will set the value of one-way delay as zero; the communication time has to be long enough, and the UAs cannot be located behind the same NAT. Contrarily, our RTP-M has no such restrictions.

Finally, RTP-M will output the voice quality as a graph which is more intuitive for human beings. We hope that RTP-M can provide VoIP administrators with an objective reference for troubleshooting when users complain about the voice quality.

From the experiment result shown in Table 5, we can observe that there indeed exist errors when using RTP-M to predict the voice quality, and the error is caused by the extra time for parsing SIP OPTIONS requests and generating response. Hence, in the future an error model may be created and we may adjust the value of R according to some formula to make RTP-M a more precise approach, even though its error is now only 0.6%.

References

[1] ITU-T, "Method for subjective determination of transmission quality," ITU-T Recommendation P.800, August 1996.

[2] ITU-T, "Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs," ITU-T Recommendation P.862, February 2001.

[3] ITU-T, "The E-Model, a computational model for use in transmission planning," ITU-T Recommendation G.107, March 2005.

[4] R. G. Cole, J. H. Rosenbluth, "Voice over IP Performance Monitoring," ACM Computer Communication Review, vol. 31, no. 2, pp.9-24, April 2001.

[5] J. Rosenberg, H. Schulzrinne, G. Camarillo, A. Johnston, J. Peterson, R. Sparks, M. Handley, E. Schooler, "SIP: Session Initiation Protocol," IETF RFC 3261, June 2002.

[6] H. Schulzrinne, S. Casner, R. Frederick, V. Jacobson, "RTP: A Transport Protocol for Real-Time Applications," IETF RFC 3550, July 2003.

[7] H. Schulzrinne, S. Casner, "RTP Profile for Audio and Video Conferences with Minimal Control," IETF RFC 3551, August 2002.

[8] M. Handley, V. Jacobson, C. Perkins, "SDP: Session Description Protocol," IETF RFC 4566, July 2006.

[9] J. Rosenberg and H. Schulzrinne, "An Offer/Answer Model with the Session Description Protocol (SDP)," IETF RFC 3264, June 2002.

[10] P. Srisuresh, K. Egevang, "Traditional IP Network Address Translator (Traditional NAT)," IETF RFC 3022, January 2001.

[11] The RTP Proxy Server [http://www.rtpproxy.org/]

[12] Multi Router Traffic Grapher

[http://oss.oetiker.ch/mrtg/]

[13] 蔡政霖, “具備分散式特色之網路電話加壓測試,” 國立暨南國際大學通訊工程所, June 2008.

[14] David L. Mills, “Network Time Protocol (Version 3) Specification, Implementation and Analysis,” IETF RFC 1305, March 1992.

[15] VQManager

[http://www.manageengine.com/products/vqmanager/index.html]

[16] H. Schulzrinne, S. Casner, R. Frederick, V. Jacobson, “RTP: A Transport Protocol for Real-Time Applications,” IETF RFC 3550, July 2003.

[17] ITU-T, “Packet-based multimedia communications systems,” ITU-T Recommendation H.323, December 2009.

[18] Daniel Collins, “Carrier Grade Voice over IP, 2nd Ed.,” McGraw-Hill, September 2002.

[19] ITU-T, “Single-ended method for objective speech quality assessment in narrow-band telephony applications,” ITU-T Recommendation P.563, May 2004.

[20] Mohammad Abareghi, M. Mehdi Homayounpour, Mehdi Dehghan, Anahita Davoodi, “Improved ITU-P.563 Non-Intrusive Speech Quality Assessment Method For Covering VoIP Conditions,” International Conference on Advanced Communication Technology, pp.354-357, February 2008.

[21] Floriano De Rango, Mauro Tropea, Peppino Fazio, Salvatore Marano, “Overview on VoIP: Subjective and Objective Measurement Methods,” International Journal of Computer Science and Network Security, vol. 6, no. 1B, pp.140-153, January 2006.