A Real Time Voice Transmission Method for Voice Privacy between CDMA Mobile and PSTN Terminal

Yong-Seok Park[1], Kyoung-Seob Kim[2], Jaehwan Ahn[1],
Choon-Soo Kim[1] and Jae-Cheol Ryou [2]
[1]National Security Research Institute
E-mail : {parkys, jaehwan, jbr}@etri.re.kr
[2]Internet Intrusion Response Technology Research Center,
Chungnam National University
E-mail : kikysu@empal.com, jcryou@home.cnu.ac.kr

Abstract

To realize the voice privacy between CDMA mobile phone and PSTN terminal, the voice frames shall be transmitted transparently between the heterogeneous networks. For satisfying this requirement, we propose the method which transmits voice frames by the CDMA circuit data channel in real time.

In this paper we analyze the causes of voice delay which occurs during voice transmission by circuit data channel, propose methods which overcome the voice transmission delay and prove proposed methods by the experiment.

1. Introduction

The CDMA mobile phone service attracts a lot of interests of the world with the help of many advantages. The commercial service was being provided successfully and made many technical advances.

The following Figure 1 shows the current voice communication method between the CDMA mobile phone and the PSTN terminal.

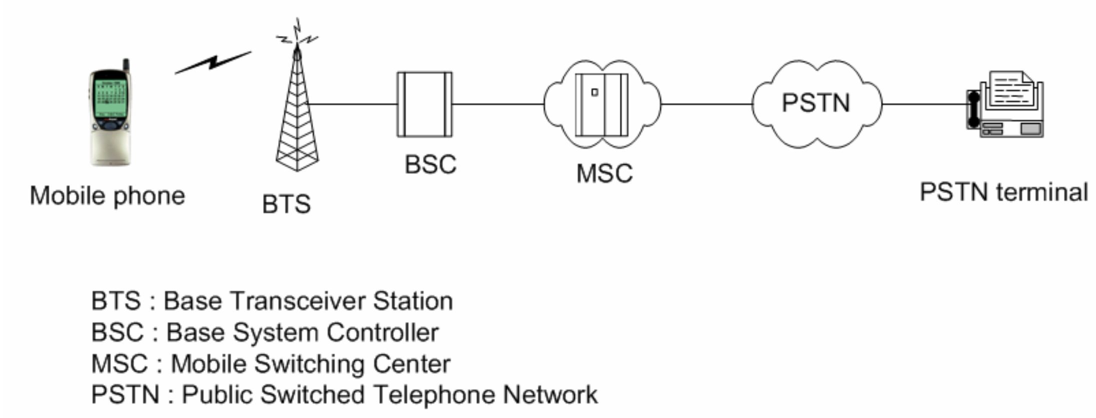


Figure 1. General voice communication between CDMA mobile and PSTN

As the CDMA mobile's voice output (Cellular : EVRC, PCS : QCELP) has the procedure of the voice signal conversion at the BSC for the connection to the PSTN like the Figure 2, the end-to-end voice privacy between the CDMA mobile and the PSTN is impossible with the current voice service architecture.

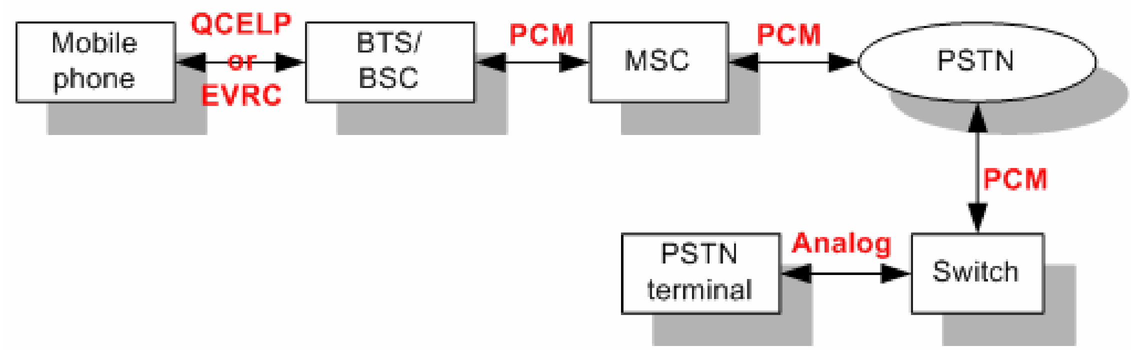


Figure 2. The signal conversion between CDMA mobile and PSTN

So this paper proposes the voice transmission method through the CDMA circuit data channel which supports the vocoder bypass for the end-to-end voice privacy between the mobile phone and PSTN modem.

2. The service analysis of the CDMA wireless data

2.1. The general aspects of the wireless data

The CDMA wireless data services are classified as the packet data service and the circuit data service. The circuit data service means that the data transmission is processed with the communication line in which the call setup has been made. The packet data service means that the split packet data are transmitted with the different routes. In this paper, the circuit data service is used for the voice traffic transmission between the mobile and PSTN.

The CDMA wireless data services use the international standard TIA/EIA/IS-707. The corresponding protocol should be implemented between the IWF(Inter Working Function) which process the wireless data protocol in the network system and the mobile phone.

2.1.1. The characteristics of voice and data. Because a data is very different from a voice, the methods of transmitting and processing of these have many differences. In the case of voice, the delay is not permitted but a little error can be ignorable. On the contrary, the delay is permitted but an error should be minimized in the data signal. So for the error reduction it is required the retransmitting scheme. Table 1 shows the differences between voice signal and wireless data.

Table 1. The differences between voice and wireless data

	Voice	Wireless data
Store process	No Store	Store and forward
Error correction	FEC (Forward Error Correction)	ARQ (Automatic Repeat Request)
FER(Frame Error Rate)	10^{-2}	10^{-6}

2.1.2. Vocoder bypass. In the case of M2L(mobile to Land) voice communication, voice transcoding is performed in BSC vocoder. Because of the transcoding, the end-to-end voice privacy between the mobile and PSTN is impossible with the current CDMA voice service. So, the vocoder bypass should be supported to make the end-to-end voice privacy possible. Current commercial mobile network supports the vocoder bypass for the case of data service. In the case of data service between mobile phone and PSTN modem, vocoder bypass is carried out under the service negotiation of network and mobile. (See the Figure 3).

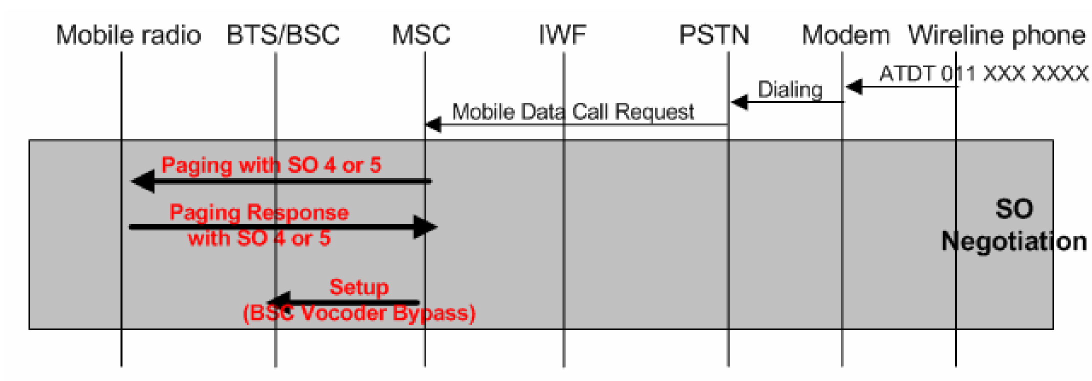


Figure 3. BSC vocoder bypass by service option

2.1.3. Service option(SO). SO is a very important parameter which defines the characteristics of media in CDMA mobile communication. It is used as the indicator parameter of voice and data, and is used to decide the vocoder bypass. The typical service option to provide CDMA circuit data service is described in Table 2.

Table 2. SO for CDMA circuit data service

SO	Services
4	Asynchronous data service (9.6 kbps)
5	Group-3 Fax (9.6 kbps)
12	Asynchronous data service (14.4 kbps)
13	Group-3 Fax (14.4 kbps)

2.2. Circuit data service

In this paper, SK Telecom's network is used as test network. Figure 4 shows the circuit data network and protocol stack of SKT. In this service, the IWF has a modem and controls it. It is interoperated with the modem in the PSTN network. The proprietary protocol is used between mobile phone and IWF.

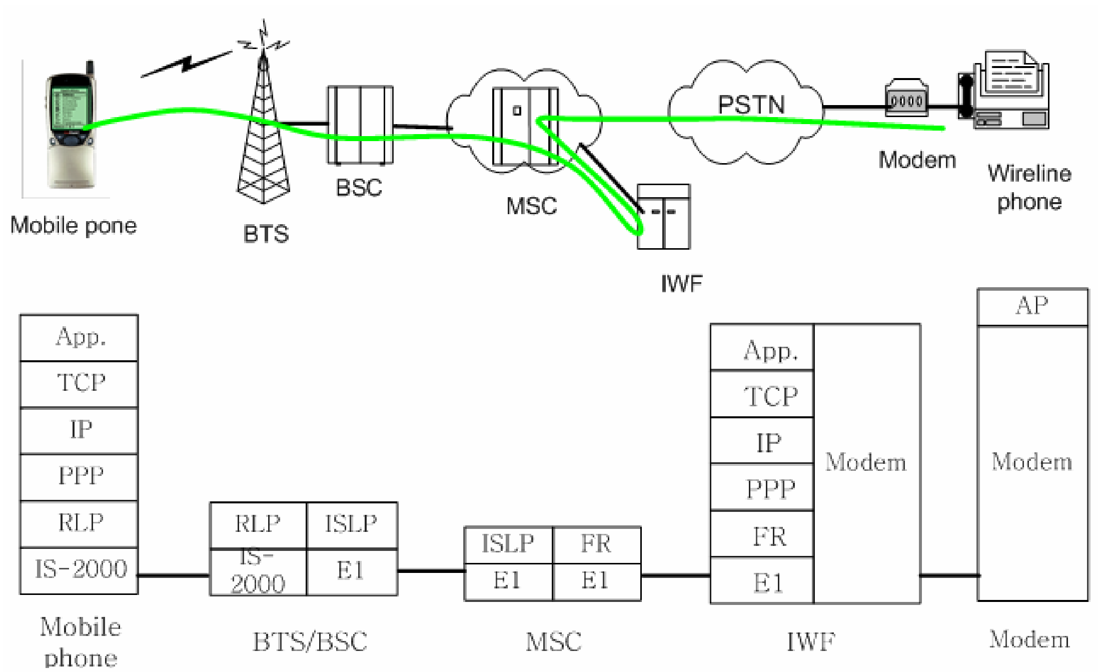


Figure 4. The SKT circuit data service network

TCP(Transmission Control Protocol) corresponds to transmission layer, IP(Internet Protocol) to network layer, PPP(Point to Point Protocol) to link layer, and IS-2000 and RLP(Radio Link Protocol) to relay layer. Above all protocols should be implemented in mobile phone and IWF.

RLP is used between mobile phone and base station. It maintains the data signal quality in poor wireless region by requiring retransmission when an error happens. Because voice service requires real time transmission, it is not necessary the data integrity by such as retransmission scheme. But in the case of data service, data integrity is important. So error free data should be transmitted by retransmission scheme in RLP. In other words, the main role of RLP is to reduce the retransmission rate in TCP by reducing the error occurrence in wireless region. Generally FER in wireless region is 10^{-2}. The FER reduces to 10^{-6} when three times retransmission in RLP is used. The data integrity is assured by the RLP, but the total data transmission rate decreases when the retransmission happens frequently by the error in the wireless region. This causes the considerable delay when the voice is transferred in the data channel.

TCP/IP/PPP layers are conformed between mobile phone and IWF. This is not for the internet connection, but for the assurance of data transmission credibility between mobile phone and IWF.

3. Proposed method of Mobile to Land real time voice transmission

3.1. The analysis of voice delay in the commercial circuit data network

The Figure 5 shows the delay time when the Mobile to Land voice transmission is carried out in the commercial SKT circuit data service network.

T1 is the time to gather voice frame in the mobile's application layer. The voice frame is outputted from mobile's vocoder every 20 ms. As the Table 3 shows, the voice frame length varies according to vocoder rate.

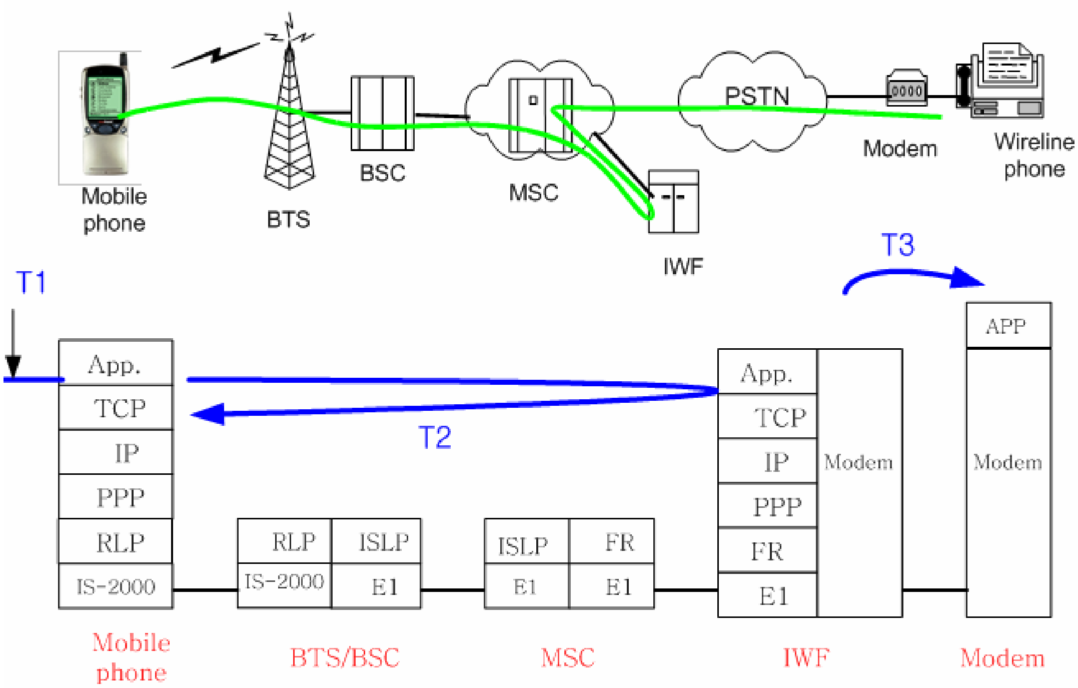


Figure 5. The delay times in Mobile to Land voice transmission

Table 3. The voice frame length (8K EVRC)

Vocoder rate	Outputted voice frame (bytes)
1	20
1/2	10
1/4	5
1/8	2

Actually, two bytes header is added to the every voice frame in mobile's application layer.

In this proposed method, several voice frames should be gathered in the mobile's application layer and then sent to TCP layer. The reason is as following.

The TCP/IP/PPP protocol headers are added to the voice frame generated in the application layer through each layer. As the Equation 1 shows, the total frame headers are 48 bytes to send one voice frame.

Block size of one frame
= one voice frame (22 byte : when vocoder rate is 1)
+ TCP header (20 bytes)
+ IP header (20 bytes)
+ PPP header (8 bytes)
= 70 bytes (1)

So to transmit 70 bytes voice block in 20 ms, the transmission rate of 28 kbps is necessary, but the rate which RLP provides is 14.4 kbps. So the data transmission through RLP is not possible. To transmit data through 14.4 kbps RLP, several voice frames should be gathered to minimize the TCP/IP/PPP protocol header burden.

$$\frac{48bytes + (22bytes + \#of\ frames)}{(\#of\ frames \times 20ms)} < 14.4kbps \quad (2)$$

By the Equation 2, the minimum number of frames to be gathered in the mobile's application layer to transmit frames with 14.4 kbps RLP is 4 frames. Consequently, the time to gather n voice frames in mobile's application layer is calculated as Equation 3.

$$T1 = n \times 20ms (n \geq 4) \quad (3)$$

T2 is called RTT (Round Trip Time) since it is the time duration from sending the data in the mobile phone's TCP to receiving the acknowledgement of IWF's TCP. If T2 is known, the transmitting time from the mobile phone to IWF can be calculated to be T2/2. RTT can be calculated in the TCP protocol software in the mobile phone. TCP uses it as the reference value for the timeout of the retransmission. Since the TCP retransmission of the mobile phone is done when the ACK from the IWF is not received within RTT, the total voice delay increases radically. The RTT value varies according to the network condition, it is not fixed. Especially it increases when the retransmission happens frequently because of the error rate increase in the wireless region. The main reason of the voice delay is the RTT in the Mobile to Land voice transmission, so it is necessary to find out the method to remove the delay caused by the RTT increase.

T3 is the time duration to transmit the data from the IWF modem to PSTN modem. Though the value of T3 varies somewhat according to the PSTN status, it is assumed to be constant in this paper. So, the total delay is like Equation 4 when the voice data is transmitted using the circuit data service in the Mobile to Land without the commercial network modification.

$$Total\ delay \cong T1 + T2/2 + T3 \quad (4)$$

3.2. The proposed method of real time voice transmission

As the Equation 4 show, the delay is inevitable when the voice is transmitted using the circuit data service in Mobile to Land. Although T1 and T3 are constant, T2 increases when the retransmission happens frequently by the error rate increase in the wireless region. In this case, the voice delay makes it impossible to use in the real time voice communication. In this clause the method to decrease the voice delay is proposed without the commercial network modification.

3.2.1. TCP control flag technique for improving the processing delay in TCP. As explained previously, the voice frames gathered in the mobile application layer during T1 are sent to TCP, the TCP header is added and is sent to IP. But a monitoring result shows the phenomenon that the frames received to TCP are not delivered to IP immediately, but sent with the next frames. This is explained in Figure 6. In normal case, the block 1 gathered during T1 is sent to TCP at T1 time, and after the TCP processing it is sent to IP immediately at T1 time. The block 2 is sent to TCP at 2T1, but it is not sent to IP immediately, instead it is sent with block 3 at 3T1. In other words, the additional time delay of T1 happens after the block 2. If the processing of sending i blocks to IP is not immediately done, additional i x T1 time delay happens. To solve this problem, the delay diminishing method using the control flags which are supported in TCP is proposed.

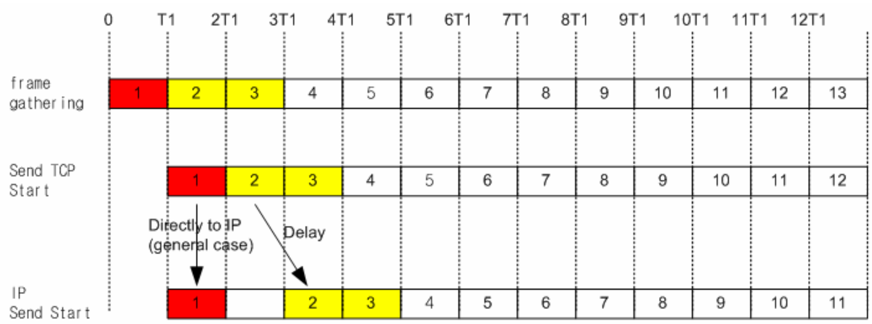


Figure 6. The reason of additional delay by the TCP processing delay

The following TCP control flags are described in the TCP protocol defined in IEEE RFC 793 standard.

```
struct /* control flags */
{
        Unsigned int force:1;
        Unsigned int clone:1;
        Unsigned int retran:1;
        Unsigned int active:1;
        Unsigned int synack:1;
        Unsigned int rtt_run:1;
        Unsigned int congest:1;
} flags;
```

The field "force" in the TCP control flags can be set to TRUE or FALSE. This value is the parameter to determine whether the data is sent to IP from TCP by compulsion. To set "force=TRUE" per voice block sent to TCP makes it solve the delay problem in TCP.

3.2.2. The block constitution algorithm adaptive to voice frame rate. As explained previously, the time T1 and T3 are fixed, T2 varies according to the network condition. Because the RTT is the main factor in the total end-to-end delay time, the optimization of the RTT is the key point. Actually, the RTT value increases in large when the retransmission happens frequently by the error increase in wireless region. Once the RTT values increases, the total end-to-end delay is fixed to this value after the maximum RTT value happens.

To solve these problems, the method not to retain the maximum delay is necessary. To do this, it is proposed the method to constitute the block adaptively to the vocoder output rate like the Figure 7 in the mobile’ application layer.

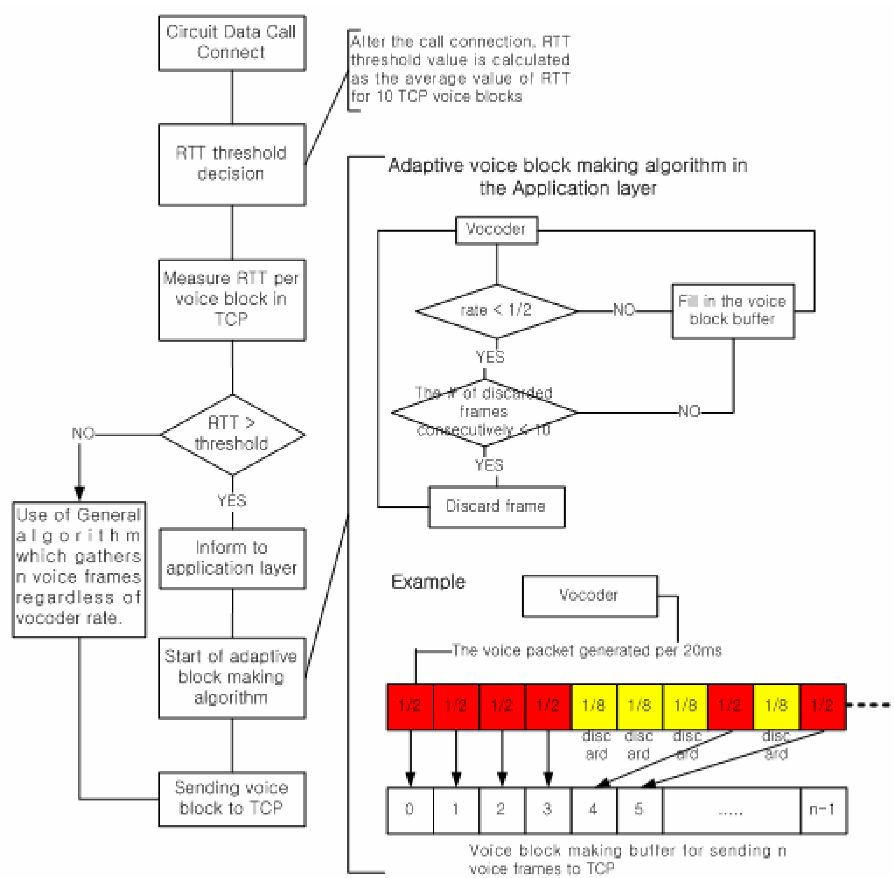


Figure 7. The generation procedure of voice frame which is adaptive to the vocoder rate

To compensate the delay occurred by RTT increase, it is used the technique of abandoning the data when the vocoder output rate is 1/8. The voice can be heard in human’s ear when the rate is over 1/2, so 1/8 rate frames cannot be heard. It does not affect the voice quality when 1/8 frames are not transmitted. In result the data which is not worth of information is not delivered to TCP, and is not transmitted. So the delay occurred by RTT increase is compensated. But when the adaptive voice block is constituted, the number of frames continuously abandoned should be checked because abandoning the frames under the 1/2 rate continuously 10 times makes it possible to be recognized in human’s ear.

4. Experiments and results

4.1. Test bed for Mobile to Land real time voice transmission

The proposed methods are proved by the experiment to connect the mobile to PSTN in SKT commercial circuit data service network.

In this paper we made the test bed to test the M2L real time voice transmission as Figure 8.

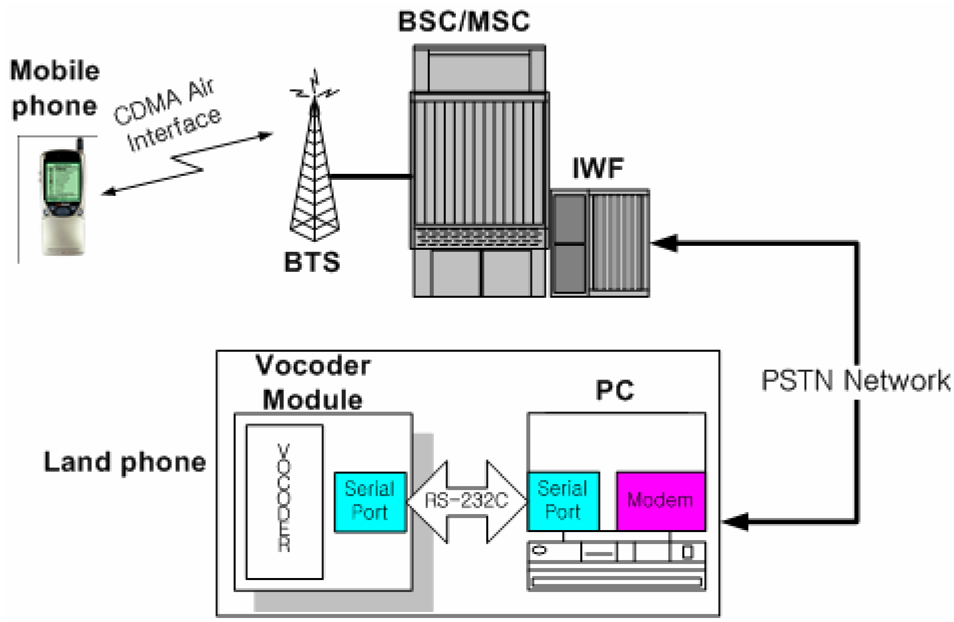


Figure 8. The test bed for the circuit data voice transmission in Mobile to Land

The mobile phone in Figure 8 is a SK telecom's CDMA phone. It has a software to control a circuit data call and protocol stack to process TCP/IP/PPP/RLP. The data service software of mobile phone which uses Qualcomm MSM5100 modem chip is modified to implement our proposed scheme.

The Land phone in Figure 8 consists of PC which includes a commercial PSTN modem and vocoder module which connected with the PC serial port. Incoming EVRC frame is received by PC modem, delivered to vocoder module through a serial port and decoded to voice. Outcoming voice is encoded to EVRC in vocoder module and transmitted to PSTN modem.

4.2. Experiment results

The EVRC vocoder of mobile phone does not generate 1/4 rate EVRC frame. Instead 1, 1/2 and 1/8 rate EVRC frames are generated. In this paper the maximum rate of vocoder output is set to 1/2 for the purpose of guaranteeing transmission margin by decreasing the voice information amount.

Previously it is said that the number of the vocoder output frames to be gathered in the mobile's application layer should be over 4. In this paper the number of frames to be gathered is set to 10. This number is set considering the buffer size used for the data service in mobile phone and the transmission rate of 14.4 kbps. The T1 to gather 10 voice frames in the mobile application layer is 200 ms (10 times of 20 ms). As the Figure 9 shows, the RTT(T2) measured in the TCP protocol software of the mobile phone is 700 ~ 800 ms.

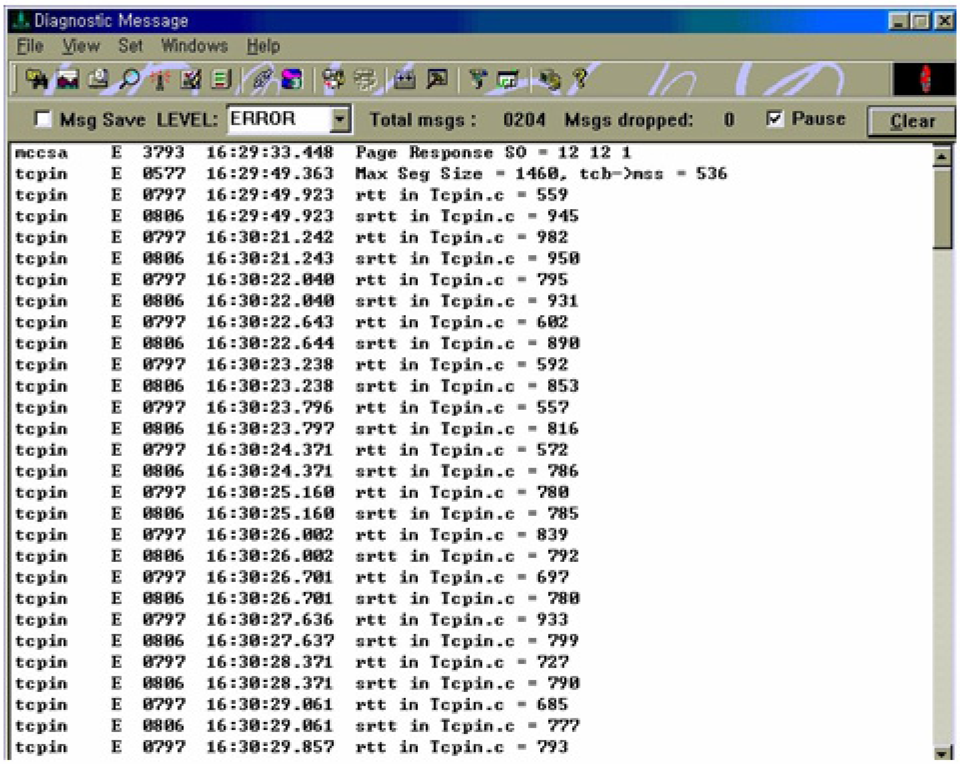


Figure 9. The RTT measured by mobile phone monitoring tool

So, when the voice is transmitted through the circuit data service, the estimated time of the voice delay is like the Equation 5.

$$\begin{aligned} &\textit{estimated delay} \cong T1(200ms) \\ &\quad + T2/2(350 \sim 400ms\ \textit{in general}) \\ &\quad + T3(\alpha) \\ &\quad \cong (550 \sim 600ms) + \alpha \end{aligned} \qquad (5)$$

(T3 : IWF Processing time + the transmission time from IWF modem to PSTN modem)

But the real voice delay in the SKT circuit data network is different from the estimated. In the beginning of the circuit data call, the delay of 800 ~ 900ms shows. In the middle of data call, the delay of 2 seconds is fixed. So, the normal voice communication is not possible. The reason of this phenomenon is the RTT increase caused by TCP retransmission. According to the our measurement using the mobile phone monitoring tool, as the retransmission rate caused by the error in wireless region increases, the delay increases to 1500 ~ 2500 ms. Once the RTT increases, total transmission delay is fixed to 2 seconds.

But, this voice delay decreases significantly by using our proposed scheme. Preventing transmission delay in TCP and using the voice block constitution algorithm adaptive to RTT variation makes it possible to overcome the delay problem and to transmit voice data in real time.

Figure 10 is a capture screen of the oscilloscope which measures the delay from the mobile phone's microphone input to the PSTN phone's earphone output. Because our proposed schemes are used to overcome the delay issue, the measured delay time is 500 ~ 600ms. It is a considerably decreased value, and we can know that a real time voice communication is possible.

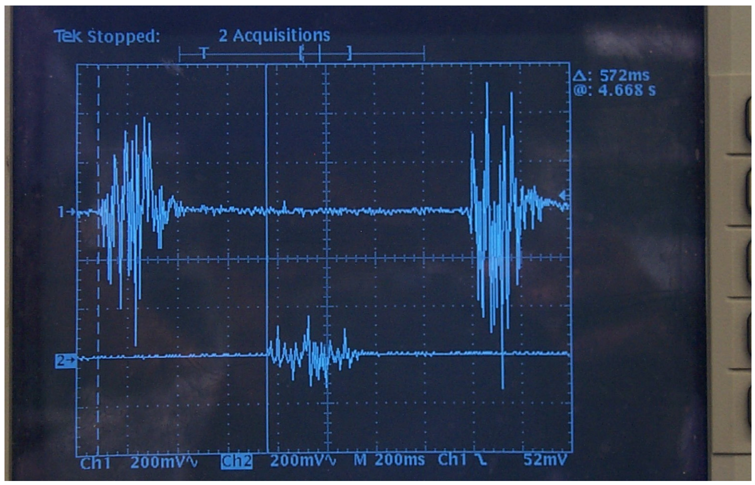


Figure 10. The measured delay between mobile phone and PSTN phone

5. Conclusion

To realize the voice privacy between CDMA mobile phone and PSTN terminal, the voice frames shall be transmitted transparently between the heterogeneous networks. So, in this paper we proposed the method which transmits voice frames by the CDMA circuit data channel to support the vocoder bypass in real time.

The reason for the voice delay was explained with the analysis of the circuit data service protocol, the technique to handle the TCP control flags was used for solving the transmission delay problem in TCP, and the delay caused by RTT increase was cleared with the help of the voice block constitution algorithm which varies to the vocoder rate in the mobile's application layer. The proposed method in this paper is able to provide the end-to-end voice privacy because the voice transmission is possible in the real time without the modification of the CDMA network.

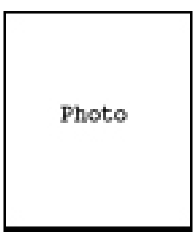
6. Acknowledgment

The fifth author of this research was supported by the MIC(Ministry of Information and Communication), Korea, under the ITRC(Information Technology Research Center) support program supervised by the IITA(Institute of Information Technology Assessment)" (IITA-2006-(C1090-0502-0020))

7. References

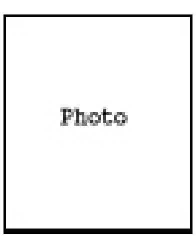
[1] Physical Layer Standard for cdma2000 Spread Spectrum Systems, 3GPP2 C.S0002-D v2.0, September 2005.
[2] Upper Layer(Layer 3) Signaling Standard for cdma2000 Spread Spectrum Systems, *3GPP2 C.S0005-D v2.0*, September 2005.
[3] (14.4 kbps) Data SOs for Spread Spectrum Systems - STU III Transparent + Non-Trans, *3GPP2 C.S0017-0*, December 1999.
[4] Data Service Option for Wideband Spread Spectrum System, *3GPP2 C.S0017-0 v5.0*, February 2003.
[5] Data Service Options for Spread Spectrum Systems: Introduction and Service Guide, *3GPP2 C.S0017-001-A v1.0*, July 2004.
[6] Data Service Options for Spread Spectrum Systems: AT Command Processing and the Rm Interface, *3GPP2 C.S0017-003-A v1.0*, July 2004.
[7] Data Service Options for Spread Spectrum Systems: Async Data and Fax Services, *3GPP2 C.S0017-004-A v1.0*, July 2004.
[8] Addendum for cdma2000 RLP and Additional Packet Data Support, *3GPP2 C.S0017-0-1*, December 1999.
[9] Data Service Option for Wideband Spread Spectrum System – Addendum2, *3GPP2 C.S0017-0-2 v2.0*, August 2000.
[10] Transmission Control Protocol. *IEEE RFC 793*.
[11] Internet Protocol , *IEEE RFC 791*.
[12] The Point-to-Point Protocol (PPP), *IEEE RFC 1661*.

Authors



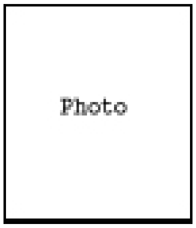
Yong-Seok Park

Received a B.S. degree in electronic engineering form Kyungpook National University, Korea, 1995 and M.S. degree in electronic engineering form Kyungpook National University, Korea, 1997. Now he is working as a senior member of engineering staff at NSRI(National Security Research Institute). His research interests are mobile communication security and DRM.



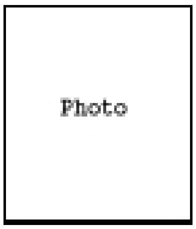
Kyoung-Seob Kim

Received a B.S. degree in electronic & telecommunication engineering from KwangWoon University, Korea, 1994 and M.S. degree in computer engineering from Korea University, Korea, 2004. Currently pursuing Doctor's degree in Computer Engineering from ChungNam National University, Korea.



Jaehwan Ahn

Received a B.S. degree in electronic engineering from Sung Kyun Kwan University, Korea, 1999 and M.S. degree in information and communication engineering from GIST(Gwangju Institute of Science and Technology), Korea, 2002. Now he is working as a researcher in NSRI(National Security Research Institute). His research interests are mobile system security and smart card security.



Choon-Soo Kim

Received the B.S. degree in electrical engineering from Soongsil University, Korea, 1987, M.S. degree in electrical engineering from Soongsil University, Korea, 1989 and Ph.D. degree in electrical engineering from Soongsil University, Korea, 1997. Now he is working as a principal member of engineering staff at NSRI(National Security Research Institute). His research interests are communication security protocol.

Jae-Cheol Ryou

Received a B.S. degree in industrial engineering from Hanyang University, Korea, 1985, M.S. degree in computer science from Iowa State University, USA, 1988 and Ph.D. degree in electrical engineering and computer science from Northwestern University, USA, 1990. He joined the faculty of Chungnam National University, Korea, in 1991. Now he is working as a director at the IIRTRC. His research interests are computer and communication security and electronic payment system.