Journal of Information Hiding and Multimedia Signal Processing ©2010 ISSN 2073-4212
Ubiquitous International Volume **1**, Number **3**, July **2010**

Reinforcement of VoIP Security with Multipath Routing and Secret Sharing Scheme

Ryouichi Nishimura, Shun-ichiro Abe, Norihiro Fujita and Yôiti Suzuki

Research Institute of Electrical Communication, Tohoku University
2-1-1 Katahira, Aoba-ku, Sendai 980-8577, Japan
{ryou,abe,yoh}@ais.riec.tohoku.ac.jp, norihiro.fujita@gmail.com

Nobuyuki Enomoto, Tsutomu Kitamura and Atsushi Iwata

System Platforms Research Labs, NEC Corporation
1753 Shimonumabe Nakahara-ku, Kawasaki, 211-8666, Japan
n-enomoto@db.jp.nec.com, t-kitamura@aj.jp.nec.com, a-iwata@ah.jp.nec.com

Received January 2010; revised May 2010

Abstract. *A technique to enhance the security of vocal communication over an open network is proposed in this paper. This technique combines a secret sharing scheme and a multipath routing technique on network communication. The secret sharing scheme, originally proposed to convey information securely from one person to another, divides original information into sets of partial data. Each set of partial data is designated as shared data. In principle, nobody can obtain any information of the original from a subset of the shared data. Only a person who collects all shared data can reconstruct the original information. Furthermore, a multipath routing technique, by which a single set of data is transferred from one host to another through multiple network paths, was developed originally for load sharing and high reliability. The proposed method therefore divides speech data using the secret sharing scheme and transfers the shared data using the multipath routing technique to realize secure voice communication over the network. Advantages of the method are that it requires no secret key, which plays an essential role in IPsec, and moreover, it can be used in conjunction with IPsec to reinforce the security of VoIP.*
Keywords: Secret sharing scheme, Multipath routing, VoIP, Speech compression coding

1. **Introduction.** There are many security concerns in VoIP, such as denial of service, eavesdropping, alteration of voice stream, toll fraud, redirection of call, accounting data manipulation, caller ID impersonation, and unwanted calls and messages [6]. From the view point of speech communication, confidentiality of conversations over telephones should be protected. In conventional public switched telephone networks (PSTN), entire communication paths were administered by a few authorized telephone companies. It was therefore difficult for a malicious person to wiretap conversations over telephones because persons who were allowed to access the network were carefully restricted. For the internet protocol telephone, or Voice over Internet Protocol (VoIP), which has recently shown rapid growth, multiple intermediates exist between the two endpoints (telephones). Accordingly, the risk of man-in-the-middle attack increases.

Current VoIP uses IPsec as a solution to the security problem [9]. This security protocol is strong and works well but has some drawbacks for the use in VoIP. Complicate settings needed on the both endpoints and the necessity of knowing of the settings each other

could prevent the dissemination of the Voice over IPsec (VoIPsec). Overhead added to the packet by the VoIPsec is larger compared to the typical size of packet for voice traffic, resulting in the wasted bandwidth [4]. Cryptography methods used in the IPsec, most of which are based on the key exchange, are vulnerable against the man-in-the-middle attack in the sense that the users have no chance to be aware that their conversations are in jeopardy.

There are other VoIP security methods that are not using IPsec. An example of such a method is that developed by combining secret sharing scheme and multipath routing [13]. Secret sharing scheme can reduce the risk of eavesdropping by imposing risks, which are different from those in the VoIPsec, on the malicious person conducting the man-in-the-middle attack. For collecting all the shared data, the malicious person needs to crack multiple routers in the different paths or to crack one of the routers closest to the endpoints. Otherwise he/she has to collude with others to make multiple routers under control. These requirements would improve chances of detecting this illicit attempt. Multipath routing is, on the other hand, a technique by which packets are delivered through disjoint multiple paths from a source node to a destination node; in contrast, in the current network, all the packets are usually delivered via a single path. Multipath routing has potential for achieving bandwidth widening, load balance, alleviation of congestion, increase in reliability, and so on. An efficient way of finding disjoint paths was presented in [22], and its efficiency was demonstrated by comparing it with a conventional method. However, the method [13] will inevitably increase the total amount of data to be delivered due to the additional redundancy provided for security. This does not affect systems in which small amount of data are transferred, as in the case of secret keys [13]. However, in VoIP applications, massive data are to be transferred. Hence, this method is not effective when applied directly to VoIP applications.

This paper proposes a method, in which the VoIP security is achieved by using secret sharing scheme in conjunction with multipath routing. Secret sharing scheme and multipath routing are both likely to increase the data amount to be transferred. The remarkable feature of the proposed method is that it can suppress the data increase while achieving its purpose, which is confidentiality of conversations. This paper is organized as follows. Problems related to VoIP communication are discussed in Section 1. Section 2 provides an outline of the proposed method and an introduction of the secret sharing scheme and multipath routing technique, both of which are the bases of the proposed method. The problem encountered when the secret sharing scheme is applied directly to audio signals is discussed in Section 3. Feasibility of the proposed method in terms of time delay is tested using a real network in Section 4. Discussion and concluding remarks are provided in Section 5 and 6, respectively.

2. **Audio Secret Sharing and its application to VoIP.**

2.1. **Outline of the proposed method.**

The proposed method is realized as a combination of a multipath routing technique [2, 23] and a secret sharing scheme [14, 21]. The secret sharing scheme has different cryptographic characteristics from those used in the IPsec:

- It requires neither key exchange nor scalable table sharing, and yet it retains the capability of sufficient security.
- It should use multiple paths to deliver an encrypted message.
- It has few calculations and has the characteristic that any smaller set of the data than the whole does not provide useful information for reconstructing the original (The latter part is common to the IPsec).

These characteristics enhance its robustness against the man-in-the-middle attack. Regarding communication along multiple network paths, a multipath routing technique is available in which a set of multipath routers selects multiple paths to exchange packets that are relevant to a single session between them.

Consequently, the proposed method is outlined schematically as presented in Fig. 1. Using the proposed method, a malicious person must compromise at least one router for each of all the multiple paths to eavesdrop on the conversation. It is noteworthy that the method is available in conjunction with IPsec to reinforce the security of the current VoIP communication systems because an eavesdropper must obtain the secret keys of the sender and recipient and simultaneously hack multiple routers on different networks. In addition, when different secret keys are assigned to each IPsec communication corresponding to each path, wiretapping the conversation becomes even more difficult because the eavesdropper has to obtain those multiple secret keys.

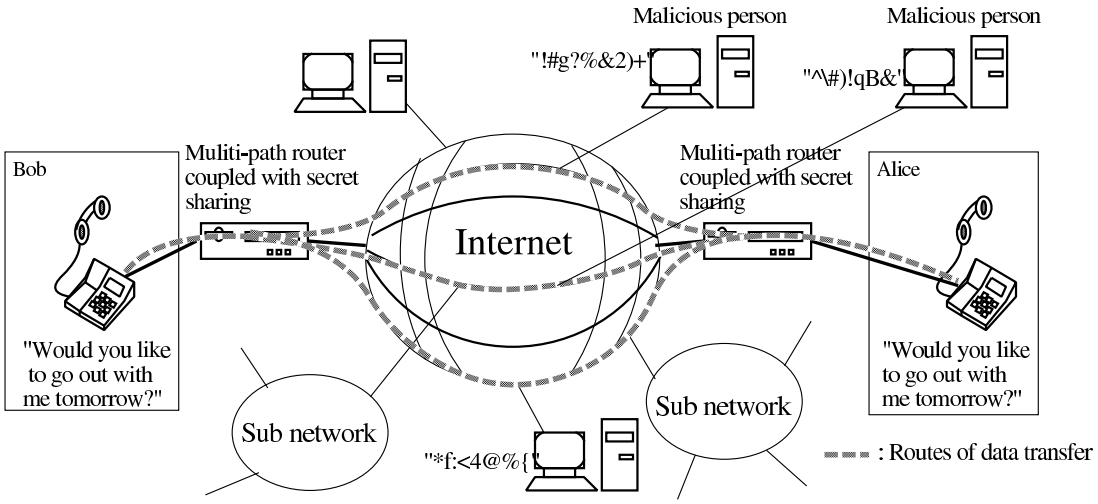
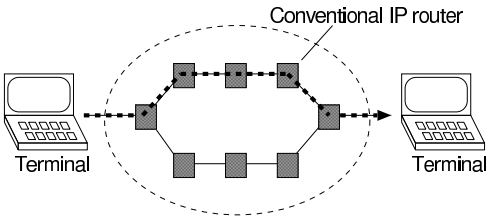


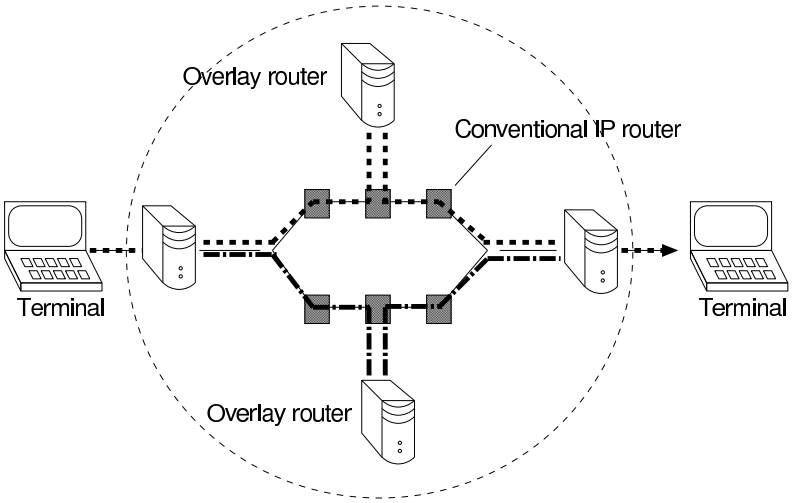
FIGURE 1. Proposed secure VoIP based on a secret sharing scheme and multipath routing

2.2. **Multipath routing.** IP networks, such as the Internet or IP-VPN, use hop-by-hop routing, in which routers individually determine the next router to send packets to. Therefore, only a single path is usually constructed to communicate between a pair of terminals, as depicted in Fig. 2(a). Meanwhile, it is also possible to use more than two paths for communication between a pair of terminals using an overlay routing technique [1], as depicted in Fig. 2(b). The overlay routing technique is implemented in layers higher than the fourth layer, i.e. a transport layer, to control the route. Therefore, by installing some overlay routers in geometrically distant places and integrating information related to IP routing of the network and deployment of the routers, disjoint routing can be realized [3]. Several studies are carried out on finding disjoint paths [11, 15, 22], and in the disjoint routing, packets relevant to a single session are transferred to the distance via the most disjoint routes. If the hosts of the server and the receiver have multiple network interfaces using which they connect to multiple Internet service providers (ISPs) simultaneously, then multihoming could enable end-to-end disjoint multipath routing [16], as shown in Fig. 2(c).

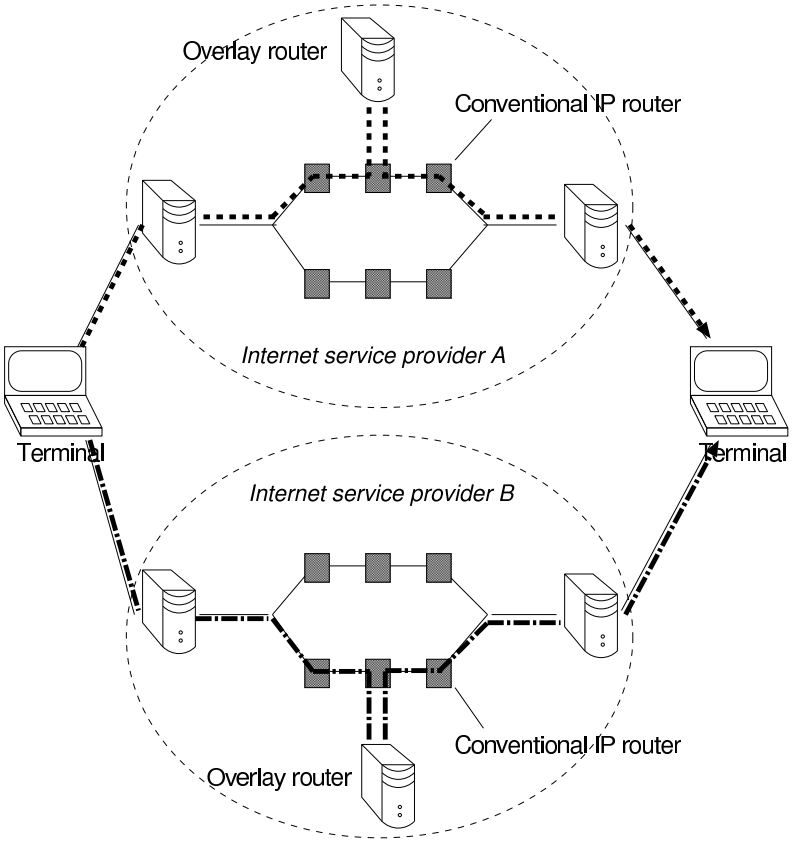
2.3. **Secret sharing scheme.**



(a) Single path routing (conventional IP network)



(b) Multipath routing



(c) Multipath routing though multiple networks

FIGURE 2. Packet communication over the network

2.3.1. *Shamir's secret sharing scheme.* A secret sharing scheme was proposed independently by Shamir [21] and Blakley [5] in the late 1970s. Shamir's scheme was based on polynomial interpolation, while Blakley's was based on geometrical projection. In this section, Shamir's scheme is reviewed.

Given k points in the two-dimensional plane with distinct x_is, one and only one polynomial $q(x)$ of degree $k-1$ exists such that $q(x_i) = y_i$ for all i. Therefore, to divide data D, which is a number, into n numbers $D_i (i = 1, \ldots, n)$, we select a random $k-1$ degree polynomial $q(x) = a_0 + a_1 x_1 + \ldots + a_{k-1} x_{k-1}$ in which $a_0 = D$, and evaluate it as

$$D_1 = q(1), D_2 = q(2), \ldots, D_n = q(n). \tag{1}$$

Given any subset of k of these values and their indices, we can derive coefficients of $q(x)$ through interpolation and evaluate $D = q(0)$. Otherwise, no chance to calculate D exists. This secret sharing scheme has a characteristic that the number of necessary data amongst n data to reconstruct the original is greater than k. We represent this characteristic as the (k, n) threshold scheme hereinafter, as used in the original paper by Shamir.

The Vernam cipher, on which the exclusive-OR operation plays an important role [24], can provide a characteristic similar to that in Shamir's secret sharing and is therefore applicable in the proposed method.

2.3.2. *Visual secret sharing.* The proposed method can employ the visual secret sharing scheme [14] as well. This scheme is suitable for our method when one would like to encrypt a message bit by bit. In addition, an efficient means exists to generate shared data for a pair of small k and n. For example, a (2, n) threshold scheme is solvable using the following collections of $n \times n$ matrices:

$$C_0 = \left\{ \begin{array}{c} \text{all matrices} \\ \text{obtained by permuting} \\ \text{the columns of} \end{array} \begin{bmatrix} 1 & 0 & 0 \ldots 0 \\ 1 & 0 & 0 \ldots 0 \\ \vdots & & \vdots \\ 1 & 0 & 0 \ldots 0 \end{bmatrix} \right\} \tag{2}$$

$$C_1 = \left\{ \begin{array}{c} \text{all the matrices} \\ \text{obtained by permuting} \\ \text{the columns of} \end{array} \begin{bmatrix} 1 & 0 & 0 \ldots 0 \\ 0 & 1 & 0 \ldots 0 \\ \vdots & & \vdots \\ 0 & 0 & 0 \ldots 1 \end{bmatrix} \right\}, \tag{3}$$

where digits of '0' and '1' respectively, correspond to white and black pixels. Every row vector of the matrices can be a piece of shared data. A logical sum of any two rows of C_0 yields a vector containing only one digit of '1'. On the other hand, when choosing any two rows from C_1, their logical sum yields a vector containing two digits of '1'. As presented in Eqs. (2) and (3), every piece of shared data of either C_0 or C_1 contains one digit of '1' that appears at an arbitrary column. Therefore, it is impossible to tell which of C_0 or C_1 a piece of shared data belongs to.

An example of the (2, 2) threshold scheme is depicted schematically in Fig. 3, where a logo image is divided into two salt-and-pepper noise images. It is impossible to obtain any information about the original image from any shared image, but the original image readily appears when both shared images are superimposed and mutually aligned.

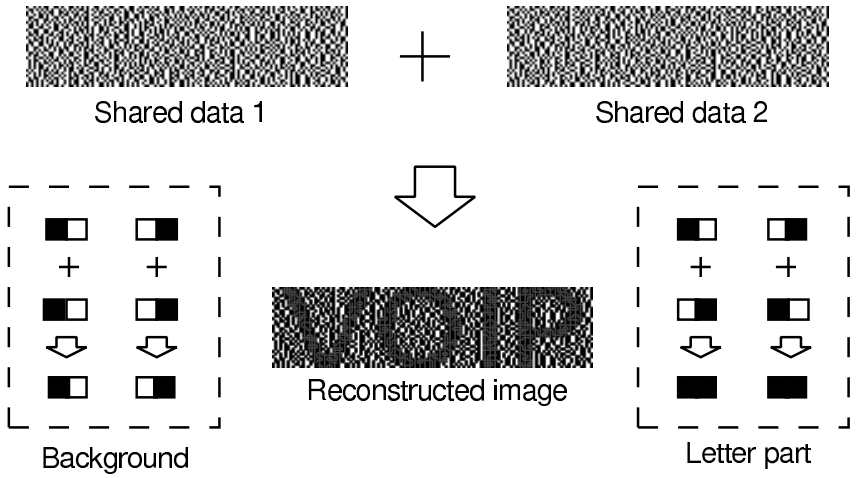


FIGURE 3. Example of visual secret sharing in which an image is divided into two pieces of shared data

Similarly, a (3, 3) threshold scheme is solved using the following scheme:

$$C_0 = \left\{ \begin{array}{c} \text{all matrices} \\ \text{obtained by permuting} \\ \text{the columns of} \end{array} \begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix} \right\} \tag{4}$$

$$C_1 = \left\{ \begin{array}{c} \text{all matrices} \\ \text{obtained by permuting} \\ \text{the columns of} \end{array} \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix} \right\}. \tag{5}$$

3. **Data size reduction.**

3.1. **Speech compression coding with secret sharing scheme.** Adopting a secret sharing scheme increases the network traffic. This data increase might not matter so much for current wideband networks. However, from a practical perspective, it presents a burden to users who must pay a fee to Internet service providers to use multiple networks. Some methods, which were originally proposed for image data, attempted to reduce the data size by applying secret sharing scheme only to significant portions [12, 8]. This approach is promising also for audio signals. Speech compression coding is a technique to reduce the amount of data necessary to represent a speech signal. Therefore, it is apparent that applying a secret sharing scheme to the output of speech compression coding can reduce the amount of data to be transferred in the proposed method. Furthermore, because speech compression coding reduces the verbal redundancy of speech, applying a secret sharing scheme to a small number of coded bits must worsen speech intelligibility.

The threat of eavesdropping becomes greater in multi-hop wireless links, where unspecified hosts inherently play the role of a router. Gibson *et al.* proposed an encryption method that emphasizes scalable speech coding [7]. Scalable coding is a method by which a signal is represented as superposition of the core layer and several supplemental layers. The core layer contains most principal parts of the contents; the supplemental layers contain information to refine the representation of the contents. Therefore, encrypting the core layer alone is efficient to ensure a high level of protection against eavesdroppers and to reduce the network traffic load, thereby lowering power consumption of computers constructing multi-hop wireless links. Although this method is promising for scalable coding, speech compression coding techniques currently used in VoIP are mainly ADPCM [28] and CS-ACELP [30, 18] as well as PCM. We therefore investigated ADPCM and

Table 1. Experimental conditions

Listeners	Four male graduate students in their twenties
Speech signals	80 excerpts from the set of 1000 syllabic-balanced sentences [17]
Repetition	20
Sound level (L_{Aeq})	60 dB

CS-ACELP for use in the proposed method. In addition to these codecs, LD-CELP [29] was investigated because of its simple encoding and decoding structures.

3.2. **ADPCM.** As described in ITU-T G.726 [28], ADPCM is the speech compression coding method used for 40, 32, 24, and 16 kbps. Among these bit rates, 32 kbps of ADPCM, in which every sample with a resolution of 16 bits is transformed into 4 bits, is commonly used. Therefore, we specifically addressed the ADPCM of 32 kbps and investigated which of the four bits is most effective to protect speech information.

This investigation was carried out through listening tests in which test stimuli were distorted by randomizing one bit of each four-bit ADPCM code. The experimental conditions are portrayed in Table 1. Four male graduate students in their twenties with normal hearing ability participated in the experiments. Regarding source signals, 80 sentences excerpted from the set of 1000 Japanese syllabic-balanced sentences, which had a sampling rate of 44.1 kHz and a resolution of 16 bits, were used. These speech signals were converted into 64-kbps PCM to be fed to an ADPCM encoder and further processed to produce the test stimuli. Listeners heard test stimuli diotically via headphones in a soundproof room. They were instructed to write down the sentences as they were able to hear them. Moreover, they were asked to write down speculated sentences when the test stimuli were so distorted by the bit-randomization operation that they became extremely difficult to understand. Listeners were allowed to listen to the test signals as many times as they liked.

In the experiments, intelligibility was defined as the correct answer rate of independent words, i.e., nouns, verbs, adjectives, and adverbs, which appeared in the tested sentences. The results are shown in Fig. 4. It is apparent from Fig. 4 that encrypting only the first bit, which corresponds to the sign bit, of each four-bit ADPCM code is sufficient to reduce speech intelligibility.

3.3. **LD-CELP.** As described in ITU-T G.728 [29], LD-CELP is speech compression coding for 16 kbps. Block diagrams of the LD-CELP encoder and decoder are shown in Figs. 5 and 6, respectively. Input is assumed to be in the PCM format or in a signal coded with 64 kbps of A-law: a sampling rate of 8 kHz and resolution of 8 bits. The LD-CELP encoder outputs a 10-bit code for every five samples of the input signal. Consequently, the bitrate of the output is reduced to one-fourth of the input signal, resulting in 16 kbps. The output code consists of three gain bits and seven vector bits. The vector bits show an index to indicate a proper vector from 128 candidates stored in the VQ codebook. The gain bits show an index to determine the gain for the selected vector. Consequently, the block diagram of the encoder of the proposed method can be depicted as Fig. 7.

Listening tests were carried out to investigate how many bits must be secret to reduce speech intelligibility sufficiently. Six male students in their twenties with normal hearing ability voluntarily participated in the tests. The bits to be protected were chosen from the MSB of each gain and vector bits, as shown in Fig. 8. The chosen bits were replaced with random binary digits. The experimental conditions are the same as those for ADPCM,

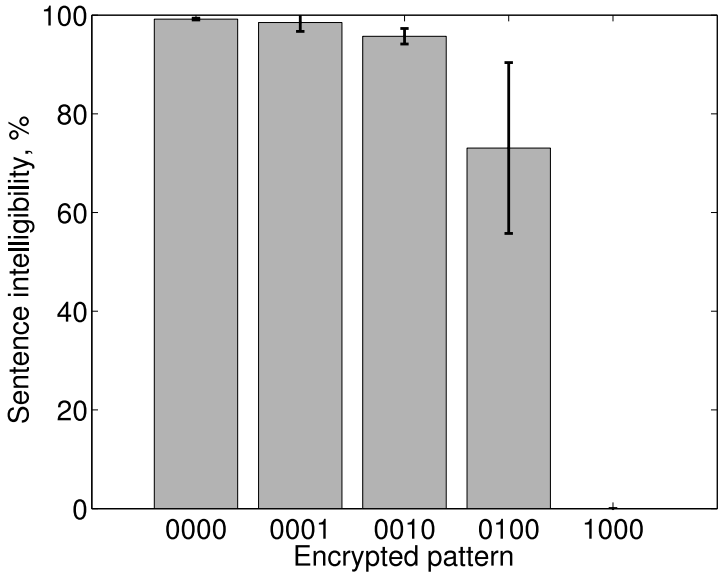


FIGURE 4. Intelligibility of ADPCM with bit-randomization operation. The four digits beneath each bar indicate which of the four bits was subject to the bit-randomization operation.

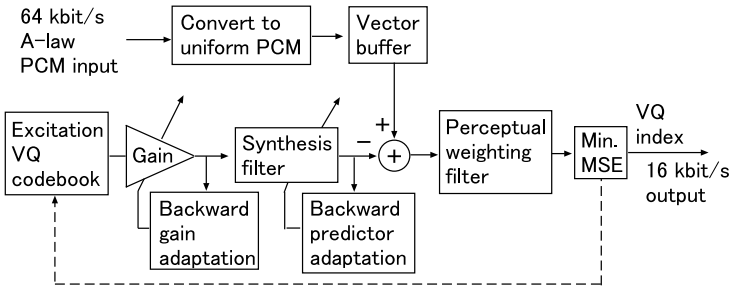


FIGURE 5. Block diagram of LD-CELP encoder

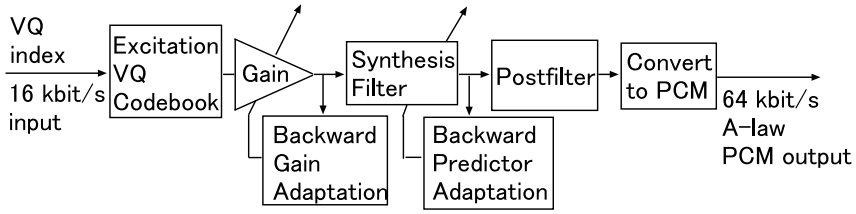


FIGURE 6. Block diagram of LD-CELP decoder

described in the previous section, except that six male graduate students participated in the experiments and that 320 sentences excerpted from the set of 1000 Japanese syllabic-balanced sentences were used.

Table 2 shows results of the experiments. Columns show how many bits in the vector bits were replaced with random digits. Rows, on the other hand, show how many bits in the gain bits were replaced with random digits. To clarify the difference, cases for $s_g = 0$ and $s_v = 0$ are depicted in Fig. 9 with error bars depicting the standard deviation. As presented in Table 2 and Fig. 9, applying secret sharing to three appropriate bits is sufficient to reduce speech intelligibility effectively. Moreover, the vector bits are more

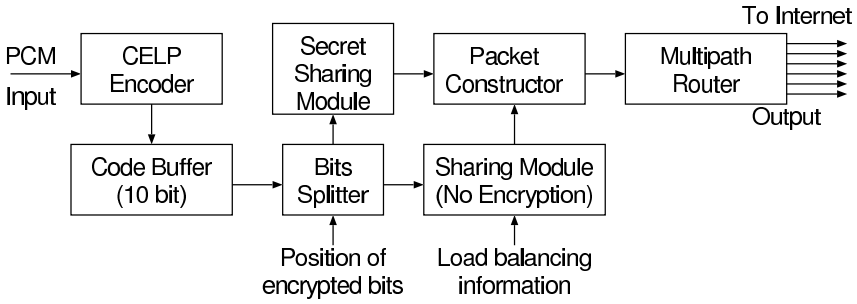


FIGURE 7. Block diagram of the proposed speech encoder containing blocks for secret sharing and multipath routing

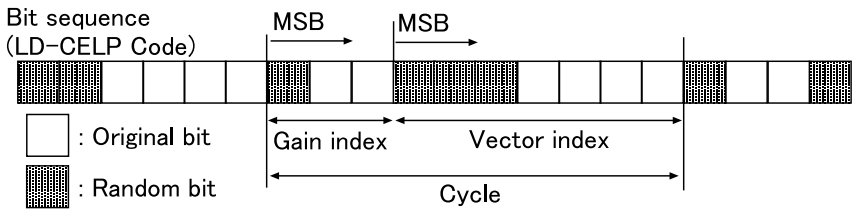


FIGURE 8. Application of bit randomization to an LD-CELP code

TABLE 2. Intelligibility of coded speech signals when some bits of the vector index (columns) and gain index (rows) are replaced with random digits.

$(s_g,\ s_v)$	0	1	2	3	4	5	6	7
0	99.1	63.6	4.7	0.0	0.0	0.0	0.0	0.0
1	95.9	21.3	2.1	0.0	0.0	0.0	0.0	0.0
2	38.9	9.5	0.0	0.0	0.0	0.0	0.0	0.0
3	27.8	4.2	0.0	0.0	0.0	0.0	0.0	0.0

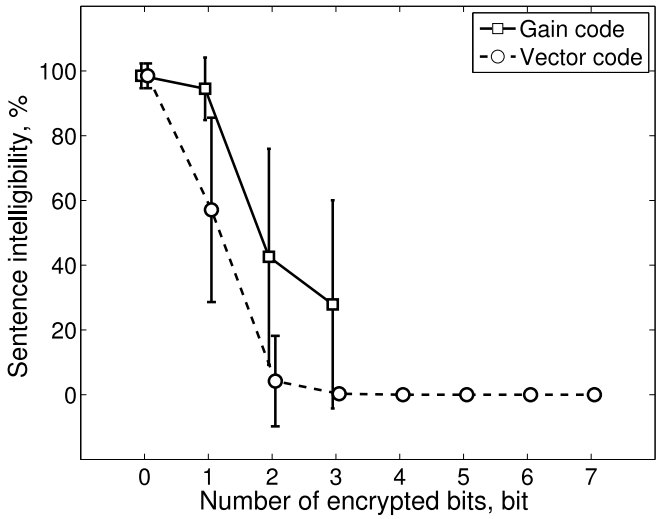


FIGURE 9. Results of speech intelligibility tests for LD-CELP with bit randomization

effective to reduce intelligibility than the gain bits. This superiority of the vector bits to the gain bits for reduction of intelligibility was revealed by the analysis of variance (ANOVA), in which two factors and four levels ($s_g = 0$, $s_v = 0, 1, 2, 3$) and ($s_g = 0, 1, 2, 3$, $s_v = 0$) were tested. A significant two-way interaction between the two factors was found. Therefore, a test of simple main effect was carried out. The test results revealed significant differences ($p < 0.01$) for each combination of ($s_g = 0$, $s_v = n$) and ($s_g = n$, $s_v = 0$) where $n = 1, 2, 3$.

3.4. **CS-ACELP.** Because of its high compression efficiency, CS-ACELP, which is specified in ITU-T G.729, has become a common codec for use with VoIP. Therefore, finding out proper encryption bits for CS-ACELP in the proposed method is important from a practical perspective. Moreover, a highly efficient codec effectively concentrates primal components onto a few bits. This fact implies that the number of bits to be encrypted for reducing intelligibility might be less than in the case with other codecs such as LD-CELP. For the reasons described above, some studies have specifically addressed speech protection for CS-ACELP.

Servetti *et al.* investigated the extent to which speech intelligibility was reduced by encrypting partial bits of CS-ACELP codes [20, 19]. Their results revealed that encryption of properly selected 45% of the overall bits was sufficient for protection comparable to full encryption through listening tests. Similarly, encryption of properly selected 30% of the overall bits was necessary to reduce speech intelligibility sufficiently. However, they suggested no means of transmitting the protected data securely from one side to the other.

A code of CS-ACELP contains numerous parameters, as shown in Table 3, although LD-CELP contains only two. Consequently, it is not feasible to test all combinations of CS-ACELP parameters. Therefore, we first evaluated the individual contribution of each bit to the waveform of speech signal. Then we conducted subjective tests by increasing the number of bits that were subject to the operation.

Table 3. Bit allocation of CS-ACELP

Parameter name	Number of bits	Group
Line spectrum pairs	18	A
Adaptive-codebook delay	13	B
Pitch-delay parity	1	C
Fixed-codebook index	26	D
Fixed-codebook sign	8	E
Codebook gains (stage 1)	6	F
Codebook gains (stage 2)	8	G
Total	80	

To evaluate the individual contribution, "bit inverse sensitivity" was used as a measure. The bit inverse sensitivity was proposed in [10] and was defined as the extent to which the SNR or cepstrum distance changes when only one specified bit of CS-ACELP code is inverted. We carried out listening tests using 100 female speech data excerpted from the set of 1000 Japanese syllabic-balanced sentences to equalize test conditions with others. Seven male and three female students in their twenties with normal hearing ability voluntarily participated in the tests. Results of the tests are depicted in Figs. 10 and 11 for the signal-to-noise ratio (SNR) and cepstrum distance (CD), respectively, and showed a similar tendency, as described in the original paper [10]. Each bit of CS-ACELP code was reordered based on results of the bit inverse sensitivity tests, resulting in a list shown

in Table 4. Alphabets shown in the right columns indicate the group of parameters in Table 3 to which the bit belongs.

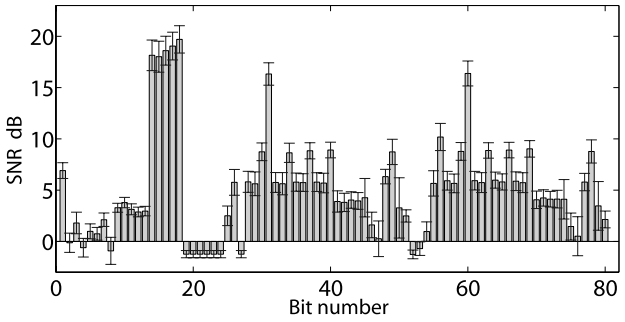


FIGURE 10. Sensitivity to bit inversion in terms of SNR

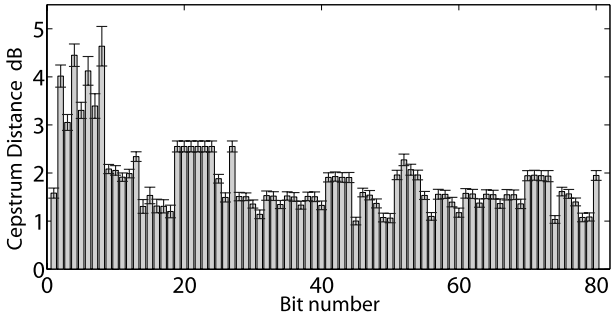


FIGURE 11. Sensitivity to bit inversion in terms of CD

TABLE 4. Order of the effective bit for destructing speech waveforms

#	bit	Group	#	bit	Group
1	8	A	11	27	B
2	4	A	12	13	A
3	6	A	13	52	D
4	2	A	14	9	A
5	7	A	15	53	D
6	19	B	16	10	A
7	20	B	17	12	A
8	21	B	18	54	D
9	22	B	19	51	D
10	24	B	20	75	G

Subjective tests under the same conditions as those for LD-CELP were executed. The results are depicted in Fig. 12. As portrayed in that figure, applying secret sharing to 16 bits, namely 20% of the overall bits of CS-ACELP, is sufficient to reduce speech intelligibility to 0.4%.

4. **Evaluation in the real network.**

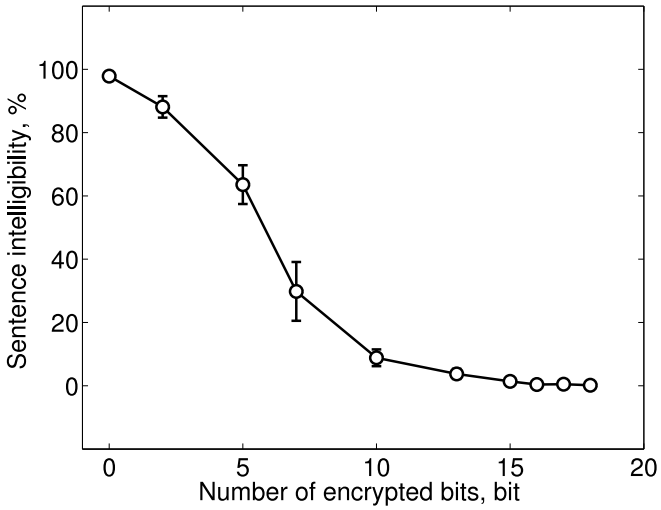


FIGURE 12. Results of speech intelligibility for CS-ACELP with bit randomization

4.1. **System configuration.** The performance of the proposed method was investigated using a real network. Two host computers, which were connected to a LAN of Tohoku University, served as VoIP communication terminals in the test. A preliminarily recorded noise-free speech signal, which was an excerpt from the same set of speech sentences as those used in the listening tests, was used to estimate the net performance of the proposed methods of encoding and decoding. The speech signal was processed by the proposed encoder based on Shamir's secret sharing. The signal was then sent to overlay routers in different networks to accomplish multipath routing. One router was placed in NEC Inc., located about 400 km far from the host computers. The other router was in the same room as the host computers but was connected to a commercial wireless network. Therefore, these two routers were on different paths from the viewpoint of the network and via networks out of the University LAN.

The buffer size for waiting packets is a parameter that can easily control the robustness against jitter and packet loss by sacrificing the delay time. According to the statements related to the delay time on VoIP in ITU-T G.114 [27], in most applications, it is necessary to achieve a delay time of 150 ms to obtain the satisfactory level and 400 ms to obtain the minimum level. In the present system, it takes approximately 70 ms for the encoding and decoding operations, including A/D and D/A. Therefore, buffer sizes of 80 ms and 330 ms were selected for the test.

Differences in QoS of the multiple paths though which the shared data travel might affect the received sound quality. In order to obtain useful information to measure this effect, tests were conducted for the following two cases depending on the allocated routers: in one case, the shared data are directed to two different routers in different networks and hence travel along paths having different QoS; in the other case, all the shared data are directed to a single router and hence travel along a path having the same QoS.

Perceptual evaluation of speech quality (PESQ) [31] was measured. The test conditions are summarized in Table 5.

4.2. **Test results.** The test results are depicted in Fig. 13. Both buffer sizes' delay times were examined only for a condition in which two different routers were used. Otherwise, the buffer size of 330 ms was tested where one router relays both pieces of shared data using two different communication sessions. Accordingly, four tested cases were considered: First, both wired and wireless paths were used with an 80-ms buffer ('Mix8' in Fig. 13).

TABLE 5. Summary of test configuration

Secret sharing	Shamir's secret sharing scheme
Speech codec	LD-CELP
Speech duration	1 min and 45 s
Measure	PESQ [31]
# of paths	2
Buffer size	80 ms, 330 ms
Locations of routers	1. One on wired network and other on wireless network 2. Both on wired network 3. Both on wireless network

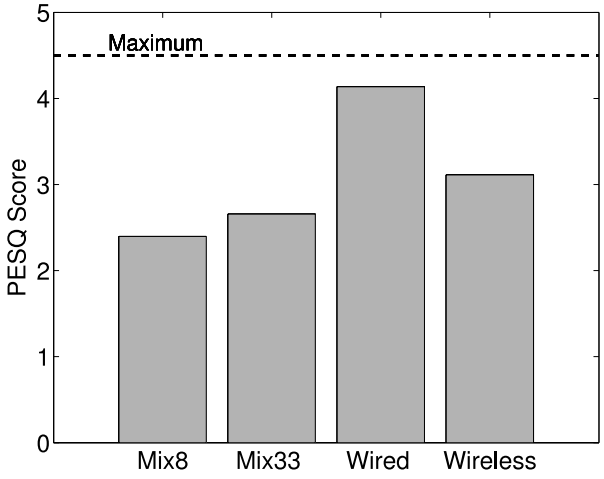


FIGURE 13. PESQ for each tested condition: Mix8, combination of wired and wireless paths with 80-ms buffer; Mix33, combination of wired and wireless paths with 330-ms buffer; Wired, wired path with 330-ms buffer; and Wireless, wireless path with 330-ms buffer.

Second, both wired and wireless paths were used with a 330-ms buffer ('Mix33'). Third, only the wired path was used with a 330-ms buffer ('Wired'); here, two logical sessions were established for delivering the shared data through a single physical path. Finally, only the wireless path was used with a 330-ms buffer ('Wireless'); here, the shared data were delivered in the same manner as that in the case of 'Wired'.

These results demonstrate that when the router was on the wired network, good performance was achieved. In contrast, conditions involving the wireless network showed bad results. The difference between the two buffer sizes in terms of PESQ was slight. There is also a difference between 'Mix33' and 'Wireless', which implies that the use of mixed paths with different QoS would reduce the received sound quality to some extent. Nonetheless, the effect of the path with the lowest QoS appears to be dominant.

5. **Discussion.** The tests using the real network showed that conditions involving a wireless network yielded bad results. To investigate the reason for that, temporal variation of MOS-LQO, which was calculated from PESQ using the transformation function indicated in ITU-T P.862.1 [32], for the conditions with 300-ms buffer size was calculated and is shown in Fig. 14. In the top panel, MOS-LQO remains at a low level in the latter part

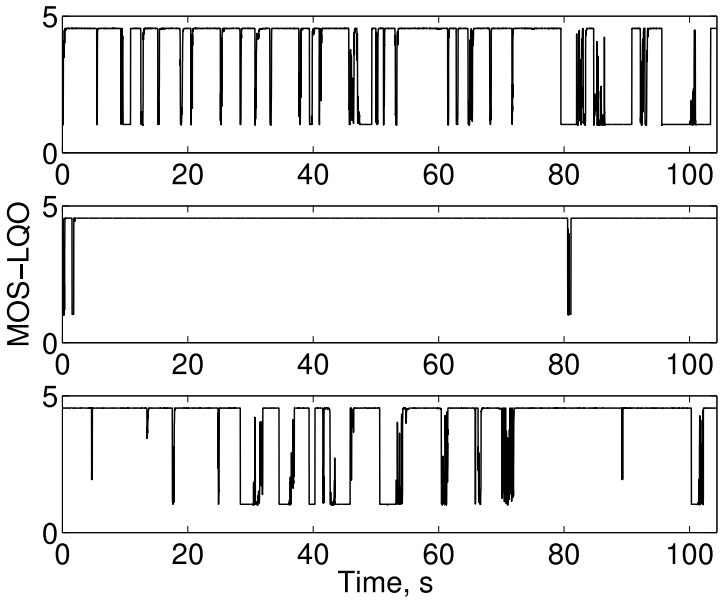


FIGURE 14. Segmental MOS-LQO for cases with 330-ms buffer size: upper, combination of wired and wireless networks; middle, wired network; bottom, wireless network.

of the signal. Similar low-level parts are apparent also in the bottom panel. These occasional low MOS-LQO parts must strongly affect the overall PESQ, resulting in low PESQ values for these two conditions even though quite good MOS-LQO were obtained for most other parts of speech. Considering the middle panel, which includes no wireless network, shows few such low MOS-LQO parts, packet loss must be the reason. Sound quality of LD-CELP markedly declined because of packet loss. Therefore, bad results in PESQ are understandable. Consequently, other speech compression coding would be advantageous in this sense.

As described previously, PCM and CS-ACELP are widely used in current VoIP communication. These encoders are used in current VoIP communication because they are more robust against packet loss than either LD-CELP or ADPCM. As presented in the block diagrams in Figs. 5 and 6, the LD-CELP encoder and decoder contain backward adaptation blocks. The role of these blocks is to predict the current coefficients using those for the preceding several blocks: ADPCM has similar blocks. These blocks therefore might impart successive bad effects on decoding of the subsequent frames once an improper code is received. This structure of LD-CELP or ADPCM is helpful for the present purpose in that the speech intelligibility is expected to be reduced drastically when a malicious person receives only some fragments of the transferred data. Both PCM and CS-ACELP, on the other hand, have no such feedback loop. Consequently, this structure of PCM or CS-ACELP might also weaken the effect of speech intelligibility reduction in the proposed method because bit randomization by secret sharing has little negative influence on the next frame. In fact, an experiment conducted for CS-ACELP showed that encryption of about 45% of the bitstream was necessary to realize content protection equivalent to full encryption of the bitstream [20], although 30% was sufficient for LD-CELP and 25% for ADPCM.

Packet loss often occurs while transmitting packets over the Internet. Even in a simple VoIP system, which usually transfers packets through a single path, packet loss results in the degradation of the received speech quality, and packet loss concealment (PLC) is

required to improve this quality. Multiple description (MD) coding can be used for this purpose [26, 25]. While the MD coding decomposes the original signal into signals as much alike as possible to the original, the secret sharing scheme decomposes the original signal into signals as less alike as possible to the original. Yet in the both methods, collecting all the divided data enables reconstruction of the original. The proposed method transfers packets through multiple network paths, and the loss of even one packet out of the ones that are relevant to a specified time frame results in complete loss of information on that frame due to the character of secret sharing; therefore, the probability of occurrence of missing frame is expected to increase in the proposed method. However, the application of a (k, n) threshold scheme [5, 21] can overcome this problem because it requires only a certain number of packets, which is less than the total packets, to reconstruct the complete original information.

6. **Conclusions.** Combination of the secret sharing scheme and the multiple-routing technique can provide VoIP communications with confidentiality in a different manner from the VoIPsec. Partially applying the secret sharing to speech compression codes reduces the increase of data that are transferred though the network in the proposed method. In fact, the secrecy of 25% for ADPCM, 30% for LD-CELP, and 45% for CS-ACELP is sufficient to reduce speech intelligibility to a certain low level. Further investigation would be necessary to generalize this result over other languages besides Japanese. Robustness against more sophisticated attacks than merely randomizing missing bits also remains as a subject for future investigation. It is possible for the proposed method to increase the security level simply by extending the application of secret sharing to more of the data. However, it should be accompanied by an increase in the amount of transferred data. In addition, the development of packet loss concealment applicable to the proposed method would be necessary for its practical use.

Acknowledgment. The authors thank Profs. Hiroki Shizuya and Masahiro Mambo at Tohoku University for their introduction of the secret sharing scheme and Prof. Akira Arutaki at Kyushu Institute of Technology for his effort in realizing the joint work by NEC Inc., and Tohoku University. This study was partly supported by the Strategic Information and Communications R&D Promotion Programme (SCOPE) of the Ministry of Internal Affairs and Communications of Japan.

REFERENCES

[1] D. Andersen, H. Balakrishnan, F. Kaashoek, and R. Morris, Resilient Overlay Networks, *ACM Symposium on Operating Systems Principles (SOSP)*, pp. 131–145, October 2001.

[2] E. Ayanoğlu, C.-L. I, R. D. Gitlin, and J. E. Mazo, Diversity Coding for Transparent Self-Healing ad Fault-Tolerant Communication Networks, *IEEE Trans. Communications*, vol. 41, no. 11, pp. 1677–1686, 1993.

[3] A. C. Begen, Y. Altunbasak, and O. Ergun, Multi-Path Selection for Multiple Description Encoded Video Streaming, *Int. Conf. on Communications, Anchorage*, vol. 3, pp. 1583–1589, May 2003.

[4] R. Barbieri, D. Bruschi, and E. Rosti, Voice over IPsec: Analysis and Solutions, *Proc. of the 18th Annual Computer Security Applications Conference (ACSAC)*, pp. 261–270, 2002.

[5] G. R. Blakley, Safeguarding cryptographic keys, *AFIPS 1979 National Computer Conference*, vol. 48, pp. 313–317, June 1979.

[6] D. Butcher, X. Li, and J. Guo, Security Challenge and Defense in VoIP Infrastructures, *IEEE Trans. Systems, Man, and Cybernetics – Part C: Applications and Reviews*, vol. 37, no. 6, pp. 1152–1162, November 2007.

[7] J. D. Gibson, A. Servetti, H. Dong, A. Gersho, T. Lookabaugh and J. C. De Martin, Selective Encryption and Scalable Speech Coding for Voice Communications over Multi-Hop Wireless Links, *Proc. of Military Communications Conference (MILCOM)*, vol. 2, pp. 792–798, October 2004.

[8] M. Hashimoto, Y. Minami, K. Matsuo, and A. Koike, A Study on Data-Size-Reduciton Methods for Hierarchical Secret Image Sharing Method, *IPSJ SIG Technical Report*, vol. AVM-52, pp. 27-32, 2006.
[9] A. B. Johnston and D. M. Piscitello, *Understanding Voice over IP Security*, Artech House Telecommunications, 2006.
[10] A. Kataoka and S. Hayashi, A cryptic encoding method for G.729 using variation in bit-reversal sensitivity, *IEICE Trans. Information and Systems*, vol. 87, no. 6, pp. 1224–1232, 2004.
[11] S.-J. Lee and M. Gerla, Split Multipath Routing with Maximally Disjoint Paths in Ad hoc Networks, *Proc. of Int. Conf. on Communications*, vol. 10, pp. 3201-3205, June 2001.
[12] C.-C. Lin and W.0H. Tsai, Secret image sharing with capability of share data reduction, *Opt. Eng.*, vol. 42, no. 8, pp. 2340–2345, 2003.
[13] W. Lou and Y. Fang, A Multipath Routing Approach for Secure Data Delivery, *Proc. of Military Communications Conference (MILCOM)*, vol. 2, pp. 1467–1473, October 2001.
[14] M. Naor and A. Shamir, Visual Cryptography, *Proc. of Workshop on the Theory and Application of Cryptographic Techniques (EUROCRYPT)*, pp. 1–12, May 1994.
[15] R. G. Ogier, V. Rutenburg and N. Shacham, Distributed Algorithms for Computing Shortest Pairs of Disjoint Paths, *IEEE Trans. Information Theory*, vol. 39, no. 2, pp. 443–455, 1993.
[16] D. S. Phatak and T. Goff, A Novel Mechanism for Data Streaming Across Multiple IP Links for Improving Throughput and Reliability in Mobile Environments, *Proc. of IEEE Conference on Computer Communications (INFOCOM)*, pp. 773–781, June 2002.
[17] S. Sakamoto, Y. Suzuki, S. Amano, K. Ozawa, T. Kondo, and T. Sone, New lists for word intelligibility test based on word familiarity and phonetic balance, *J. Acoust. Soc. Japan*, vol. 54, no. 12, pp. 842–849, 1998.
[18] R. Salami, C. Laflamme, J.-P. Adoul, A. Kataoka, S. Hayashi, T. Moriya, C. Lamblin, D. Massaloux, S. Proust, P. Kroon, and Y. Shoham, Design and Description of CS-ACELP: A Toll Quality 8 kb/s Speech Coder, *IEEE Trans. Speech and Signal Processing*, vol. 6, no. 2, pp. 116–130, 1998.
[19] A. Servetti and J. C. De Martin, Perception-based Selective Encryption of G.729 Speech, *Proc. of Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 1, pp. 621–624, May 2002.
[20] A. Servetti and J. C. De Martin,, Perception-Based Partial Encryption of Compressed Speech, *IEEE Trans. Speech and Audio Processing*, vol. 10, no. 8, pp. 637–643, 2002.
[21] A. Shamir, How to share secret, *Communications of the ACM*, vol. 22, no. 11, pp. 612–613, 1973.
[22] D. Sidhu, R. Nair and S. Abdallah, Finding Disjoint Paths in Networks, *Proc. of Conference on Communications Architecture & Protocols*, Zürich, pp. 43–51, September 1991.
[23] A. Tsirigos and Z. J. Haas, Multipath Routing in the Presence of Frequent Topological Changes, *IEEE Communications Magazine*, vol. 39, no. 11, pp. 132–138, 2001.
[24] G. S. Vernam, Secret signaling system, U. S. Patent 1310719, 1919.
[25] Y. Wang, A. R. Reibman and S. Lin, Multiple Description Coding for Video Delivery, *Proceedings of IEEE*, vol. 93, no. 1, pp. 57–69, 2005.
[26] J. K. Wolf, A. D. Wyner and J. Ziv, Source Coding for Multiple Descriptions, *Bell Sys. Tech. J.*, vol. 59, no. 8, pp. 1417–1426, 1980.
[27] ITU-T G. 114, One-way transmission time, the International Telecommunication Union, May 2003.
[28] ITU-T G. 726, 40, 32, 24, 16 kbit/s Adaptive Differential Pulse Code Modulation (ADPCM), the International Telecommunication Union, December 1990.
[29] ITU-T G. 728, Coding of speech at 16 kbit/s using low-delay code excited linear prediction, the International Telecommunication Union, September 1992.
[30] ITU-T G.729, Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear prediction (CS-ACELP), the International Telecommunication Union, March 1996.
[31] ITU-T P. 862, Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end quality assessment of narrow-band telephone networks and speech codecs, the International Telecommunication Union, February 2001.
[32] ITU-T P. 862. 1, Mapping function for transforming P.862 raw result scores to MOS-LQO, the International Telecommunication Union, November 2003.