Prevention of Anomalous SIP Messages

Ming-Yang Su and Chung-Chun Chen

***Abstract*—Voice over internet protocol (VoIP) communication services have been widely used in recent years. VoIP is a real-time communication service prone to attack; once attacked, it is more difficult to be addressed in real time as compared to off-line Email. The attacker may transmit large sum of online request messages to the SIP server to cause congestion or even breakdown of the machine to make it unable to provide services. The proposed defense mechanism will filter malformed messages inconsistent with SIP protocol and judge whether SIP server suffers any flooding attack. Once attack is detected, this study will immediately update the black-list on the SIP server and use the SIP server built-in function to block the online requests of specific users on the black-list.**

***Index Terms*—Flooding attack, malformed messages, session initiation protocol (SIP), voice over internet protocol (VOIP).**

I. Introduction

VoIP (voice over internet protocol) is the technology to transmit voice information via IP network; it is the so-called Internet telephony. With the continuous progress in network technology, VoIP applications are becoming more widely used. VoIP is constructed on the existing TCP/IP, any TCP / IP security threats will pose a threat to VoIP including the most common flooding attack. SIP (session initial protocol) protocol [1]–[3] is call protocol of VoIP established with the open source code. SIP is defined on the Application Layer of the OSI model with the main purpose to establish, modify and interrupt multimedia session. SIP message uses the text-based, similar to HTTP (Hyper Text transmission p rotocol) protocol; caller and callee identification address is in the form similar to E-mail.

The main components of SIP include: User agents (UA) and servers. User agent refers to the component at the user end, it can be a software Soft-Phone or a hardware SIP telephone. User agent client (UAC) refers to the party initializing the connection, namely, the caller while user agent sedrver (UAS) refers to the callee. UAC is responsible for producing requests and UAS is responsible for producing response according to the request. Before UA (including UAC and UAS) using SIP services, it should be registered in SIP Server. UA will transmit the REGISTER message to Server, Server will judge whether the user is authorized to register after receiving the message. If the user is authorized, it will respond with 200-OK. In general, the Server will set a validation procedure, namely, the response 401-Unauthorizedmessage, and provide encrypted digest message in the header field of response message, for example, the nonce value for the encryption password. After receiving the message, UA will re-transmit to the REGISTER and use the newly received nonce value to produce the encrypted digest to be put in the header field. After receiving the request signals, the Server will compare the encrypted digest with the original digest and send the response of 200-OK to UA if no error is found to complete the registration procedure.

SIP call connection is established as shown in Fig. 1. UAC will send INVITE message to SIP Server. After analysis, SIP Server will send the message to UAS. Upon reception, UAS starts to ring the bell and respond the message of 180-Ringing. After receiving the message, the SIP Server will send back to UAC for notification of bell ringing and waiting for response. The UAS user then sends 200-OK to UAC. UAC will send back ACK request message to UAS, and thus the two parties will establish the RTP (Real Time Transport Protocol) [4] channel for talking. The voice packets transmitted via the RTP channel will not be transmitted by the SIP Server. When UAC is to end the call, it will send the BYE message to the SIP Server. SIP Server will transmit the message to UAS, which will send back the message of response 200-OK.

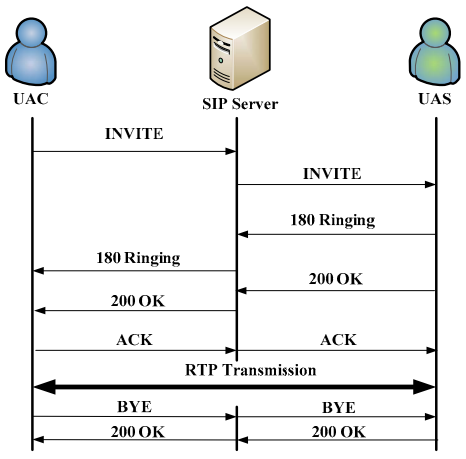


Fig. 1. SIP connection establishment.

SIP protocol message is usually transmitted by UDP. Each message is composed of Command Line, Header, and Body. Command Line is to confirm the type of message. There are six types of requests and 6 types of responses. The six types of request messages include INVITE, BYE, ACK, OPTIONS, CANCEL and REGISTER. The six types of response messages are usually represented by codes including 1XX (notification and response, for example, is to call 180-Ring), 2XX (successful response, such as 200-OK), 3XX (call forward response), 4XX (call failed, such as the 401-Unauthorized), 5XX (server failure), 6XX (global

Manuscript received January 20, 2013; revised March 15, 2013. This work was supported in part by the National Science Council with contracts NSC 101-2622-E-130-001-CC3 and NSC 101-2221-E-130-016.

The authors are with the Department of Computer Science and Information Engineering, Ming Chuan University, Taoyuan, Taiwan (e-mail: minysu@mail.mcu.edu.tw, hatsumi.17y@gmail.com).

DOI: 10.7763/IJFCC.2013.V2.203

failure). when the caller sends INVITE request message, it will transmit the SDP (Session Description Protocol) [5]. SDP includes the caller end media format, address, port. When is the callee responds, it determines whether to accept or reject the message. In this way, you can achieve the consistent transmission of media needs.

This study uses the SIP standard format of RFC 3261 [3] as the basis to filter out the malformed packets in advance. However, the anomalous message detection cannot effectively prevent the message flooding attack. Therefore, the Chi-square test [6] is used to further judge whether SIP server suffers flooding attack. The remainder of this paper is organized as follows: Section II reviewed related works and introduces the Chi-square test technology; Section III describes the research methods; Section IV presents the experimental architecture and the data analysis; Section V offers the conclusion.

II. Related Works

This section describes some of the relevant literature on the attack using malformed messages and SIPS message flooding attack, and introduces the Chi-square test technology used in this study. Flooding attack is the most commonly seen in the attack on the SIP server. The attacker sends large sum of SIP messages to the server, resulting in great consumption of resources to break down the normal operation of the machine, such as the Invite flooding, Cancel flooding, and Bye flooding. Zhou *et al*. [7] focused on the prevention of SIP Register servers from CPU exhausted DoS attack. Register server should conduct identity confirmation and decryption/ encryption computation. In the case of large sum of registration requests at the same time, the Register server may not be able to process the normal registration procedure or even break down. The method used in [7] is to use the previous legitimate call records' IP addresses for judgment when the connection request increases considerably under the possible DoS attack. If there is no past record, the user will be regarded as an attack and the provision of relevant services will be suspended. In [7], the authors used the Bloom filter algorithm [8] to judge the IP address. Bloom filter is a simple algorithm almost consuming no CPU resource and is used to determine whether the new data (for example IP address) have been stored in the memory.

This study uses a statistically known as Chi-square test [6] mechanism to design and detect whether SIP server is under the INVITE/CANCEL/BYE flooding attack. Therefore, this section will briefly introduce the Chi-square test. Among statistical data, some are quantitative data and some are categorical data. The main parameters of quantitative data are average and variance while the categorical data's main parameter is proportion. The processing categorical data will be categorized into different types according to their attributes and the distinguished categories are measured or represented by the sequence. The sequence scale value is random and therefore, such data cannot be described by average or variance like the quantitative data. Instead, they can be described by the proportion and sequence of various categories. Therefore, the hypothesis test of such categorical data is to test the category proportion or level by mainly using the Chi-square test.

Chi-square test -test of goodness of fit is to test the closeness of "observed frequency" and "expected frequency" of the null hypothesis H_0 to determine whether it is consistent with a specific distribution. In other words, it is to compare the differences of the "observed frequency" obtained by comparing the samples of categorical data and "expected frequency" when null hypothesis H_0 is supported. Chi-square test value, i.e., x^2 is calculated as shown below:

$$x^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \quad (1)$$

$$E_i = np_i \quad (2)$$

where O_i is the No. i Group of samples' observed frequency, E_i is the expected frequency, k is the number of categories, n is the number of samples, p_i is the probability of categories. When O_i and E_i differs greatly, chi-square value will increase, representing that the sample data cannot support null hypothesisH_0, and thus H_0 should be rejected, namely,

if $x^2 > x^2_{k-m-1,\alpha}$, then reject H_0 if $x^2 \le x^2_{k-m-1,\alpha}$, then accept H_0 (3)

where $x^2_{k-m-1,\alpha}$ is known as the Chi-square critical values, which can be learnt from the table as shown in Table I. Chi-square critical values k-1-m is known as the *df* (degree of freedom), m is the number of estimated parameters. When computing the expected frequency, no parent parameter should be estimated in advance, hence, m = 0; i.e., *df* is k-1. In general, α is set as 0.05, representing the tolerance of 5% error. The rate of error α in this study is set as 0.05 in general statistics.

TABLE I: CHI-SQUARE DISTRIBUTION CRITICAL VALUES

Deg. of Freedom	Probability (α)										
	0.95	0.90	0.80	0.70	0.50	0.30	0.20	0.10	0.05	0.01	0.001
1	0.004	0.02	0.06	0.15	0.46	1.07	1.64	2.71	3.84	6.64	10.83
2	0.10	0.21	0.45	0.71	1.39	2.41	3.22	4.60	**5.99**	9.21	13.82
3	0.35	0.58	1.01	1.42	2.37	3.66	4.64	6.25	7.82	11.34	16.27
4	0.71	1.06	1.65	2.20	3.36	4.88	5.99	7.78	9.49	13.28	18.47
5	1.14	1.61	2.34	3.00	4.35	6.06	7.29	9.24	**11.07** 20.52		15.09
6	1.63	2.20	3.07	3.83	5.35	7.23	8.56	10.64	12.59	16.81	22.46
7	2.17	2.83	3.82	4.67	6.35	8.38	9.80	12.02	14.07	18.48	24.32
8	2.73	3.49	4.59	5.53	7.34	9.52	11.03	13.36	15.51	20.09	26.12
9	3.32 14.68	4.17	5.38	6.39	8.34	10.66	12.24		16.92	21.67	27.88
10	3.94 15.99	4.86	6.18	7.27	9.34	11.78	13.44		18.31	23.21	29.59
	Non-significant								Significant		

The Chi-square test application is shown in an example as follows: when throwing a dice for 300 times, and the times of each point in appearance are as shown in Table II, is the dice throwing a fair one?" In this example, the null hypothesisH_0 is that "is it a fair dice", according to (1), chi-square value computation process is as shown in Table III, and the final

chi-square value is 7.0308. In this example, the number of categories k = 6, *df* = k-1-m, m is the number of estimation parameters. When computing the expected frequency, as there is no parent parameter to be estimated in advance, i.e., m=0 (there are 6 faces of the same probability as it is known), therefore, *df* = k-1 = 5. In the case of the preset error tolerance rate α=0.05, by checking Table I, the Chi-square critical value is11.07. As 7.0308 < 11.07, therefore, H_0 is acceptable, i.e., it is a fair dice.

TABLE II: TIMES OF EACH POINT IN APPEARANCE

Point	1	2	3	4	5	6
Times of appearance	40	60	60	45	45	50

TABLE III: CHI-SQUARE VALUE COMPUTATION PROCESS

	O	P	E	(O-E)	$(O-E)^2$	$(O-E)^2/E$
1	40	1/6	300*0.166=49.8	40-49.8=-9.8	96.04	1.93
2	60	1/6	300*0.166=49.8	60-49.8=10.2	104.04	2.09
3	60	1/6	300*0.166=49.8	60-49.8=10.2	104.04	2.09
4	45	1/6	300*0.166=49.8	45-49.8=-4.8	23.04	0.46
5	45	1/6	300*0.166=49.8	45-49.8=-4.8	23.04	0.46
6	50	1/6	300*0.166=49.8	50-49.8=0.2	0.04	0.0008
Σ	300	1				7.0308

III. RESEARCH RESULTS

The main functions of the proposed system include the malformed message detection and INVITE/CANCEL/BYE flooding attack detection mechanisms. When network packets come in, the system will intercept the SIP packets to the Malformed Detection module. If the packets are detected as anomalous, the attack source will be sent to the SIP server for the updating of the black-list. Otherwise, it will send the packets to the Chi-square test for statistical detection to determine whether the system is under the flooding attack. If it is confirmed as a flooding attack, the attack message is then stored in the database and the attack source is sent to the SIP server to update the server built-in black-list. SIP malformed detection is implemented by following the principles as stipulated in RFC 3261 as shown in Table I. If the packet format is inconsistent with the norms, it is regarded as an attack. In this section, it is mainly conducted by string comparison as it is simpler. This section illustrates the Chi-square test.

VoIP flooding attack detection is based on the Chi-square Goodness of Fit Test. If the number of packets of various categories is consistent with the expected proportional distribution, it is regarded as normal; otherwise, it is regarded as anomalous. When SIP message passes the malformed message detection, it will enter the range of VoIP flooding inspection. This study checks whether the number of packets of INVITE/CANCEL/BYE of each extension machine is beyond the threshold or not. If it is beyond the threshold, the system will enter into the statistical state and conduct the Chi-square test. If it is beyond the Chi-square critical value, it is then regarded as an attack. The setting of the threshold is basically the statistical average value in a unit time, the number of packets of INVITE/CANCEL/BYE on individual extension machine should be more than the value to proceed to the procedure of Chi-square test. The Chi-square testing results are then determines whether the SIP server built-in black-list should be updated to lock the extension machine. The main steps of the system are detailed as below:

TABLE IV: SIP MESSAGES AND MEANINGS COUNTED

	Source	Corresponding message
INVITE	Caller	
Ringing	System	INVITE
Code183	System	INVITE/Code468
ACK	Caller	Code468
CANCEL	Caller	
Code487	System	CANCEL
OK	System	CANCEL
BYE	Caller	
BYE	System	BYE(Caller)
OK	System	BYE

Step 1. SIP Packet Characteristic Acquisition

This system counts the number of SIP messages of each extension machine. Table IV shows all the types and meanings of SIP messages. INVITE is Caller-sent message, Callee may respond in one of the three states including: online, offline or busy. When Callee is online, it will respond with the message of Ringing, which is transmitted via the SIP server to Caller. When Callee is offline, SIP Server will respond with the message of Code183, and the Caller responds the corresponding ACK. when Callee is busy, Callee responds Code468 message to SIP sever, SIP server transmits Code183 to Caller, which responds with the corresponding message of ACK. CANCEL is Caller message to end the connection with the extension machine. Code487 is system response after receiving the CANCEL message. OK is the system message after sending out Code487. BYE is Caller message to end the call. OK is the system response message after receiving the Caller-sent BYE message.

Step 2. Normal Threshold Setting

The connection request beyond threshold should be listed as the object of observation for the following Chi-square test. Threshold is the records of the hourly Call view based on the normal calls of the business days each week including the times of INVITE, CANCEL and BYE messages. By adding up the same period of the day, the average μ and standard deviation σ are obtained as shown in (4) and (5). This study sets the threshold of each time segment as μ+3σ. This is based on the basic principle: random variables mostly are

distributed along both sides of the average and the most of the values concentrate on the range of average ±3 standard deviations. The counting of Call view in this study is accumulated, the threshold of the first hour is computed by the number of INVITE/CANCEL/BYE during the period of 0:00am~1:00am, the threshold of the second hour is computed by the times of INVITE/CANCEL/BYE in the period 0:00 am ~ 2:00 am, and so forth. The 24 normal thresholds are as shown in Fig. 2. In operation, the system uses the corresponding threshold value according to the time point of the event.

$$\mu = \frac{day1 + day2 + day3 + day4 + day5}{5} \quad (4)$$

$$\sigma = \sqrt{\frac{1}{5}(day1-\mu)^2 + (day2-\mu)^2 + (day3-\mu)^2 + (day4-\mu)^2 + (day5-\mu)^2} \quad (5)$$

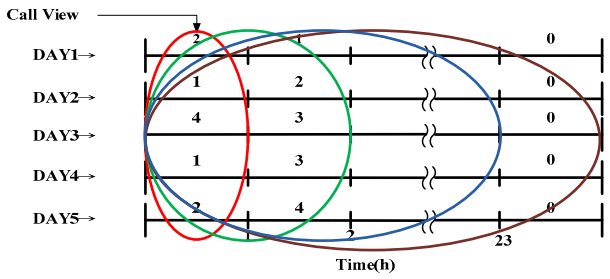


Fig. 2. Computation of normal threshold.

Step 3: Chi-Square Test

If the sum of the INVITE/CANCEL/BYE messages of a Caller is beyond the normal threshold, the caller will be listed as a subject of observation for Chi-square test. If the chi-square values are more than the Chi-square critical values, it will be regarded as an attack and the system will update the SIP server built-in black-list in real time to lock down the extension. The INVITE Chi-square tests are illustrated as shown below. The parts of CANCEL and BYE are similar, so they are omitted here.

Regarding the INVITE message Chi-square test, three types of SIP messages are considered in this study: #INVITE, (#Code183 + #Ringing), and (#ACK + #Ringing), Therefore, in Chi-square test, the number of categories k = 3, *df* is k - 1 = 2; under the hypothesis of the 5% error tolerance, i.e., α = 0.05, Chi-square critical values by referring to Table I are $x^2_{df,\alpha} = x^2_{2,0.05} = 5.991$. Therefore, by acquiring and counting the number of the three types of SIP messages in unit time from the network flow, the results are inputted into the Chi-square test equation of (1). If it is smaller or equal to (≦) 5.991, it is confirmed there is no INVITE flooding attack; otherwise, if the value is greater than 5.991, it is confirmed that there is the INVITE flooding attack.

Regarding the computation of chi-square value, we set E_{ext} = (#INVITE + (#Code183 + #Ringing) + (#ACK + #Ringing) - # (Code183|486) - #(ACK|486)) × 1/3, as the callee may be in one of the three possible cases.

Case 1. Callee online: SIP server sends Ringing to caller, resulting in #INVITE: (#Ringing): (#Ringing) = 1 : 1 : 1, then E_{ext} = (#INVITE + (#Ringing) + (#Ringing)) ×1/3.

Case 2.Callee offline: SIP server sends Code183 to caller, which sends ACK to SIP server, resulting in #INVITE: (#Code183): (#ACK) = 1 : 1 : 1, then, E_{ext} = (#INVITE + (#Code183) + (#ACK)) × 1/3.

Case 3. Callee busy: SIP server sends Ringing to the caller, resulting in #INVITE : (#Ringing) : (#Ringing) = 1 : 1 : 1, in addition, callee will send Code486 to SIP server, and the SIP server sends Code183 to caller, which sends ACK to SIP server. As Code486 triggers Code183 and ACK, they are recorded as (Code183 | 486) and (ACK | 486).

In summary of the above three cases, we have #INVITE : (#Code183 + #Ringing) : (#ACK + #Ringing) = 1 :1:1, and E_{ext} = (#INVITE + (#Code183 + #Ringing) + (#ACK + #Ringing) - #(Code183|486) - #(ACK|486)) × 1/3. Therefore, when conducting the INVITE Chi-square test, by acquiring and counting the number of SIP messages including #INVITE, (#Code183 + #Ringing), and (#ACK + #Ringing) in unit time, we can compute the chi-square valuex^2 as shown below:

$$x^2 = \frac{(\#INVITE - E_{ext})^2}{E_{ext}} + \frac{((\#Code183+\#Ringing) - E_{ext})^2}{E_{ext}} + \frac{((\#ACK+\#Ringing) - E_{ext})^2}{E_{ext}}$$

If $x^2 \le x^2_{2,0.05} = 5.991$, it is regarded as normal, otherwise, $x^2 > x^2_{2,0.05} = 5.991$, it is determined as an attack. The current #INVITE is set as the threshold of the extension. The purpose is to prevent the continuous increase in chi-square value due to number of #INVITE. In other words, attack may occur in a certain unit time and may stop in the following unit time. The value of #INVITE will continue to grow and the system will believe that the attack is going on to result in false positives.

IV. Experimental Results

The experimental environment of this study is as shown in Fig. 3. The Attacker simulation computer has Windows XP, Pentium(R) 4 3.00GHz and 512MB RAM with two free software packages including SIPp [9] and SiVuS [10]. The two computers simulating UAC and UAS have Windows XP, Pentium(R)4 3.00 GHz and 512MB RAM. VoIP PBX computer Elastix [11] with Intel(R) Core (TM)2 Quad 2.66GHz and 4GB RAM, as well as the installation of IDS. The IDS detection mechanism is written in JAVA and JNETPCAP [12], JNETPCAP provides the online packet acquisition function. Before using this system, the general flow sum of each extension is collected, and the general flow volume is simulated by SIPp. This study uses SIPp to simulate 50 extensions as UAC and set that 50 extensions have randomly produced 10-30 Calls every day. The number of calls in every hour in each extension is recorded to set the normal thresholds of each extension at different time periods according to the method as shown in Step 2 of the previous section.

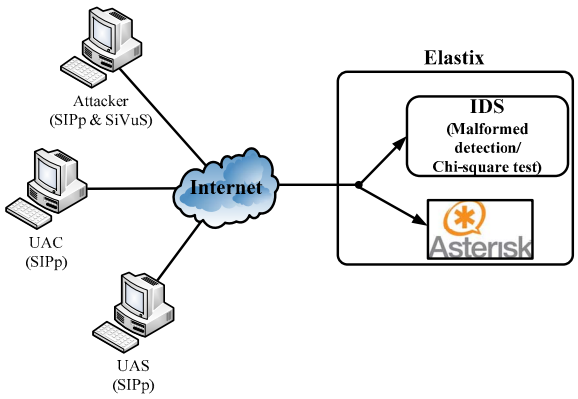
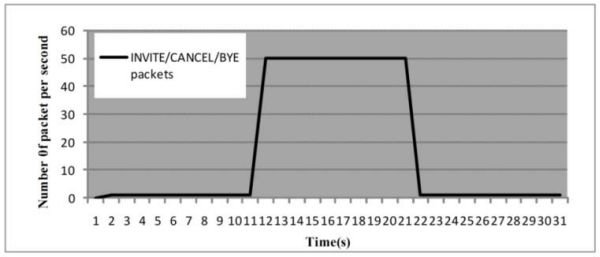
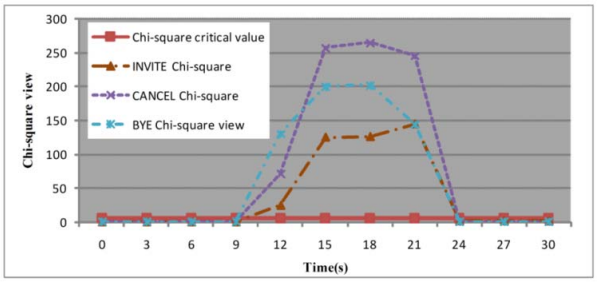


Fig. 3. Experimental environment.

This study uses SIPp as the tool of flooding attack of INVITE/CANCEL/BYE. As shown in Fig. 4(a), after 10 sec, the amount of messages increases dramatically and large volumes of INVITE/CANCEL/BYE messages are kept until the 20th second. This study computes the chi-square value every three seconds and the changes in chi-square value are as shown in Fig. 4(b). The chi-square values as shown in the figure reflect that some message (INVITE/CANCEL/BYE) considerably increases and the chi-square value decreases when the message returns back to normal. The red lines as shown in Fig. 4(b) are the preset Chi-square critical values. When the chi-square value is beyond the critical value it is determined as an attack.



(a)



(b)

Fig. 4. Chi-square test data(a) message transmission (b) INVIT/CANCEL/BYE chi-square value.

V. Conclusions

The proposed system in this study equipped the function of INVITE/CANCEL/BYE flooding attack detection. It applied Chi-square test to detect the flooding of a caller in certain unit time. Once an attack is detected, it will immediately update the SIP server built-in black-list to prevent the caller from attack. Chi-square test is to find out the proportional relationship of fixed response messages when SIP VoIP are connected and use the Chi-square test of goodness of fit characteristic, i.e., consistency with the expected proportional distribution or not, to detect whether each characteristic's proportion is anomalous. According to the experimental results, the detection method of this study can effectively detect the malformed messages and the malicious flooding attack.

References

[1] D. Butcher *et al.*, "Security challenge and defense in VoIP infrastructures," *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, vol. 37, pp. 1152-1162, 2007.
[2] S. El Sawda and P. Urien, "SIP security attacks and solutions: a state-of-the-art review," in *Proc. Information and Communication Technologies, 2006. ICTTA '06. 2nd*, pp. 3187-3191, 2006.
[3] J. Rosenberg. RFC 3261. [Online]. Available: http://www.ietf.org/rfc/rfc3261.txt.
[4] H. Schulzrinne *et al.*, "RTP: a transport protocol for real-time applications," in *IETF,RFC1889*, January 1996.
[5] M. Handley and V. Jacobson, "Session description protocol(SDP)," in *IETF,RFC2327*, April 1998.
[6] Y. Tianlu *et al.*, "A novel voip flooding detection method basing on call duration," in *Proc. 2010 First International Conference on Pervasive Computing Signal Processing and Applications (PCSPA)*, pp. 1158-1162, 2010.
[7] C. V. Zhou, C. Leckie, and K. Ramamohanarao, "Protecting SIP server from CPU-based DoS attacks using history-based IP filtering," *IEEE Communications Letters*, vol. 13, no. 10, pp. 800-802, 2009.
[8] B. H. Bloom, "Space/time trade-offs in hash coding with allowable errors," *Communications of the ACM*, vol. 13, no. 7, pp. 422–426, 1970.
[9] SIPP. [Online]. Available: http://sipp.sourceforge.net/.
[10] SiVuS. [Online]. Available: http://nil.uniza.sk/sip/tools/sivus-voip-vulnerability-scanner.
[11] Elastix. [Online]. Available: www.elastix.org.
[12] JNetPcap OpenSource. [Online]. Available: http://jnetpcap.com/.

Ming-Yang Su received his B.S. degree from the Department of Computer Science and Information Engineering of Tunghai University, Taiwan in 1989, and received his M.S. and Ph.D. degrees from the same department of the National Central University and National Taiwan University in 1991 and 1997, respectively. He is an IEEE member, and currently a professor of the Department of Computer Science and Information Engineering at the Ming Chuan University, Taiwan. His research interests include network security, intrusion detection/prevention, malware detection, mobile ad hoc networks, SIP security and wireless sensor networks.

Chung-Chun Chen received the M.S. degree in 2011, from the Department of Computer Science and Information Engineering of Ming Chuan University, Taoyuan, Taiwan. His research interests are in the areas of network security, SIP security, and intrusion detection/prevention.