Signal Processing and Renewable Energy

December 2017, (pp.37-50) ISSN: 2588-7327 eISSN: 2588-7335

# Detecting Flooding Attacks on IMS Networks Using Kullback-Leibler Divergence and Triple EWMA

Noorallah Hemmati Doust<sup>1</sup>, Mansour Nejati Jahromi<sup>1\*</sup>

<sup>1</sup> Department of Electrical Engineering, Shahid Sattary Aeronautical University of Science and Technology Tehran, Iran.

#### Abstract

The IP Multimedia Subsystem (IMS) is a platform for the exchange of multimedia communications that was proposed by 3GPP as of the year 2002. The 3GPP proposal called for the integration of mobile cellular networks and internet technology using a completely IP-based structure. The IMS uses the protocols defined by the IETF, such as SIP, RTP and others. SIP is the backbone of the IMS network, where it is used for signaling and multimedia services control. However, security vulnerabilities are inherent in such integration. When the IMS architecture is opened for easy network access and the use of SIP, it is far more vulnerable to SIP flooding attacks. This has presented a significant security problem in new networks. In the presented method for detection, network traffic is captured in two phases, being the training phase and the test phase. The distance between the probable distributions of SIP messages in these two phases is then measured using the Kullback-Leibler divergence. Then, an adaptive threshold is defined for the Kullback-Leibler divergence which, when passed, means that an attack has occurred. The adaptive threshold is accounted for by the use of a Triple Exponential Moving Average (TEMA), and the performance of the presented detection method in various situations of normal traffic and massive attacks is evaluated. The parameters  $\alpha$ ,  $\beta$ ,  $\varepsilon$ , and  $\gamma$  are used for estimating the threshold and setting a safe margin for authorized traffic. In addition, the effect of changes of the estimate and setting parameters is evaluated.

**Keywords:** IP Multimedia Subsystem, SIP protocol, Flooding attack, Attack detection, Adaptive threshold.

# **1. INTRODUCTION**

At the time the IP Multimedia Subsystem

\*Corresponding Author's Email: nejati@aut.ac.ir

(IMS) was introduced, the 3GPP organization proposed a security mechanism to protect networks from unauthorized access and interference. However, due to the extension and complexity of the next generation of networks, there exist new threats and attacks against which this mechanism cannot confidently stand up. Flooding attacks that are generated by sending volumes of messages as a flood, targeting networks and aiming to consume and crush network resources, are one such threat. In order to differentiate network problems from normal behavior, the use of a detection threshold is highly important. The detection threshold for these attacks can be fixed or adaptive although, given the variable nature of network traffic, the adaptive threshold is typically more suitable than the fixed threshold. In the layer structure of the IMS and standardization of existent services for establishing a session via IMS, various protocols handle special functions so that one complete call can be established. Some of these protocols are SIP, RTP, TCP/UDP, and AAA. SIP is an application layer and text-based protocol which was standardized by the Internet Engineering Task Force (IETF) for managing sessions.

# 2. IMS ARCHITECTURE

The IP Multimedia Subsystem (IMS) is a divergence communication architecture that merges cellular and internet technologies for the delivery of smooth voice, video and data in a unit network [1]. The architecture of this network is open, IP-based and independent of access technology. The IMS uses IETF-defined protocols such as SIP, TCP, and UDP to facilitate the integration of mobile and internet services. These protocols are exposed to various types of security threats that cause IMS vulnerabilities [2]. In this network SIP and DIAMETER protocols are used for signaling between internal and external elements. The IMS network is independent of

access technology. This means that the IMS network can be accessed by various carriers through IP networks [3].

In the following, some notifications about this network will be mentioned. Some features of this network will be described. The nodes of the IMS core subsystem are the Home Subscriber Server (HSS); the Subscriber Location Function (SLF); the Call Session Control Function (CSCF); Application Server(s) (AS); Media Resource Function (MRF), which is divided into a media resource function controller (MRFC) and a media resource function processor (MRFP); the Breakout Gateway Control Function (BGCF); and the Public Switched Telephone Network (PSTN) gateway to a signaling gateway (SGW) and a media gateway (MGW). The HSS is the main storage server for information related to the user, including all user information necessary for exchanging multimedia data. Networks with more than one HSS require an SLF. An SLF is a simple database that maps users' addresses to the HSS's. The Proxy CSCF (P-CSCF) is the first contact point between an IMS terminal and IMS network that is identified by DHCP and PDP that can be called a CSCF. The Interrogating CSCF (I-CSCF) is the entrance node to the home scope. For this reason, it is also called a CSCF entrance Gateway. The Serving CSCF (S-CSCF) continues the established session with the network operator. As long as a user is accessing the service, it provides service to the user. The Services Access-Service Capability Server (SA-SCS) creates the performance of an interference up to the OSA server working range and includes all properties of the OSA. The IP Multimedia Service Switching Function (IM-SSF) allows the use of Customized Applications for Mobile networks Enhanced Logic (CAMEL) services which is related to GSM

in the IMS. The MRFC controls the media plan resources. The MRFP provides services such as media exchange, conference establishing, and voice/video email. Basically, the BGCF is a SIP server that provides the path between networks and routes calls based on telephone numbers. It works with calls that originate at an IMS terminal and are connected to a user through a circuit switched (CS) network such as a PSTN or a public land mobile network (PLMN). The SGW sets up the signaling map of CS networks and in doing so, performs protocol conversions. The Media Gateway Controller Function (MGCF) is the main node of the PSTN Gateway that performs protocol conversions and creates the mapping of SIP forward to the Bearer-Independent Call Control (BICC). MGW performs the media role of a PSTN gateway with the CS network. The service exchange platform in IMS is called the Session Description Protocol (SDP) and provides the basis of the creation, provision, control, accounting, and management of communication services up to the end user [4].

# **3. SIP PROTOCOL**

The SIP protocol is used for the commencement, management, and ending of sessions between two or more applications and is a type of client-server protocol. The use of SIP specifically for voice conversations based on Voice over Internet Protocol (VoIP) has enjoyed increasing growth, such that SIP has become the IMS signaling protocol proposed for next generation networks. The architecture of this protocol consists of two logical entities: the user agent and the server agent. In turn, user agents are divided into two groups: user agent entities and server agent entities that issue requests and answer leads. Servers are divided into several categories: registrar servers having the duty of registering users, and proxy servers that seek intended users and are responsible for routing delivery requests to the intended user [5]. SIP protocol messages include two types of request and answer. The most important SIP request codes are as follows:

- INVITE: is a request containing destination information, and when the PBX IP phone signals the call and determines the destination number and address, a message is sent to the subscriber user. In the case of an accepted call offer, the destination phone will ring.
- ACK: is a response that confirms the accurate arrival of SIP packets by using the TCP protocol.
- BYE: is a request to terminate a call that is sent by one user to another and this message consists of a notification of the absence of the disconnecting user.
- CANCEL: is a message issued by the server proxy communicating a cut in the sustain state, and issuing the BYE message. Both CANCEL and the BYE are request messages.
- OPTION: is a request asking the user or server to express its abilities.
- REGISTER: is a message communicating information from client to server in order to introduce the client to the server. This message includes information such as IP, OPTION, and PORT.
- 200 OK: is a confirmation of successful SIP requests.

#### 4. IMS NETWORK SECURITY

Since the IMS architecture is open and IPbased, the hacker's access for performing the attack on this network is unimpeded. As a presenting technology, certainly, the SIP standard will be a target for attacks and, as a result, IMS technology will almost always inherit these problems. Based on the 3GPP series of technical specifications, the IMS framework offers many security specifications such as authentication and cryptography. However, it does not support a method for the protection of IMS resources against flooding attacks.[6] In fact, further attempts have been made to control access and encrypt communications, but new threats such as Denial of Service and spamming are not mentioned. As targets of flooding attacks, network resources such as memory, processors, and bandwidth are being consumed. As a result, legal users are either prevented from receiving services, or their performance is minimized [1].

#### 4.1. Flooding Attacks on the IMS

Key entities in the IMS architecture include the call session control function (CSCF) and the home subscriber server (HSS). The Proxy-CSCF (P-CSCF) is the first element of the IMS that users' calls pass through and all input and output IMS calling passes through this entity. In other words, flooding attacks on the IMS almost always target the SIP protocol, therefore in the completed research about flooding attacks on the IMS, studies usually concentrated on the performance of the SIP protocol through the P-CSCF. The registration process in the IMS is begun with the registration of a user's message to the P-CSCF, which is then sent to the I-CSCF and S-CSCF.

The denial of service attack can occur in several ways [6]. The types of attacks include the following:

The flooding attack: this attack targets the resource server (CPU, line capacity or memory).

The abuse attack: this is a hacker's use of a SIP-modified message for the abolition or change in direction of calls, or abuse of service. These attacks typically affect small groups of users.

The unintentional attack: this attack occurs when an attacker targets support services (DNS, call payment services and others) for destroying or restricting service.

The effect of The DoS attack depends on its target. Targeting a specific user only causes the denial of service to one user, but when a SIP server is targeted, no user can access VoIP services. In recent years, due to an increase in the occurrence, effect and complexity of DoS attacks, these attacks have been recognized as presenting a severe problem. However, intentional and unintentional attacks should be distinguished from each other. The effects of intentional attacks include CPU depletion, memory depletion, bandwidth depletion, and the abuse of amplification, buffer overflow and protocols. Unintentional attacks usually result in implementation errors, population congestion, and incorrect configurations. The flooding attacks on the IMS are performed through the distributed denial of service attack or by using the sending of excessive SIP messages.

The distributed denial of service attack is performed by using a combination of elementary attacks such as flooding attacks, IP flooding attacks, smurf attacks, SYN attacks and others [7]. As the SIP flooding attack is easy to achieve, it is in the category of the most frequently occurring attacks [8]. Types of SIP flooding attacks can be divided into SIP flooding REGISTER attacks, SIP flooding INVITE attacks, and the flooding of other SIP messages [9]. In the case of flooding attack, the attacker sends a vast mass of one or more of SIP messages in order to consume server resources. The depletion of server resources by these messages depends on the configuration of the server and its processing of each message. As the attacker increases the number of offending messages, it causes more failed calls and missed packets so that the effect on the server disorder grows exponentially.

# **4.2. Flooding Attack Detection Using the Kullback-Leibler Divergence**

In the presented algorithm for detection, the initial SIP traffic rate is captured for a period of T0 seconds. This is the "training phase". Then, the "testing phase" begins and lasts for another T0 seconds. In this training and testing phase's process, the number of SIP messages, such as INVITE, 200 OK, ACK, BYE, and others, are accounted in the nth time slot. P is the probability distribution in the training phase, and Q is the probability distribution during the next seconds (the testing phase). The difference between these probabilities is then counted using the Kullback-Leibler divergence. If this difference, called the Kullback-Leibler distance, becomes greater than the obtained threshold, it means that an attack occurred. Figure 1 shows the flowchart of this algorithm. The Kullback Leibler distance between these two phases is counted by relation (1) as below, where  $D_{KL}$  is the Kullback-Leibler distance between two probability distributions. . The Kullbackand Leibler divergence is asymmetric, meaning that:  $D_{KL}(P||Q) \neq D_{KL}(Q||P)$ . Accordingly, the following relationship is used for detection:

$$D_{KL} = D_{KL}(P||Q) + D_{KL}(Q||P)$$
  
=  $\sum_{i=1}^{n} p_i \ln \frac{p_i}{q_i} + \sum_{i=1}^{n} q_i \ln \frac{q_i}{p_i}$  (1)

In the mathematical method presented for detection, one parameter is often considered to be the dividing line between a normal and an abnormal state. This parameter's measurement can then be used to monitor system performance. Normally, the measurement should not become more or less than a certain limit, which limit being the detection threshold. There are two types of thresholds: the fixed threshold and the adaptive threshold. Due to changes in network traffic that normally occur and which cause corresponding changes to Kullback-Leibler distances over time, attempts to obtain a fixed threshold are neither effective nor practical. Accordingly, an adaptive threshold is more suitable. Adaptive or dynamic thresholds are calculated based on an estimate of the parameter measurements at the current and previous times, using estimation theory. In this algorithm, the threshold is calculated using a modified exponentially weighted moving average (EWMA) known as a triple exponential moving average (TEMA). The basis of using the EWMA is the prediction of a parameter by its current and previous values. Relationships (4) through (9) show how to calculate an adaptive threshold.

$$D_{KL}(P||Q) = \sum_{i=1}^{n} p_i ln \frac{p_i}{q_i}$$
(2)

$$D_{KL}(\varrho \| P) = \sum_{i=1}^{n} q_i ln \frac{q_i}{p_i}$$
(3)

$$s_t = \alpha \frac{x_t}{c_{t-L}} + (1-\alpha) (s_{t-1} + b_{t-1})$$

$$\tag{4}$$

$$c_t = \gamma \frac{x_t}{s_t} + (1 - \gamma) c_{t-L} \tag{5}$$

$$b_t = \beta (s_t - s_{t-1}) + (I - \beta) b_{t-1}$$
(6)

$$F_{t+m} = (s_t + mb_t)_{C_{t-L}} + 1 + (m-1)$$
(7)

When considering the amount of x in time t + 1, the relationship is as follows:

#### 4.3. Adaptive Threshold



Fig. 1. Flooding attack detection algorithm.



Fig. 2. IMS network.

$$F_{t+1} = (s_t + b_t)c_{t-L} + 1$$
(8)

Therefore, the adaptive threshold to be gained through relationship (9) is:

$$KLD_{t+1}^{th} = (s_t + b_t)c_{t-1} + \varepsilon \,\delta_t$$

$$= KLD_{t+1}^{est} + M$$
(9)

For detection, the threshold should be more than the normal distance value. Accordingly,  $\lambda$  and  $\varepsilon$  are applied in order to set a safe margin for the threshold value. In general, the  $\alpha$ ,  $\beta$ ,  $\varepsilon$  and  $\gamma$  parameters can be adjusted to achieve a high rate of accuracy. In relationship (9), M is a parameter for increasing the estimated threshold distance over the actual distance.

- $x_t$  = Kullback-Leibler distance value at t time
- $\delta_t$  = Difference between the Kullback-Leibler distances at the final time with the current distance.
- *b<sub>t</sub>* = Estimation of the Kullback-Leibler distance at *t* time.
- c<sub>t</sub> = A sequence of correlated coefficients at t time.
- *s<sub>t</sub>* = The estimation of the Constant part of the Kullback-Leibler sequence at *t* time.
- *ε*, *M* = The parameters used to regulate the threshold.
- $\alpha, \beta, \lambda$  = The parameters used to estimate the threshold.

# **5. SIMULATION TESTBED**

For evaluating the methods of flooding attack detection, a testbed is needed to produce SIP traffic. In implementing the SIP protocol, there are two users: one being the caller, and another the recipient of the call. The central agent is the SIP server. The attacker can send messages to the server as another, third-party user. Figure 2 shows the topology of this network. However, the possibility exists of sending SIP messages from the user agent to



Fig. 3. Simulation testbed.

Response code	The means of transmitted response code	Example
1XX	Received	180: means ringing
	message is in	phone, 100 means try
	processing	for establishing call
2 <i>XX</i>	Request	2000K: successful end
	completed	of request. If user waits
		for response and doesn't
		receive 2000K, will
3 <i>XX</i>	Depresentation	resend this message.
344	Representation of exchange	300: movement of op- posite user
4 <i>XX</i>	Request has	407: indicating the re-
444	encountered	transmission of request
	with an error	with authentication
	and need modi-	header, 408: constitutes
	fication	expiration of request.
5 <i>XX</i>	As reasons	500: indicative of inter-
	server is unable	nal error server
	to	
	complementa-	
	tion of a vali-	
	date	
	request	
6 <i>XX</i>	General error of	
	system	

the server agent and, since in this work the goals are to observe the SIP traffic, test the proposed algorithm for attack detection and attack facility in traffic, tests have used direct connections between two users. Figure 3 presents this scenario. In performing an experiment after installing SIPp software on two Windows 7 operating systems, defining one of them as a user agent and the other as a server agent, and setting up Kali Linux software on a Debian Linux operation system as attack platform, the SIP traffic is produced by SIPp software and its default scenarios.

# 6. SIMULATION RESULTS

After the implementation of a simulation testbed, the production of SIP traffic, and the creation of an attack on normal traffic, the performance of the presented algorithm in various network situations was evaluated and the results are presented in the next section. In an experiment scenario, all time slots are deemed to be 300 seconds and attacks are created by using INVITE messages over 150 seconds. As observed in figures 4 through 12, when an attack occurred, the Kullback-Leibler distance suddenly increased, and this quick change of distance caused it to pass the adaptive threshold. Detection method performance is highly dependent on threshold estimation and, indeed, threshold estimation parameters are adjusted to set up a compromise between false alarm rates (accuracy) and detection percentages (precision).

# 6.1. Evaluation of Detection Method Performance with Various Attack Rates in Network Traffic

In evaluating the performance of detection methods in various situations of network traffic and attack ratios, normal network traffic was considered in three ratios: 10 (low), 50 (moderate), and 100 (high). Then, three INVITE flooding attacks were launched against this traffic, with masses of 10000 (low), 100000 (moderate), and 1000000 (high). The results of these experiments are shown in Fig. 4 to 9, and are described as follows:

1- If the threshold estimate is near the actual value, the attacks with low ratios are



Fig. 4. Attack detection with traffic rate of 10 calls per second and attack mass of 10000 INVITE messages.



Fig. 5. Attack detection with traffic rate of 10 calls per second and attack mass of 100000 INVITE messages.



Fig. 6. Attack detection with traffic rate of 10 calls per second and attack mass of 1000000 INVITE messages.



Fig. 7. Attack detection with traffic rate of 50 calls per second and attack mass of 10000 INVITE messages.



Fig. 8. Attack detection with traffic rate of 50 calls per second and attack mass of 100000 INVITE messages.



Fig. 9. Attack detection with traffic rate of 50 calls per second and attack mass of 1000000 INVITE messages.



Fig. 10. Attack detection with traffic rate of 100 calls per second and attack mass of 10000 INVITE messages.



Fig. 11. Attack detection with traffic rate of 100 calls per second and attack mass of 100000 INVITE messages.



Fig. 12. Attack detection with traffic rate of 100 calls per second and attack mass of 100000 INVITE messages.

detected but the false alarm rate increases. In other words, legal use of the network is falsely flagged as an attack. In contrast, an increase in the adaptive threshold allows lowlevel attacks to pass undetected. Therefore, a threshold should be obtained that carefully balances between detection accuracy (the false alarm rate) and detection precision. The combination of these two parameters with the speed of detection is considered to be the performance evaluation criterion.

2- The selection of time intervals is important in detection because this method depends on changes in traffic. Accordingly, the time interval should be set in such a way that it shows the timing of the attack. In addition, the choice of time intervals greatly affects detection speed.

3- There are two disadvantages in this method: first, if an attacker slowly increases the rate of his attack, since the adjustment of the threshold is based on traffic, the threshold

will remain higher than the traffic and the attack will not be detected. Second, although the volume of traffic will exceed the threshold after an attack is carried out and an alarm will be issued, due to the threshold being estimated based on traffic, after an elementary attack a pathway for carrying out a low-level attack will remain open.

# 6.2. Evaluating the Effect of Detection Parameters on Detection Method Performance

In evaluating the effect of detection parameters on detection method performance, first, the normal network traffic with rates of 50 calls (300 messages) per second and attack traffic with masses of 100000 INVITE messages were considered. Then, after changing the detection parameters, the performance was evaluated with results that are presented in Table 2.

Parameter	Decreasing up to min	Opti- mum	Increasing up to max
$\beta_k$	Decrease threshold level before and after attack and bring to fixed value	1	Increased the threshold level and if it be- comes pretty high, after attack threshold will be ineffective
$L_k(1)$	Threshold level not to be lower than certain value	0.001	Raise the threshold level and if be pretty high, threshold before attack will be very high and after attack negative
$T_k(1)$	Threshold level decreased and false error rate to be increased	0.001	Raise threshold level before attack
$S_k(1)$	Threshold level to be decreased	1	Threshold level increased and as be more, threshold level before attack fixed and after attack will be raise
a <sub>k</sub>	Threshold level increased and caused to ineffectiveness of detection	0.1	Reduce threshold level and bring to nega- tive value then false alarm rate become 100%
b <sub>k</sub>	Increased the threshold level and bring to fixed value	0.05	Raise threshold level before attack and no change after attack
g <sub>k</sub>	Doesn't change the threshold level before attack and decreased after attack but not bring to zero	0.05	Raise the threshold after attack and no change before attack

Table 2. Effect of parameters change of performance of detection method.

#### 7. CONCLUSION

In light of the fact that next-generation networks are moving to increase accessibility and improve the quality of user communications, and that IMS networks are best able to respond to users' current requirements, it is clear that the IMS network is the most suitable platform for providing service to the next generation of networks. On the other hand, with the increasing use of internet technology, the IMS network has also become vulnerable to various types of attacks. Therefore, the maintenance of security is very important. The flooding attack is one of the attacks that has often troubled communications networks and, with increasing access to a new generation of networks; its risk of occurrence is increasing. Various methods have been presented for resisting this type of attack, and in this study, its detection by use of the Kullback-Leibler divergence is presented. Adaptive thresholds were also estimated using the triple exponential moving average (TEMA), which offers better estimation accuracy than previous methods and successfully resists problems during disturbances in network traffic.

Based on the completed simulations, it is clear that, when network traffic rates are low, attacks with high rates are easily detected and will not have much effect on the network.

When the normal traffic rates are high, however, an attack will tie up network traffic much more easily. This means that as network use continues to increase, if unchecked, attacks will have a greater and greater effect on the network. Although high and low rates of normal traffic and attacks don't have a considerable effect on the presented detection method, sampling time intervals and estimation and adjusting coefficients that are appropriate to network traffic should obtained. The simulation results and evaluation of the performance of the presented algorithm in various situations show the efficiency of this method as well.

# REFERENCES

- [1] Yacine Rebahi, Muhammad Sher, Thomas Magedanz, "Detecting Flooding Attacks Against IP Multimedia Subsystem (IMS) Networks", IEEE AICCSA, 2008.
- [2] Z. Chen, W. Wen, D. Yu, "Detecting SIP Flooding Attacks on IP Multimedia Subsystem (IMS)", International Conference on Computing, Networking and Communications (ICNC), 2012.
- [3] Rogier Noldus, Ulf Olsson, Catherine Mulligan, Ioannis Fikouras, Anders Ryde, Mats StilleIMS, "Application Developer's Handbook Creating and Deploying Innovative IMS Applications", Academic Press is an imprint of Elsevier, chapter 8, First published 2011.
- [4] Elham Nosrati, Evaluation and Simulation of the Security Threats of IMS Network and Propose the Methods for Detection and Prevention, Islamic Azad University South Tehran Branch, 2010.
- [5] Mojtaba Jahanbakhsh, Sayed Vahid Azhari, Maryam Homayouni, Ahmad Akbari, "Evaluating the Various Configurations of SIP Signaling Network by Using of Measuring the Quality calls Parameters", Journal of Information and Communication Technology, First year, number 1 and 2, autumn and winter 2008.
- [6] N. Chaisamran, T. Okuda, S. Yamaguchi, "Using a Trust Model to Reduce False Positives of SIP Flooding Attack Detection in IMS", IEEE 37th Annual Computer Software and Applications Conference Workshops, 2013.
- [7] M. Voznak, J. Safarik, "DoS Attacks Targeting SIP Server and Improve-

ments of Robustness", International Journal of Mathematics and Computers in Simulation, 2012.

- [8] Elham Nosrati, Aazar Saadaat Kashi, Yashar Darabian, S. Navid Hashemi Tonekaboni, "Register Flooding Attacks Detection in IP Multimedia Subsystems by Using Adaptive z-score CUSUM Algorithm", Proceeding of the 5th International Conference on IT and Multimedia at UNITEN (ICIMU 2011) Malaysia, 2011.
- [9] Jin Tang, Yu Cheng, Yong Hao, "Detection and Prevention of SIP Flooding Attacks in Voice over IP Networks", Proceeding IEEE INFOCOM, 2012.