Hybrid Scaling Based Dynamic Time Warping for Detection of Low-rate TCP Attacks

Won-Ho So* ***Lifelong Member***, Kyoung-Min Yoo** ***Regular Member***,
Young-Chon Kim** ***Lifelong Member***

ABSTRACT

In this paper, a Hybrid Scaling based DTW (HS-DTW) mechanism is proposed for detection of periodic shrew TCP attacks. A low-rate TCP attack which is a type of shrew DoS (Denial of Service) attacks, was reported recently, but it is difficult to detect the attack using previous flooding DoS detection mechanisms. A pattern matching method with DTW (Dynamic Time Warping) as a type of defense mechanisms was shown to be reasonable method of detecting and defending against a periodic low-rate TCP attack in an input traffic link. This method, however, has the problem that a legitimate link may be misidentified as an attack link, if the threshold of the DTW value is not reasonable. In order to effectively discriminate between attack traffic and legitimate traffic, the difference between their DTW values should be large as possible. To increase the difference, we analyze a critical problem with a previous algorithm and introduce a scaling method that increases the difference between DTW values. Four kinds of scaling methods are considered and the standard deviation of the sampling data is adopted. We can select an appropriate scaling scheme according to the standard deviation of an input signal. This is why the HS-DTW increases the difference between DTW values of legitimate and attack traffic. The result is that the determination of the threshold value for discrimination is easier and the probability of mistaking legitimate traffic for an attack is dramatically reduced.

Key Words : Low-rate TCP attack, Dynamic Time Warping, Denial of Service, Network Security.

Ⅰ. Introduction

The Internet has been required to provide various services and to evolve into broadband networks due to dramatic increases in the number of users and multimedia traffic volume. This impels network service providers to ensure that their network services converge on broadband IP based Internet. Specially, BcN (Broadband Convergence Network) related technologies have been introduced, focusing on a variety of research topics. In order to efficiently construct BcN as a next-generation network, requirements such as the quality of service, network survivability, SLA (Service Level Agreement), and network security must be satisfied. Network security will be an important component of BcN in the near future [1].

Denial of Service (DoS) attacks consumes resources in networks, server clusters, or end hosts. The malicious objective of these attacks is to prevent or severely degrade service to legitimate users. Examples of DoS attacks include TCP SYN attacks, ICMP flooding, and DNS flood attacks. These generate high volumes of traffic, similar to directing a sledge-hammer at a target victim. Therefore, a DoS attacker can often be detected by analyzing the network traffic[2,3].

Recently, low-rate TCP attacks called "shrew

※ 본 연구는 정통부 및 정보통신연구진흥원의 지원을 받아 수행된 연구결과임, <08-기반-13, 정보통신연구기반조성사업>
* 순천대학교 컴퓨터교육과 컴퓨터네트워크연구실 (whso@sunchon.ac.kr)
** 전북대학교 전자정보공학부 차세대통신망연구실 ({mini7029, yckim}@chonbuk.ac.kr)
논문번호 : KICS2007-08-376, 접수일자 : 2007년 8월 24일, 최종논문접수일자 : 2008년 6월 11일

attacks," was introduced in[4]. Thses attacks attempt to deny bandwidth to TCP traffic flows while occurring at a sufficiently low average rate to elude detection by counter-DoS mechanisms. Thus, a new detection mechanism is required to defend against shrew attacks.

In this paper, a mechanism for detecting low-rate TCP attacks is considered as a rapid link based detection mechanism. A DTW based detection mechanism is evaluated. This has a problem which makes it difficult to decide whether an input sample is legitimate or malicious. The reasons for the problem were found. Scaling by dividing sample data by maximum auto-correllation value increases the detection rate of attacks. Thus, the proposed Hybrid Scaling based DTW mechanism uses the scaling and considers the standard deviation of the input sample to efficiently discriminate between attacks and legitimate traffic flows.

This paper is organized as follows. In Section Ⅱ, we show the TCP retransmission timer behavior, low-rate TCP attacks, and previous defense mechanisms. We proceed with an analysis of a previous pattern matching mechanism with DTW and we show the critical problem with it in Section Ⅲ. In Section Ⅳ, the proposed mechanism based on hybrid scaling and standard deviation is introduced. Finally, we draw conclusions and suggest future work in Section Ⅴ.

Ⅱ. Low-rate TCP Attack and Defense

2.1 TCP Timeout Mechanism

TCP Reno uses the following RTO mechanism for congestion control in the Internet, which is exploited by low-rate TCP DoS attacks. Packet loss is detected via either timeout from non-received ACKs, or by receiving a triple-duplicate ACK. If there is packet loss and less than three duplicate ACKs are received, a TCP agent delays until the period of retransmission timeout, reduces its congestion window to one packet, and resends the packet. Therefore, the performance of the TCP connection decreases in terms of throughput, due to this delay. Allman and Paxson experimentally showed that TCP achieves near-maximal throughput if the lower bound for RTO is one second [4]. Figure 1 shows the behavior of the TCP retransmission timer.

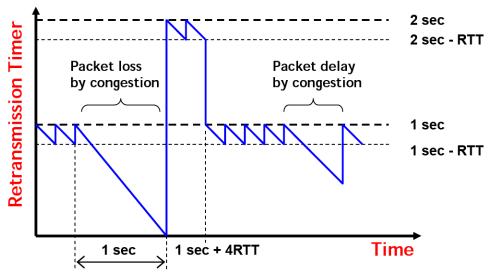


Fig. 1. Behavior of the TCP retransmission timer

2.2 Low-rate TCP Attack Model

As described in the previous section, packet loss with retransmission timeout sets the congestion window to 1 and induces the degradation of throughput. For example, consider a single TCP traffic flow and a single DoS stream. Assume that an adversary creates an initial outage at time 2 RTT via a high-rate burst of short duration. Due to this dramatic packet loss, the TCP sender waits for a retransmission timer of 1 sec, then, double its RTO. If another attack is created between time 1 and 1+4RTT, the TCP sender must delay another 2 sec.

Fig. 2 shows the function $f(T, l, R, S, N)$ of a "square wave" shrew attack, which transmits bursts of duration l and rate R in a deterministic on-off pattern with period T. Generally, a successful shrew attack has a rate R large enough to induce packet loss (i.e., R aggregated with existing traffic must exceed the link capacity). In addition, l must be long enough to induce timeout but short enough to avoid detection. Finally, T is set to the minimum RTO, which causes the timeout of legitimate traffic flows and their packet loss.

2.3 Related Defense Mechanisms

A defense to this "shrew wave" attack is to randomize RTO. Here, information can still be transmitted while the attacker is waiting and a connection can avoid successive time outs.

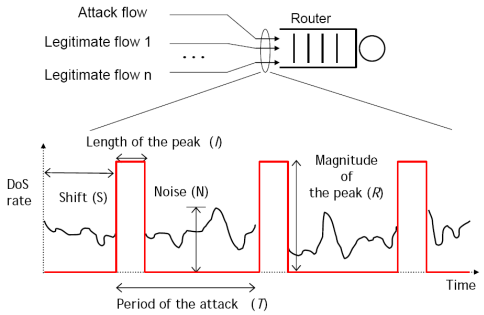


Fig. 2. Square-wave of low-rate TCP attack

Randomizing a fixed minimum RTO is an immediate solution, but the main issue is whether such an approach should be adopted. Randomizing also reduces TCP connection performance in the absense of an attack[5].

In [6], a novel scheme is proposed that does not introduce any modification to the TCP congestion control mechanism and can be implemented at a single edge router. This approach considers each arrival times of packets at the edge router. The malicious traffic flow detection sub-module of the object module computes the time difference between consecutive packets of each traffic flow. If the average time difference for the attack traffic flow is repeated periodically and its burst length is greater than or equal to the RTTs of other traffic flows, the attack traffic flow will be filtered at the victim's edge router. This scheme, however, cannot protect against other traffic flows affected by this attack at an intermediate router.

A filtering shrew mechanism identifies and detects attacks by examining the frequency domain characteristics of incoming traffic flows to a server[7]. Although it is effective in terms of detection time and traffic flow based detection, it requires substantial quantities of traffic flow information to monitor and detect the frequency of each traffic flow.

A distributed detection mechanism which uses the dynamic time warping (DTW) method is adopted, to robustly and accurately identify this type of attack [8]. Once a router detects the attack, it uses a fair resource allocation mechanism called DRR (Deficit Round-Robin) to minimize the number of affected TCP traffic flows, and provide sufficient resource protection for the affected TCP traffic flows. DTW in this approach, however, can misidentify legitimate traffic flows as attacks, due to the limitations of this algorithm. If the problem can be overcome this approach can be used over the network to provide a coordinated defense against attacks such as shrew attacks. Therefore, an analysis of the problem is required, and an advanced modification or approach must be designed, as shown in the following sections (Table 1).

Ⅲ. Drawback of Simple DTW Approach

The following subsections describe the DTW defense mechanism and evaluate it to analyze the reasons for mistaking legitimate traffic flows as attacks.

3.1 Evaluation of DTW Approach

In order to evaluate the previous DTW approach, we create and use an attack pattern template (APT), two kinds of attacks; strictly

Table 1. Comparison of Previous Mechanisms

Term \ References	[5]	[6]	[7]	[8]
Detection	×	Packet arrival	Periodic traffic flow	Periodic link flow
Determination	×	Arrival interval	Frequency analysis	DTW
Defense	×	Traffic flow filtering	Traffic flow filtering	DRR*
Additional Process	Random RTO	×	×	Trace back
Implemented Point	All TCP agents	Single router	Single router	All routers

* DRR means deficit round robin.

periodic square burst (SPSB) and random periodic general burst (RPGB), and one legitimate traffic flow (LEGI). APT is defined as a function *f*(1.2sec, 0.2sec, 1, 0, 0), as shown in Figure 2. In this case, monitoring shift is zero and the noise is also zero, because APT is the base template. SPSB is a strictly periodic signal with a single burst of length l and a period T. The initial values of l and T are randomly selected and identical values are adopted for the subsequent period. RPGB is a sine wave with a period of T which includes random noise N. The values of T and N are drawn from uniform distributions and these values may be vary from one period to the next period. The burst length l, period T, and background noise N are uniformly distributed within [0, 0.5], [1, 1.5], and [0, 0.5], respectively. The time shift is also considered as a uniform distribution within [0, T] and the magnitude of the burst is set to R=1 [8]. We generate 5000 samples for each of two kinds of attacks and legitimate traffic.

The detection mechanism proposed in [8] involves four steps; statistical sampling of incoming traffic, noise filtering, feature extraction, and signatures comparison. Incoming traffic is sampled and normalized based on the transmission capacity of the link in the *statistical sampling* step. To perform *noise filtering* prior to the feature extraction, the non-active period of the low-rate attack must be set to zero. Auto-correlation is used for the *feature extraction* and is calculated via unbiased internal normalization. Consider an input signal with n values (x_0, x_1, ⋯ , x_{n-1}) and all other x_i=0. The unbiased normalized auto-correlation $A(k)$ can be calculated as follows:

$$A(k) = \frac{1}{n-k} \sum_{i=0}^{n-k+1} x_{i+k} x_i \quad k = 0, \cdots, n-1. \qquad (1)$$

In unbiased normalized auto-correlation, we determine the period of the input signal and the auto-correlation plot is independent of the time shift value ***S***. In final *signature comparison*, dynamic time warping (DTW) is adopted and is a robust and computationally efficient method for comparing the degree of similarity between a template signature ***S*** and an input signal ***I*** [9]. The lower the value of $\boldsymbol{DTW^*(S, I)}$, the higher the degree of similarity between input string ***I*** and signature ***S***. Dynamic programming can help to calculate the minimum cost warping path in the DTW algorithm.

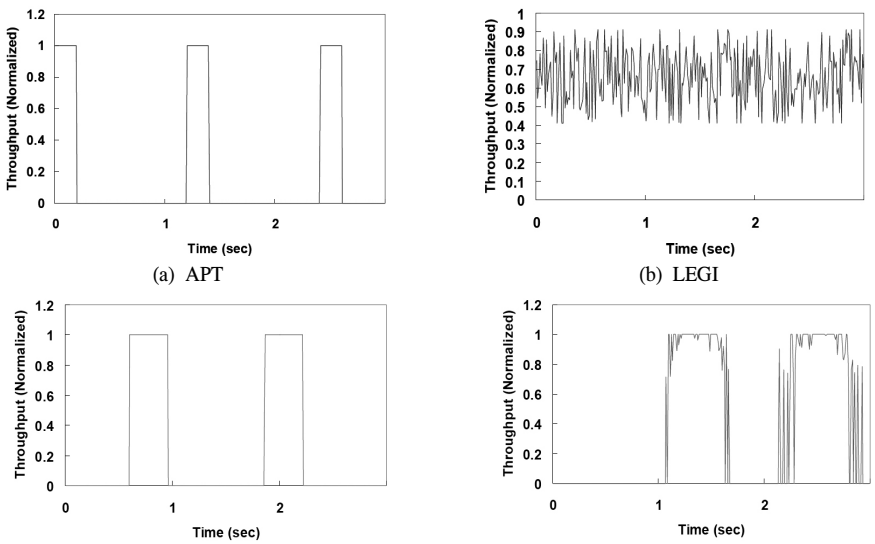


Fig. 3. Generated traffic patterns for evaluation

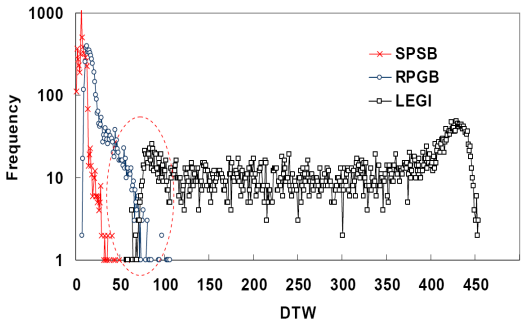


Fig. 4. Probability density functions of DTW values

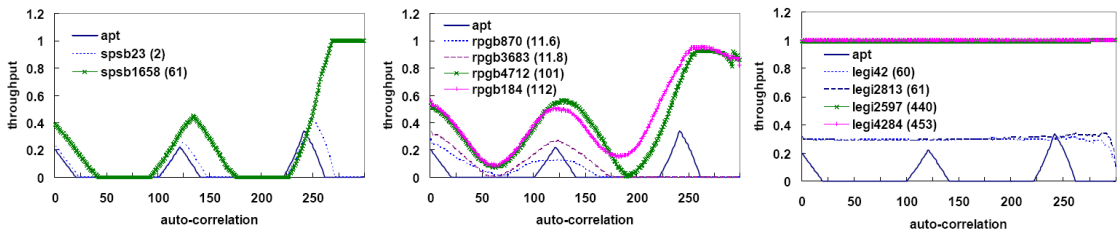
Based on our prior assumption, we evaluate the previous mechanism. Figure 4 illustrates the probability density function of DTW values for attacks and legitimate traffic, which consists of a major constant throughput with some Gaussian noise. The figure shows that the minimum DTW value for legitimate traffic is less than the maximum DTW value for attack traffic.

Therefore, some false positives and false negatives may occur during detection. In the case of SPSB, most DTW values are distributed below 50 and it is clear that the SPSB type attack is easily detected. We also find that the results for RPGB are similar to the results for SPSB. However, some RPGB attacks are calculated to have a maximum DTW value of 112. For LEGI, DTW values are between 60 and 453 and are widely distributed. Thus almost LEGI traffic flows can pass by the edge router because almost DTW values for LEGI are relatively greater than the maximum DTW value for attack patterns.

There are, however, false positives and false negatives during detection in this DTW approach. For instance, DTW values between 50 and 120 can be used as a threshold to filter attacks and legitimate traffic. This depends on the means of determining a threshold for differentiating between legitimate and attack traffic. Although there may be varied results according to the generation of Gaussian traffic or real-world incoming traffic, and we should investigate this in future research, the problem of the overlap between the maximum value of an attack and the minimum value for legitimate traffic should be effectively overcome.

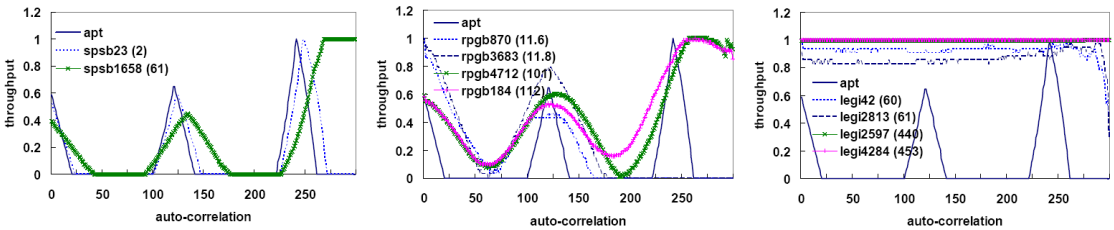
We analyze the problem of DTW values overlapping by comparing the results of auto-correlation for each input signal. Figure 5 describes the differences between auto-correlation values among two attacks and legitimate traffic. Figure 5(a) includes SPSB sample files numbered 23 (spsb23) and1658 (spsb1658) and their minimum and maximum DTW values are 2 and 61, respectively. In the cases of RPGB and LEGI as depicted in Figure 5(b) and 5(c), we select two sample files for each case, and perform a simultaneous comparison. A comparison of rpgb4712(101) (which means the DTW value is 101), rpgb184(112), legi42(60), and legi2813(61) is performed. The RPGB samples are morphologically similar to the attack pattern template, but their DTW values are sufficiently high for the detector to regard them as legitimate traffic. The cases of LEGI samples are contrary to the RPGB cases. Therefore, they can be filtered at the detection router when they go by it.

A mechanism may be required to increase the similarity between the auto-correlation of attack traffic and APT, and to decrease it between legitimate traffic and APT.



(a) APT vs. SPSB (b) APT vs. RPGB (c) APT vs. LEGI

Fig. 5. Comparison of auto-correlation values among different input signals



(a) APT vs. SPSB (b) APT vs. RPGB (c) APT vs. LEGI

Fig. 6. Comparison of auto-correlation values among different input signals with simple scaling

3.2 Scaling Auto-correlation Values

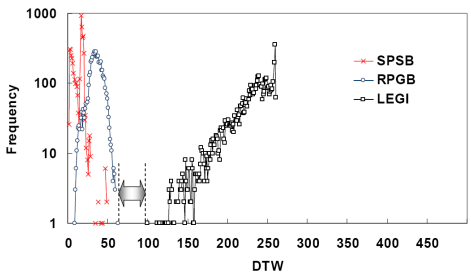
As previously described, the overlapping of probability density functions induces a misidentification between legitimate and attack traffic. This is unacceptable. Thus, we propose scaling the auto-correlation value, in which $A(k)$ is divided by the maximum $A(k)$. Then, the DTW value is calculated. Firstly, we adopt scaling for APT and all samples of input signals. Figure 6 shows the effect of scaling after computing the auto-correlation of ATP, SPSB, RPGB, and LEGI. Morphologically, the similarity between APT and SPSB, or between APT and RPGB is shown in Figure 6(a) and 6(b) respectively. However, the similarity between APT and LEGI in Figure 6(c) decreases. Therefore, we can expect that the DTW value of LEGI in Figure 6(c) is higher than one of LEGI without scaling.

Figure 7 depicts the probability density function and cumulative density function of DTW processing with scaling for each input sample. Note that the minimum DTW value for legitimate traffic is higher than the maximum DTW value for attack traffic. The maximum DTW values of SPSB and RPGB are 51 and 68, respectively, and the minimum DTW value of LEGI is changed from 60 to 104. Therefore, we can easily select a threshold DTW value between the maximum DTW value for attack and the minimum one for LEGI. In [10], the difference in the range is 36.

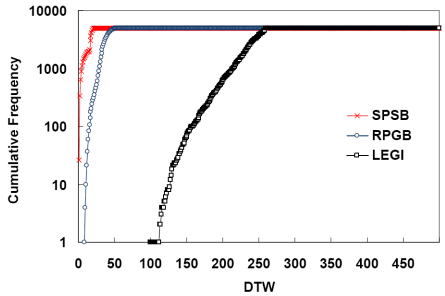
However, as mentioned before, this significant difference in the ranges of DTW values between attack and legitimate traffic might be reduced or overlapped according to the characteristics of the incoming signal. In addition, scaling for both APT and each input signal produces side-effects such as the fact that there is an increase in the DTW values for some attack samples, which means that there is the probability of misidentification of attack as legitimate traffic. Thus, the combination of scaling input traffic and APT must be considered, and other factors can be considered.

Ⅳ. Hybrid Scaling based Dynamic Time Warping

In this section, we compare the results of different scaling combinations and find another factor to significantly differentiate legitimate traffic from attacks.



(a) Probability density functions



(b) Cumulative density function

Fig. 7. Results of DTW with scaling

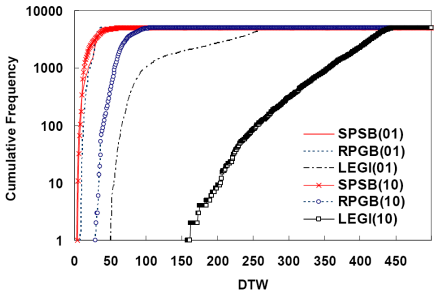
4.1 Relationship between Standard Deviation and Hybrid Scaling

The experiment on various scaling methods should consider combinations of scaling of the auto-correlation of APT and scaling of the auto-correlations of input signals SPSB, RPGB, or LEGI. Four kinds of combinations are considered; non-scaling for APT and non-scaling for input signal (SC00), scaling for APT and non-scaling for input signal (SC01), non-scaling for APT and scaling for input signal (SC10), and scaling for APT and scaling for input signal (SC11). The first and the fourth combinations were previously considered in prior subsections, and the second and the third combinations are evaluated via simulation.

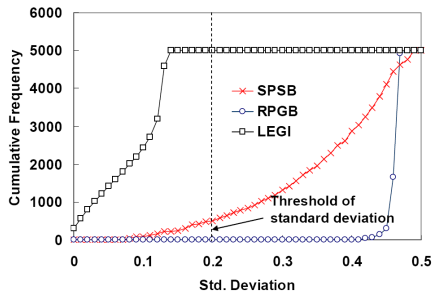
The results are shown in Figure 8(a). Based on the results, we note two important characteristics. First, SPSB is independent of scaling difference, thus, SC01 and SC10 can be used to detect SPSB type attacks effectively. Second, in the case of RPGB, we can easily detect these attacks for SC01 scaling. LEGI, however, shows contrary results SC10 makes the DTW values for LEGI. This means that legitimate traffic almost passes through alow-rate attack detection router if we use SC10 scaling. In summary, SC01 is better than SC10 for detecting two kinds of attacks effectively. For legitimate traffic, SC01 produces a more reasonable result than SC10. Although SC01 and SC10 can also be used for a detection mechanism with a difference in range, there is the problem that both SC01 and SC10 scaling also have a side-effect similar to SC11, as previously described.

Therefore, we use the standard deviation for each input and attack pattern template. From the morphological perspective, while the sample values for SPSB and RPGB are widely distributed, LEGI has a less narrow distribution. Consider an input signal with n values (x_0, x_1, ⋯ , x_{n-1}). The standard deviation can be calculated as follows:

$$\sigma = \sqrt{\left(\sum_{k=0}^{n} x_k^2 + \left(\sum_{k=0}^{n} x_k\right)^2 / n\right) / n}, \quad k = 0, \cdots, n-1. \quad (2)$$



(a) Cumulative density function for SC01 and SC10



(b) Standard deviation for each input signal

Fig. 8. Standard deviation and DTW values for two scaling types

Fig. 8(b) shows the results of the standard deviation for each input signal. We note that the value for LEGI is relatively low, and the value for RPGB is high. However, in the case of SPSB, the value for the standard deviation are uniformly distributed between 0.1 and 0.5. Therefore, we can distinguish between attacks and legitimate traffic if we set the threshold of the standard deviation δ as 0.2. Less than 10% of SPSB input signals are below δ.

Based on the results of Figure 8(a) and 8(b), we note that an input signal is legitimate traffic if the standard deviation is less than δ, thus, scaling of SC10 is adopted, in order to increase the DTW values. If the standard deviation of the input signal is greater than or equal to δ, scaling of SC01 is better, due to the high probability of detecting an input signal which is an attack.

The hybrid scaling based DTW detection mechanism is introduced in Figure 9. This mechanism includes computing the standard deviation and reasonable scaling according to the standard deviation threshold. Scaling of SC10 uses the equation, $s' = s / A_{S\_max}$, if σ is less than the

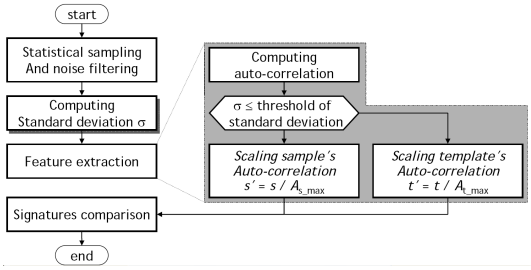


Fig. 9. Hybrid Scaling based Dynamic Time Warping Detection Mechanism

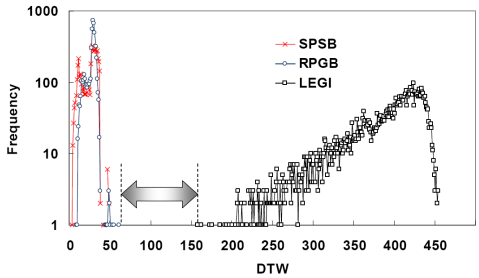
threshold δ. In the case of SC01, we use the equation $t' = t/A_{t\_max}$ prior to calculating the DTW value of the input signal for more effective detection of attacks. The former may be used in the normal network state. On the other hand, the latter can be only used when the standard deviation is greater than the threshold. Therefore, the HS-DTW mechanism does not require an edge router in scaling computation for the input signal in the normal state.

4.2 Evaluation and Discussions

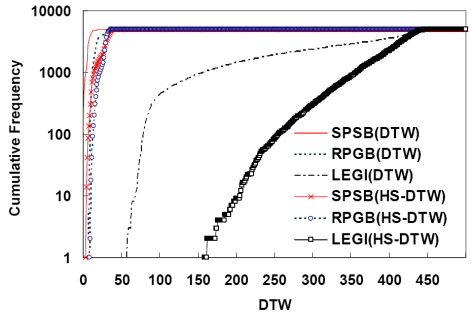
Let us consider the robustness and accuracy of using the HS-DTW method for detection of a low-rate TCP attack. The experimental configuration is identical to the previous one and the proposed mechanism shown in Figure 9 is adopted. Figure 10(a) shows the probability density function for this. We can establish that the extended difference in range is about 100, between attacks and legitimate traffic. All attacks are detectable if the threshold DTW value is set between 70 and 150. This range is critically wider than that for the original DTW mechanism. In addition, almost all DTW values for legitimate traffic are biased to high values exceeding 200.

Fig. 10(b) shows the cumulative density function for the original DTW and HS-DTW. They can be compared in terms of the difference between DTW values.

The proposed HS-DTW is flexible. It can be used according to the state of the edge router which must monitor and detect attacks. If an edge router has a heavy computational load, it can only perform standard deviation computing with a



(a) Probability density functions



(b) Cumulative density function

Fig. 10. Results of HS-DTW mechanism

computational complexity of $O(n)$. In the case of DTW, the complexity is $O(n^2)$. Thus, the HS-DTW can be considered for supporting a coordinated intrusion detection system over the Internet.

V. Conclusions

In this paper, a Hybrid Scaling based DTW (HS-DTW) mechanism is proposed for effective detection of periodic low-rate TCP attacks. In order to differentiate between attacks and legitimate traffic, the difference between their DTW values must be large as possible. Thus, after exposing the reason for the problem with the previous DTW mechanism, we introduced a scaling method which increases the difference in range of their DTW values. In addition, we considered four kinds of scaling methods and the standard deviation of sampling data. Through simulation, we showed that the proposed HS-DTW can increase the difference between DTW values for shrew attacks and legitimate more effectively than the previous approach. As a

result, we can more easily discriminate between an attack input link and a legitimate input link.

References

〔1〕 Y.S. Kim, "Technological Issues and Prospect of BcN," *Telecommunication Technology Association*, 2005

〔2〕 J. Mirkovic, J. Martin, and P. Reiher, Computer Science Department, UCLA "A Taxonomy of DDoS Attacks and DDoS Defense Mechanisms," *Technical report* #020018.

〔3〕 J. Mirkovic, "D-WARD: Source-End defense Against Distributed Denial-of-Service Attacks", Ph.D Thesis 2003.

〔4〕 A. Kuzmanovic and E. Knightly, "Low-rate TCP-targeted denial of service attacks," *In Proc. ACM SIGCOMM*, Karlsruhe, Germany, August 2003.

〔5〕 G. Yang, M. Gerla, and M. Y. Sanadidi, "Randomization: Defense against Low-Rate TCP-targeted Denial-of-Service Attacks," *In Proc. IEEE Symposium on Computers and Communications*, July 2004, pp. 345-350.

〔6〕 Shevtekar, K. Anantharam, and N. Ansari, "Low Rate TCP Denial-of-Service Attack Detection at Edge Routers," *IEEE Communications Letters*, Vol.9, No.4, April 2005.

〔7〕 Y. Chen, K. Hwang, "Collaborative detection and filtering of shrew DDoS attacks using spectral analysis," *J. Parallel Distributed Computing*, Vol.66, 2006, pp.1137-1151

〔8〕 H. Sun, J. C. S. Lui, and D. K. Y. Yau, "Defending Against Low-rate TCP Attacks: Dynamic Detection and Protection," *In Proc IEEE Conference on Network Protocols* (ICNP2004), Oct. 2004, pp. 196-205.

〔9〕 E. Keogh, "Exact indexing of dynamic time warping," *In Proceedings of the 28th VLDB Conference*, Hong Kong, China, Aug. 2002.

〔10〕 W. H. So, S. H. Shim, K. M. Yoo, B. J. Oh, Y. S. Kim, Y. C. Kim, "Scaling based Dynamic Time Warping Algorithm for the Detection of Low-rate TCP Attack," *Proceedings of IEEK Fall Conf.* 2006, Hanyang Univ., Korea, Nov. 2006.

소 원 호 (Won-Ho So) 종신회원

한국통신학회 논문지 제25권 1A호 참조

현재 순천대학교 컴퓨터교육과 조교수

<관심분야> 유·무선 MAC 및 라우팅 프로토콜 설계, QoS 및 트래픽 제어, 네트워크 보안, 임베디드시스템 기반 교육

유 경 민 (Kyoung-Min Yoo) 정회원

한국통신학회 논문지 제33권 제3호 참조

현재 전북대학교 컴퓨터공학과 박사과정

<관심분야> BcN, Network Security, AIS(Artificial Immune System) based Network Control

김 영 천 (Young-Chon Kim) 종신회원

한국통신학회 논문지 제19권 제2호 참조

현재 전북대학교 전자정보공학부 교수