The Korean Journal of Applied Statistics (2013) DOI: http://dx.doi.org/10.5351/KJAS.2013.26.2.291
26(2), 291–305

Nonparametric Detection Methods against DDoS Attack

J. L. Lee[a] · C. S. Hong[a,1]

[a]Department of Statistics, Sungkyunkwan University

(Received November 6, 2012; Revised December 5, 2012; Accepted March 21, 2013)

Abstract

Collective traffic data (BPS, PPS *etc.*) for detection against the distributed denial of service attack on network is the time sequencing big data. The algorithm to detect the change point in the big data should be accurate and exceed in detection time and detection capability. In this work, the sliding window and discretization method is used to detect the change point in the big data, and propose five nonparametric test statistics using empirical distribution functions and ranks. With various distribution functions and their parameters, the detection time and capability including the detection delay time and the detection ratio for five test methods are explored and discussed via monte carlo simulation and illustrative examples.

Keywords: Big data, change point, sliding window, discretization, detection delay time, window data.

1. 서론

인터넷의 발전과 정보화 시대의 가속화로 인해 다양한 온라인 서비스들이 생성되고 발전하고 있으며, 이러한 온라인 서비스가 기업과 국가의 중요한 경쟁력으로 부상하고 있다. 하지만 그 역기능인 사이버 공격 또한 기하급수적으로 증가하고 있어서 온라인 서비스의 경쟁력에 대한 심각한 위협이 되고 있다. 2011년 사이버공격의 수는 전년도에 비해 85% 증가하였으며, 사이버공격으로 인한 피해액은 3880억 달러(약 417조원)에 달할 정도로 심각한 문제가 되고 있는 실정이다 (Symantec, 2011).

분산서비스거부(distributed denial of service; DDoS) 공격은 적게는 수십대에서 많게는 수십만대의 좀비 PC로 하여금 특정 컴퓨터 시스템이나 네트워크에 동시에 서비스를 요청함으로써 단시간 내에 과부하를 일으켜 서비스 중단 또는 질적 저하를 유발시키는 공격이다. 특히 DDoS 공격은 불특정 다수를 대상으로 하는 것이 아니라 특정 대상을 노린 표적 공격으로써 피해가 크고 심각하다. 국내에서는 2009년 7월 7일(77 DDoS)과 2011년 3월 4일(34 DDoS)에 청와대, 국정원, 은행 및 대형 포털사이트 등의 국가 주요 전산망이 DDoS 공격에 의해 마비되는 사건이 발생하여 심각성을 인식하고 있다.

DDoS 공격은 좀비 PC에 의해 공격이 진행되므로 공격의 원점을 찾아 이를 중단시키기는 것은 거의 불가능하다. 대신 공격의 발생여부를 신속하게 탐지하여 조치하는 것이 공격 대응을 위한 가장 중요한 요소이다. 네트워크에서 발생되는 트래픽의 양이나 개수를 측정하기 위해 일반적으로 사용되는 BPS(bits

This paper was supported by 63 Research Fund, Sungkyunkwan University, 2012.
[1]Corresponding author: Professor, Department of Statistics, Sungkyunkwan University, 25-2, Sungkyunkwan-ro, Jongro-gu, Seoul 110-745, Korea. E-mail: cshong@skku.edu

per second)나 PPS(packets per second) 등은 대용량 자료(big data)이다. 이 자료에 대한 DDoS 공격을 탐지하기 위해서는 적용되는 통계 알고리즘의 계산시간과 메모리 사용공간의 효율성이 확보되어야 한다. 그리고 BPS나 PPS 등의 대용량 자료들은 이를 수집하는 네트워크의 사용환경이나 수집시간에 따라 상이한 분포적 특성을 가지고 있으므로 모수적인 방법은 적용하기 어렵다. 네트워크에서 시간이 지남에 따라 연속적으로 발생되는 대용량 자료에 적용할 수 있도록 계산 성능이 좋으면서도, 신속·정확하게 DDoS 공격을 탐지할 수 있는 통계적 방법에 대한 필요성이 제기된다.

네트워크 기반의 DDoS 공격 탐지 방법은 일반적으로 다음의 두 가지로 구분된다 (Karen과 Peter, 2007). 첫 번째는 알려져 있는 DDoS 공격들의 특징들을 분석하여 이를 데이터베이스에 등록한 후, 이러한 트래픽이 관리자가 설정해 놓은 임계값 이상 유입될 때 이를 공격으로 탐지하는 시그니쳐 기반의 탐지방식이다. 두 번째는 정상적인 상황에서의 네트워크 트래픽의 분포를 모니터링 하였다가 일정기간 유입되는 트래픽의 분포가 정상상태와 동일한지를 통계적으로 검정하여 일정한 임계치를 벗어난 경우 이를 공격으로 간주하는 이상탐지(anomaly detection) 방법이다. 기존의 시그니쳐 방식으로는 조직화, 지능화된 다양한 DDoS 공격기법을 탐지하는 데에 한계가 있으나, 이상탐지 방식은 분포의 변화점 탐지를 통해 새롭게 등장하는 DDoS 공격까지도 탐지할 수 있으므로 적절한 방법이다.

DDoS 공격탐지 방법에 대한 연구를 몇 가지로 구분하면 다음과 같다 (Carl 등, 2006). 첫째, 유입되는 트래픽의 특징(IP주소, Port번호, 프로토콜 등)을 모니터링한 후 비정상적인 트래픽이 일정수준 이상 탐지되는 경우 공격으로 간주하는 방법이다. 둘째, 시간의 순서대로 발생되는 네트워크 트래픽을 웨이블렛 변환을 활용하여 주파수와 시간의 영역으로 표현한 후 특정 시간 영역에서 주파수의 분산이 높은 경우 이를 DDoS 공격으로 탐지하는 방법이다 (Li와 Lee, 2003). 셋째는 시간의 순서대로 발생되는 네트워크상의 트래픽의 양을 측정하는 특정 확률변수의 분포에 대한 변화점이 발생하는지를 통계적으로 분석하여, 이를 통해 DDoS 공격을 탐지하는 방법으로써 본 연구에서 사용하고자 한다.

분포의 변화점을 탐지하기 위한 전통적인 방법인 CUSUM 알고리즘, 지수가중치이동평균차트(Exponentially Weighted Moving Average Charts), 일반화가능도검정, Bayesian Shiryaev-Roberts 접근법 등은 변화점 이전의 자료의 분포가 알려져 있다고 가정하는 모수적 방법이다 (Basseville와 Nikoforov, 1993; Ross 등, 2011). 순위에 근거한 비모수적 CUSUM이나 비모수적 Shiryaev-Roberts 등 (Mcdonald, 1990; Gordon과 Pollak, 1994)이 존재하나 순위 계산 시 유발되는 메모리 부하 문제로 인해 대용량자료에서는 사용하기 적절치 않다. 정상상태(stationary regime) 기간 동안 계산된 관측치의 평균 및 분산을 이용하는 비모수적 CUSUM 방법 (Brodsky와 Darkhovsky, 1993)을 활용하여 각종 네트워크 공격을 탐지하는 Wang 등 (2004), Takada와 Hofmann (2004), Siris와 Papagalou (2006) 그리고 Tartakovsky 등 (2006)의 연구도 존재하지만, 변화점을 판단하는 임계치 결정 문제와 비정상적인 상황에서도 과도하게 누적합을 수행한 결과로 인해 공격의 끝 지점을 탐지하는데 있어서 민감도가 저하된다는 단점이 존재한다 (Ming, 2011).

본 연구에서는 분포의 변화점 검정을 위한 비모수적 방법으로서 경험적 분포함수를 이용한 방법과 자료의 순위를 이용한 검정방법을 사용한다. 앞에서 언급한 바와 같이 대용량의 자료에서 경험적 분포함수와 순위를 계산하는데 있어서는 메모리 공간문제와 계산시간의 문제에 직면하게 된다. 우리는 이 문제를 해결하기 위하여 Ross 등 (2011)가 제안한 Sliding Window와 Discretization(SWD) 방법을 사용한다. 이 방법은 자료의 개수가 점차 많아지더라도 항상 일정한 개수의 자료만 메모리에 저장하도록 하며, 변화점 탐지를 위한 계산의 양 또한 획기적으로 줄임으로써 대용량 자료에서의 계산시간 및 메모리 공간의 문제를 해결한다.

본 논문은 다음과 같이 구성되어 있다. 2절에서는 분포의 변화점을 탐지하기 위한 비모수적 방법으로써 경험적 분포함수의 차이를 이용한 Cramer-von-Mises와 Kolmogorov-Smirnov 검정법을 다루며, 순서

통계량을 사용한 방법으로 분포의 위치모수의 변화를 탐지하기 위한 Mann-Whitney 검정법, 척도모수의 변화를 탐지하기 위한 Mood 검정법, 그리고 위치 및 척도가 동시에 변화한 경우에 활용할 수 있는 Lepage 검정법을 소개한다. 3절에서는 대용량 자료에서 변화점 탐지를 위한 시간적, 공간적 제약조건을 만족시키는 Sliding Window와 Discretization 방법을 설명하고, 이에 대한 통계적 모형과 검정방법을 제안한다. 4절에서는 DDoS 공격 이전의 BPS, PPS 분포와 공격 이후에 나타날 수 있는 다양한 분포의 변화를 고려하여 대용량 자료를 생성하는 모의실험을 실시하여 최적 DDoS 공격 탐지 방법들을 탐색하고 토론한다. 5절에서는 DDoS 공격이 실제 발생한 실증 자료에 제안된 방법들을 적용시켜 효율성 및 정확성을 측정한다. 마지막 6절에서 결론을 유도한다.

2. 비모수적 변화점 탐지방법

Ross 등 (2011)은 분포의 위치와 척도 모수에 변화점이 존재하는지를 탐지하기 위한 비모수적 방법으로 순서 통계량을 사용한 방법들을 제안하였다. 또한 Ross와 Adams (2012)의 최근 연구에서는 분포의 변화점 탐지를 위해 경험적 분포함수를 사용하는 비모수적 검정방법들을 제안하였는데, 순서 통계량을 사용한 방법에 비해 변화점 탐지능력이 좋은 것으로 알려져 있다.

비모수적 변화점 탐지방법들을 설명하기 위하여 대용량 자료의 t시점에서 발생된 BPS 혹은 PPS 자료를 다음과 같이 정의한다.

$$D_t = \{x_1, \ldots, x_k, \ldots, x_t\}, \quad 1 < k < t < \infty.$$

이제 $k\,(< t)$시점에서 변화점이 존재하는지를 검정하기 위하여 t시점까지 발생한 표본을 $D_t^1 = \{x_1, \ldots, x_k\}$과 $D_t^2 = \{x_{k+1}, \ldots, x_t\}$로 분리한다. 여기서 $D_t = D_t^1 \bigcup D_t^2$.

비모수적 DDoS 공격을 탐지하기 위한 분포의 변화점 탐지 문제는 D_t^1과 D_t^2간에 분포의 변화가 존재하는지를 검정하는 문제가 된다. 이를 검정하기 위한 귀무가설과 대립가설은 다음과 같이 표현할 수 있다.

$$H_0 : X_i \sim F_{D_t}^0, \qquad H_1 : \exists k < t : X_i \sim \begin{cases} F_{D_t^1}^0, & \text{if } 1 \le i \le k, \\ F_{D_t^2}^1, & \text{if } k < i \le t. \end{cases}$$

이를 검정하기 위한 경험적 분포함수를 이용하는 비모수적 방법은 다음과 같다. D_t^1과 D_t^2에 존재하는 각각의 표본들에 대한 경험적 분포함수는 각각 다음과 같이 정의된다.

$$\hat{F}_{D_t^1}(x) = \frac{1}{k}\sum_{i=1}^{k} I(X_i \le x), \qquad \hat{F}_{D_t^2}(x) = \frac{1}{t-k}\sum_{i=k+1}^{t} I(X_i \le x), \tag{2.1}$$

여기서 $I(\cdot)$는 지시함수이다.

$k\,(< t)$시점에서 변화점이 존재하는지에 대한 Cramer-von-Mises(CVM) 통계량은 경험적 분포함수의 차를 제곱한 $C'_{k,t}(x) = \sum_{i=1}^{t}[\hat{F}_{D_t^1}(x) - \hat{F}_{D_t^2}(x)]^2$로서 정의되고, t시점에서의 검정통계량은 다음과 같이 결정되며, 적절하게 선택된 h_t에 대하여 $\mathrm{CVM}_t > h_t$인 경우, 변화점이 존재하지 않는다는 귀무가설은 기각된다 (Anderson, 1962).

$$\mathrm{CVM}_t = \max_k \left[\frac{C'_{k,t} - \mu_{C'_{k,t}}}{\sigma_{C'_{k,t}}} \right], \tag{2.2}$$

여기서 $\mu_{C'_{k,t}} = (t+1)/6t$, $\sigma_{C'_{k,t}} = \sqrt{(t+1)/(45t^2) \cdot [t - (3/4)\{(t^2 + (t-k)^2)\}/\{k(t-k)\} + 1/2]}$.

경험적 분포함수를 사용하는 추가적인 검정방법으로서 Kolmogorov-Smirnov(KS) 검정통계량은 다음과 같다.

$$\mathrm{KS}_{k,t} = \max_x \left| \hat{F}_{D_t^1}(x) - \hat{F}_{D_t^2}(x) \right|. \tag{2.3}$$

KS검정통계량의 p-값은 Kim (1969)이 제안하고 Greenwell과 Finch (2004)이 근사시킨 방법으로 $p_{k,t}$를 결정한 후, t시점에서의 p-값은 $p_t = \min_k p_{k,t}$로 결정한다. 적절하게 선택된 유의수준인 α에 대하여 $p_t < \alpha$인 경우, 변화점이 존재하지 않는다는 귀무가설은 기각된다.

순서통계량을 이용한 비모수적 방법으로 다음과 같은 세 종류의 검정방법을 소개한다. 우선, D_t^1과 D_t^2의 분포의 차이가 존재하는지에 대한 순서통계량을 이용한 Mann-Whitney(MW) 검정통계량은 Wilcoxon 순위합 통계량(rank sum statistic)을 활용하여 다음과 같이 정의된다 (Gibbons와 Chakraborti, 2003).

$$U_{k,t}^{'} = \sum_{i=1}^{t} r(x_i) I_{D_t^2}(x_i) - \frac{(t-k)(t-k+1)}{2},$$

여기서 $r(x_i)$는 전체 표본에서 x_i의 순위이다. 또한 t시점에서의 통계량은 다음과 같다.

$$\mathrm{MW}_t = \max_k \left| \frac{U_{k,t}^{'} - \mu_{U_{k,t}^{'}}}{\sigma_{U_{k,t}^{'}}} \right|, \tag{2.4}$$

여기서 $\mu_{U_{k,t}^{'}} = k(t-k)/2$, $\sigma_{U_{k,t}^{'}} = \sqrt{k(t-k)(t+1)/12}$.

순서통계량을 사용하는 또 다른 검정방법으로서 Mood(MD) 검정통계량은 다음과 같이 정의된다.

$$M_{k,t}^{'} = \sum_{i=1}^{t} \left[r(x_i) - \frac{t+1}{2} \right]^2 I_{D_t^2},$$

t시점에서 두 표본들이 동일하게 분포되어 있다는 귀무가설 하에서 각점들의 기대순위는 $(t+1)/2$이다. 따라서 MD 검정방법은 각 점들의 순위가 그 기대값으로부터 얼마나 떨어져 있는지를 측정하는 검정통계량이다.

$$\mathrm{MD}_t = \max_k \left| \frac{M_{k,t}^{'} - \mu_{M_{k,t}^{'}}}{\sigma_{M_{k,t}^{'}}} \right|, \tag{2.5}$$

여기서 $\mu_{M_{k,t}^{'}} = (t-k)(t^2-1)/12$, $\sigma_{M_{k,t}^{'}} = \sqrt{k(t-k)(t+1)(t^2-4)/180}$.

마지막으로 Lepage (1971)는 두 개의 표본 간의 위치와 척도가 동시에 변화하는지에 대한 검정으로써 위치모수의 변화점을 검정하기 위한 Wilcoxon 순위통계량과 척도모수의 변화점을 검정하기 위한 Ansari-Bradley 통계량을 함께 사용한 LP 통계량을 제안했다. Ross 등 (2011)은 이러한 Lepage의 검정방법에서 Wilcoxon 순위통계량 대신 MW 통계량을 사용하고, Ansari-Bradley 통계량 대신 MD 통계량을 사용한 Lepage-type(LP) 검정통계량을 다음과 같이 제안하였다.

$$\mathrm{LP}_t = \max_k \left[\left(\frac{U_{k,t}^{'} - \mu_{U_{k,t}^{'}}}{\sigma_{U_{k,t}^{'}}} \right)^2 + \left(\frac{M_{k,t}^{'} - \mu_{M_{k,t}^{'}}}{\sigma_{M_{k,t}^{'}}} \right)^2 \right]. \tag{2.6}$$

3. SWD를 이용한 비모수 DDoS 공격탐지 방법

대용량자료의 t시점에서 분포의 변화가 존재하는지를 검정하기 위하여 $k < t$인 모든 시점에서 검정을 수행하는 방법은 다음과 같은 문제점이 존재한다. 첫째는 t가 증가함에 따라 변화점의 후보인 k 또한 증가하게 되므로 변화점 탐지를 위해 계산해야 하는 통계량의 개수 또한 증가하게 된다. 이는 실제 변화점 탐지를 위한 검정방법을 적용할 때 과도한 계산시간을 필요하게 한다. 둘째로는 t가 증가함에 따라 실제 검정에 사용되는 데이터의 개수가 증가하게 된다. 따라서 검정 방법을 적용할 때 이를 저장하기 위해 필요한 메모리의 양 또한 증가하게 된다. 이러한 문제로 인하여 DDoS 공격 탐지를 위해 대용량으로 발생되는 BPS나 PPS 자료에 이러한 방법을 적용하는 것은 불가능에 가깝다. 따라서 본 연구에서는 대용량자료에서 비모수적 변화점 탐지방법을 BPS 및 PPS 등의 자료에 적용하기 위한 Sliding Window 및 Discretization(SWD)의 방법을 제안한다.

이를 위해 t시점에서의 자료 집합을 다음과 같이 표현한다.

$$D_t = W_{n,t} \bigcup W_{n,t}^C, \tag{3.1}$$

여기서 $W_{n,t}$는 최근에 발생된 n개의 자료집합 $\{x_{t-n+1}, \ldots, x_t\}$으로서 t시점의 윈도우자료라 정의한다. $W_{n,t}^C$는 $W_{n,t}$의 여집합으로 그 이전에 발생한 데이터의 집합이며, 실제 검정에서는 메모리의 효율적 사용을 위하여 다음과 같은 동반(concomitant) 변수의 집합으로 요약된다.

$$W_{n,t}^C = \left\{ \left(S_1, c_1^{t-n}\right), \ldots, \left(S_m, c_m^{t-n}\right) \right\}, \tag{3.2}$$

여기서 $S_1, \ldots, S_m$은 처음 발생되는 s개의 자료를 사용하여 생성된 도수분포의 구간으로서 처음과 마지막 구간은 각각 $S_1 = (-\infty, a)$와 $S_m = [b, \infty)$이며 나머지 구간은 $S_j = [a + (j-2)w, a + (j-1)w)$, $j = 2, \ldots, m-1$로 생성된다. 여기서 $w = (b-a)/m - 2$이며, a와 b는 초기 s개의 자료 중 최소값과 최대값이다. 또한 c_i^{t-n}은 처음부터 $t-n$개의 데이터를 이용하여 생성된 도수분포에서 S_i 구간의 도수를 의미한다. 이때 선택된 s개의 표본의 범위(range)가 모집단의 범위에 가깝도록 선택하는 것이 바람직하다. 이를 위해서는 BPS 및 PPS에서 최소한 1일 이상의 자료를 수집한 후 이를 통해 도수분포를 작성하는 것이 좋다고 판단된다.

윈도우자료 내의 각 시점에서 그 이전과 이후의 분포가 동일한지를 검정하는 것이므로 $W_{n,t}$를 각 검정시점 k에서 $W_{n,t} = W_{n,t}^k \bigcup W_{n,t}^{n-k}$로 분리시킨다. 여기서 $W_{n,t}^k = \{x_{t-n+1}, \ldots, x_k\}$, $W_{n,t}^{n-k} = \{x_{k+1}, \ldots, x_t\}$, $t - n + 1 \le k < t$. 따라서 윈도우자료 내에서의 k시점에서의 가설검정은 다음의 두 집합 간에 분포의 차이점이 존재하는지에 대한 검정이 된다.

$$D_t^1 = \left\{ W_{n,t}^C, W_{n,t}^k \right\} \quad v.s. \quad D_t^2 = \left\{ W_{n,t}^{n-k} \right\}, \tag{3.3}$$

여기서 $D_t = D_t^1 \bigcup D_t^2$. 이러한 자료의 구조는 시간이 지남에 따라 업데이트되며 $t+1$시점에서는 $D_{t+1} = W_{n,t+1} \bigcup W_{n,t+1}^C$로 변경된다. 또한 $t+1$시점에서의 가설검정은 $D_{t+1}^1 = \{W_{n,t+1}^C, W_{n,t+1}^k\}$와 $D_{t+1}^2 = \{W_{n,t+1}^{n-k}\}$의 두 집합간에 분포의 차이점이 존재하는지에 대한 검정이 되며, 여기서 $D_{t+1} = D_{t+1}^1 \bigcup D_{t+1}^2$.

이러한 프레임워크 내에서 앞서 언급한 분포의 변화점이 존재하는지를 검정하기 위한 다섯 종류의 비모수적 검정방법을 사용할 수 있다.

3.1. 경험적 분포함수를 이용한 변화점 탐지방법

경험적 분포함수를 사용한 CVM과 KS 검정방법에서는 식 (3.3)에서의 D_t^1과 D_t^2를 다음과 같은 형태의

동일 구간에서의 도수분포를 통한 동반변수의 형태로 요약시킨다.

$$D_t^1 \approx \left\{\left(S_1, c_1^{t-n+k}\right), \ldots, \left(S_m, c_m^{t-n+k}\right)\right\}, \qquad D_t^2 \approx \left\{\left(S_1, c_1^{n-k}\right), \ldots, \left(S_m, c_m^{n-k}\right)\right\}. \tag{3.4}$$

이때 각 그룹의 도수분포를 이용한 j구간의 경험적 분포함수는 각 구간에서 다음과 같이 표현된다.

$$\hat{F}_{D_t^1}(x_j) = \frac{1}{t-n+k}\sum_{i=1}^{j} c_i^{t-n+1}, \qquad \hat{F}_{D_t^2}(x_j) = \frac{1}{n-k}\sum_{i=1}^{j} c_i^{n-k}, \tag{3.5}$$

여기서 $x_j \in S_j$. 이제 경험적 분포함수를 사용한 CVM 및 KS 통계량은 식 (2.2)와 (2.3)내에 존재하는 경험적 분포함수를 식 (3.4)의 SWD를 이용한 경험적 분포함수로 대체함으로써 구할 수 있다. CVM 검정통계량의 표준화를 위한 평균 및 분산은 k를 윈도우자료 내의 검정 시점인 $t-n+k$로 대체함으로써 동일한 방법으로 계산할 수 있다.

3.2. 순서통계량을 이용한 변화점 탐지방법

순서통계량을 이용한 MW, MD, LP 검정에서는 D_t^2에 존재하는 각 표본의 전체 표본으로부터의 순위를 윈도우자료 내에서의 순위와 그 이전에 발생한 자료를 요약해 놓은 식 (3.2)의 $W_{n,t}^c$로부터 구한 순위의 합으로써 다음과 같이 결정한다.

$$r(x_i) = r_w(x_i) + \sum_{j=1}^{m+2} c_j I(x_k > \nu_j) + 1, \tag{3.6}$$

여기서 $r_w(x_i)$는 윈도우자료 내에서의 x_i의 순위이며, $\sum_{j=1}^{m+2} c_j I(x_i > \nu_j) + 1$은 윈도우자료 이전의 자료에 의해 생성된 동반변수를 통하여 계산된 순위이다. 그리고 ν_j는 S_j의 중앙값으로서 첫 번째와 마지막 세그먼트의 중앙값은 각각 $a-w$와 $b+w$로 결정되고 가운데 구간의 중앙값은 $\nu_j = a+(j-1.5)w$로 결정된다.

이러한 방법을 사용하여 D_t^1과 D_t^2간에 분포의 차이가 존재하는지를 검정하기 위해서는 식 (2.4), (2.5)의 MW, MD 통계량에서의 순서통계량 $r(x_i)$을 식 (3.6)으로 대체하고, 윈도우자료를 사용하지 않았을 경우의 검정시점인 k를 윈도우자료를 사용할 경우의 검정시점, $t-n+k$로 변경하여 구할 수 있다. 마찬가지로 MW와 MD 통계량의 표준화를 위해 필요한 평균과 표준편차에서 k를 $t-n+k$로 대체함으로써 계산할 수 있다. 또한 LP 통계량은 이렇게 구해진 MW와 MD 통계량을 사용하여 식 (2.6)과 동일하게 계산된다.

4. 모의실험

DDoS 공격을 탐지하기 위하여 사용되는 자료인 BPS와 PPS는 네트워크의 종류와 DDoS 공격의 유형에 따라 다양한 분포의 형태를 가지고 있다. 예를 들어 정상인 상황에서 하루 동안의 BPS나 PPS의 분포를 고려하는 경우에 낮 시간동안에는 일정 규모 이상의 트래픽이 발생하다가 밤 시간이 되면서 트래픽의 양이 서서히 감소하고 새벽에는 거의 발생하지 않는 경우라면, BPS 및 PPS는 정규분포에 가까운 분포의 형태를 가지게 된다. 특정시간대에만 네트워크 사용량이 많고 나머지 시간에는 거의 사용량이 없는 경우라면 0 방향으로 치우친 로그정규분포의 형태를 따른다.

DDoS가 발생한 이후에도 다양한 형태의 분포가 존재할 수 있다. 예를 들어 공격자가 네트워크를 다운시키는 것이 목적이라면 공격이후의 분포는 가용 BPS나 PPS의 한계점에 몰려있는 퇴화분

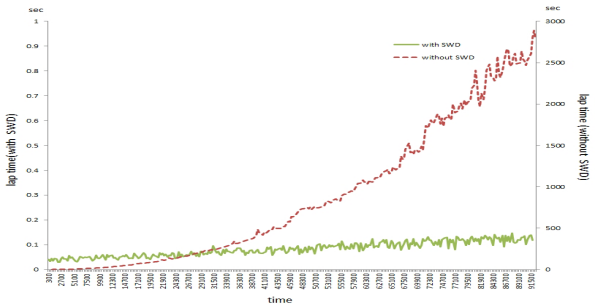


Figure 4.1. Lap time for MW detection methods.

포(degenerated distribution)에 접근한다. 네트워크를 다운시키는 것이 목적이 아니라 네트워크 서비스의 질을 저하시키는 것이 목적이라면, 정상인 경우와 분포의 형태는 같으나 평균과 분산이 변화하는 형태가 될 것이다. 특히 DDoS 공격에 의해 네트워크상에 많은 패킷들이 유입되게 되면, 일정수준 이상의 BPS나 PPS가 지속적으로 관측되므로 분산의 값은 작아진다.

우리는 이러한 DDoS 공격에 의한 분포의 변화를 검정하기 위한 구체적 모형으로서 공격이전의 BPS 및 PPS의 분포를 확률변수의 값들이 0보다 큰 정규분포와 왜도가 큰 값인 로그정규분포로 설정한다. 분포의 평균이 0 그리고 표준편차가 1이 되도록 표준화시킨 귀무가설 분포는 다음과 같다.

$$H_0^{(1)} : N(0,1), \qquad H_0^{(2)} : \frac{\text{LogNormal}(1,1/2)-3}{1.6}.$$

그리고 공격 이후의 변화된 분포인 대립가설 분포는 분포의 평균이 변화하는 경우(Type I)와 분포의 평균과 분산이 변화하는 경우(Type II), 그리고 분포의 형태가 균일분포 $U(1,4)$로 변화하는 경우(Type III)로 설정한다.

$$\begin{aligned} H_1^{\text{II}(1)} &: N(\delta,1), & H_1^{\text{II}(2)} &: \frac{\text{LogNormal}(1,1/2)-3}{1.6}+\delta_1, \\ H_1^{\text{II}(1)} &: N(\delta_1,\delta_2), & H_1^{\text{II}(2)} &: \left[\frac{\text{LogNormal}(1,1/2)-3)}{1.6}\right]\times\delta_2+\delta_1, \\ H_1^{\text{III}} &: U(1,4), \end{aligned}$$

여기서 $\delta_1 = 05, 1, 1.5, \ldots, 3.5$, $\delta_2 = 0.75, 0.5, 0.25$.

4.1. 경과 시간 비교

본 연구에서 제안한 SWD를 사용한 탐지 방법과 이를 사용하지 않은 방법의 성능을 비교하기 위하여 자료의 개수가 증가함에 따른 알고리즘의 경과시간(lap time)을 동일 컴퓨팅 환경에서 측정하였다. SWD를 사용할 경우 필요한 윈도우자료의 크기를 $n = 20$으로 설정하여 모의실험을 수행하였다. 그리고 다섯 종류의 비모수 검정방법 중 MW 방법을 사용하여 얻은 결과를 Figure 4.1로 표현하였다.

Figure 4.1은 표본자료 추출시간이 경과(수집된 표본자료의 개수가 증가)함에 따라 SWD를 사용한 경우와 이를 사용하지 않은 경우에 있어서 MW 검정방법을 실행시켜 각 시점 별 경과 시간을 측정한 결과이다. 측정된 두 방법의 경과 시간의 차이가 매우 크므로 Figure 4.1의 좌측 수직축은 SWD를 사용한 경우의 경과 시간(0부터 1초까지)을 표현하고, 우측 수직축은 SWD를 사용하지 않은 경우의 경과 시간(0부터 3,000초)을 표현하였다. 실선과 점선으로 각각 표현된 SWD를 사용한 방법과 이를 사용하지

않은 방법의 경과 시간은 자료생성 시작 후 30,000초을 초과하면서부터 급격한 차이가 나타나고 있는 것을 볼 수 있다. SWD를 사용한 방법은 대용량 자료가 생성되더라도 경과 시간은 0.2초 이내로서 매우 완만한 증가를 하고 있으나, SWD를 사용하지 않은 방법은 30,000초를 초과하면서 지수형태로서 급격하게 증가하여, 90,000초 근처에서는 2,500초(약 41.7분)가 소요되는 것을 볼 수 있다. 이러한 결과는 MW 검정 뿐 아니라 다른 검정방법에서도 동일하게 나타났으며, SWD를 사용한 경우와 이를 사용하지 않은 방법 간에서만 차이가 발생할 뿐 각 방법 내에서는 거의 차이가 존재하지 않았다. 따라서 SWD를 사용한 방법은 대용량으로 발생하는 BPS 및 PPS를 사용하여 DDoS 공격을 탐지하는데 있어서 컴퓨팅 속도 즉, 경과 시간 측면에서 매우 적절한 방법이라고 판단된다.

4.2. 탐지성능비교

공격탐지 알고리즘에 있어서 탐지 시간 이외에 탐지 알고리즘의 성능을 평가하기 위한 기준으로서 다음의 두 가지를 정의한다.

정의 4.1 평균 탐지지연(Mean Detection Delay: MDD) 시간

i번째 모의실험에서의 실제 변화점 τ_i와 탐지 알고리즘에 의해 탐지된 변화점 $\hat{\tau}_i\,(>\tau)$의 차이 시간의 평균을 평균 탐지지연(MDD) 시간이라고 하며 다음과 같이 정의한다.

$$\text{MDD} \equiv \sum_{i=1}^{n} \frac{\hat{\tau}_i - \tau_i}{n},$$

여기서 n은 모의실험을 수행한 횟수이다.

정의 4.2 1분 이내 탐지율(Detection Ratio within one minute: DRM)

모의실험을 수행한 횟수 중에서 공격발생이후 1분 이내 탐지하는 비율을 1분 이내 탐지율이라고 하며 다음과 같이 정의한다.

$$\text{DRM} \equiv \sum_{i=1}^{n} \frac{I(\hat{\tau}_i - \tau_i)}{n},$$

여기서 $I(\hat{\tau}_i - \tau_i) = \begin{cases} 1, \text{ 만약 } \hat{\tau}_i - \tau \le 60, \\ 0, \text{ 그외.} \end{cases}$

실제 네트워크 환경에서는 DDoS 공격이 시작된 후 1분 이내에 이를 탐지하지 못할 경우에 대응 조치를 수행할 시간적 여유가 없이 서비스가 불능화 상태에 빠지게 된다. 따라서 우리는 공격 탐지 방법이 의미있는 역할을 수행할 수 있는 시간적 한계를 1분으로 설정하여 알고리즘을 평가하기로 한다.

앞에서 설정한 세 가지의 분포의 변화의 유형에 대하여 SWD를 사용한 검정방법과 이를 사용하지 않는 경우의 검정방법의 MDD와 DRM을 계산하기 위한 모의실험을 실시하였다. 이를 위해 실제 변화점이 $\tau \in \{60, 120, 180, 240, 900, 86400, 604800\}$에서 발생하는 1,000개의 자료 집합을 각각 추출하였으며, 여기에 SWD를 사용하지 않은 다섯 가지의 비모수적 검정방법과 SWD를 사용한 검정방법을 적용시킨 후 MDD와 DRM을 계산하였다. 이때 각각의 검정방법에서 사용할 유의수준으로서는 $\text{ARL}_0 = 500$을 사용하였다. 여기서 ARL(average run length)은 $E(\hat{\tau}|F = F^0)$로 정의되며 유의수준 α의 역수이다. 그리고 MDD는 $\text{ARL}_1 = E(\hat{\tau} - \tau|F = F^1)$의 추정값이다 (Ross 등, 2011).

Figure 4.2의 좌측 그래프는 $H_0^{(1)} : N(0,1)$ *vs.* $H_1^{I(1)} : N(\delta_1, 1)$에 대한 CVM 검정방법의 MDD를 표현한 것으로서 실선은 SWD를 사용한 경우이며, 점선은 이를 사용하지 않은 경우의 MDD를 의미한다. 우측은 $H_0^{(1)} : N(0,1)$ *vs.* $H_1^{III} : U(1,4)$에 대하여 SWD를 사용한 것과 이를 사용하지 않은 다섯

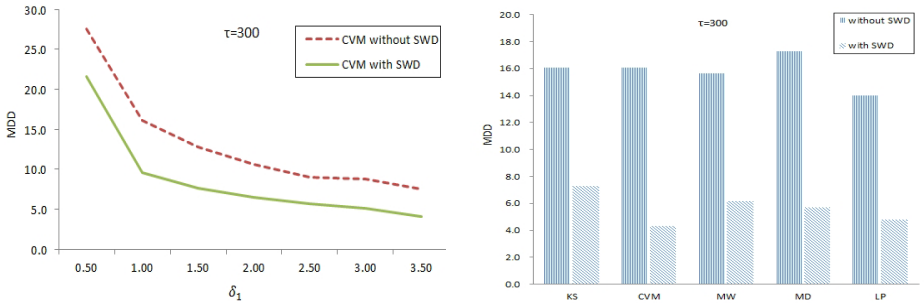


Figure 4.2. The MDD time with and without SWD.

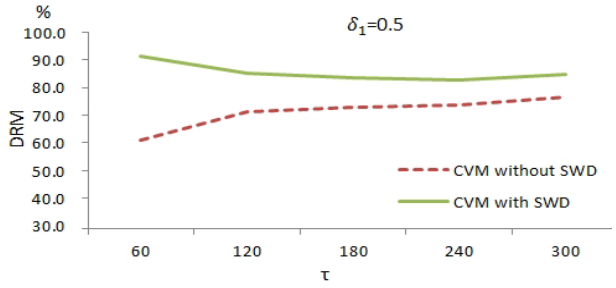


Figure 4.3. The DRM with and without SWD.

종류의 검정방법을 적용하여 구한 각각의 MDD값을 비교한 것이다. 이를 통해서 확인할 수 있는 결과는 다음과 같다. 다섯 종류의 비모수적 검정방법 모두에서 SWD를 사용한 방법이 SWD를 사용하지 않은 방법에 비해 탐지 성능이 더 뛰어나다. 이는 DDoS 공격에 의한 분포의 세 가지 변화 유형 모두에서 동일하다. 다섯 종류의 모든 비모수적 검정방법에서 유사하게 평균 및 표준편차의 변화량이 커질수록 MDD의 성능은 개선된다.

DRM은 분포의 변화에 있어서 변화량이 큰 경우, 즉 $\delta_1 > 0.5$와 $\delta_2 < 0.75$인 경우에는 모든 검정방법에서 100%에 근접한 값이 출력되어 비교가 불가능하였다. 따라서 DRM의 차이가 비교적 크게 나타난 평균이 $\delta_1 = 0.5$로 변화된 경우에 한하여 SWD를 사용한 방법과 이를 사용하지 않은 방법을 비교하고, 실제 변화점의 위치에 따른 CVM 검정방법의 DRM을 Figure 4.3에 표현하였다. 변화점의 위치에 상관없이 SWD를 사용하지 않은 방법에 비하여 이를 사용한 경우의 DRM이 크게 나타났으며, 변화점의 위치가 커짐에 따라 차이가 점차 감소되는 것을 파악할 수 있다.

SWD를 사용한 검정방법은 경과 시간과 탐지 성능에서도 SWD를 사용하지 않는 방법에 비해 우월한 방법임을 확인하였다. SWD를 사용한 경우의 경과 시간은 검정방법별로는 차이가 없는 것으로 나타났기에 탐지 성능을 통해 DDoS 공격에 의해 변화되는 분포의 유형(Type I, II, III)에 따라 어떠한 검정 방법이 최적인지를 살펴본다. 분포의 유형별 각각 1,000개의 자료집합을 추출하여 SWD를 사용하는 다섯 종류의 비모수 검정방법을 적용한 후 MDD와 DRM을 계산하였다. 자료 발생 시 실제 변화점들의 집합 및 윈도우자료의 크기, 초기 자료의 개수는 이전의 실험과 동일하게 선택하였다. 변화점이 자료발생 초기에 발생한 경우($\tau \in \{60, 120, 180, 240, 300\}$)는 동일한 패턴으로 MDD가 변화하였으며, 변화점이 자료 발생 후 1일 혹은 1주일이 경과한 후에 존재하는 경우($\tau \in \{86400, 604800\}$)에서도 동일한 패턴으로 MDD가 변화하는 것을 파악할 수 있다. 변화점이 자료발생초기에 발생하는 경우를

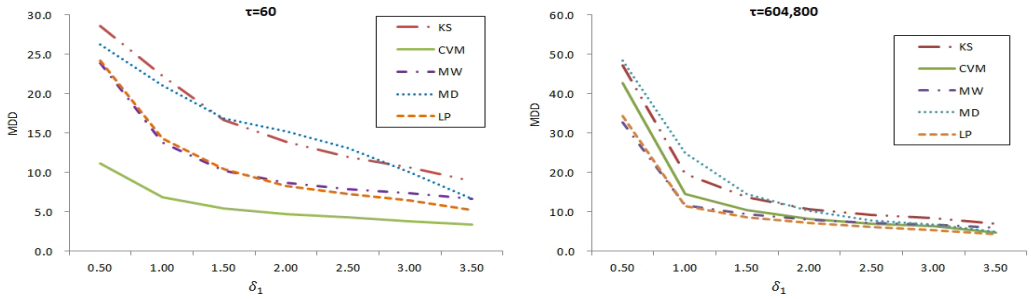
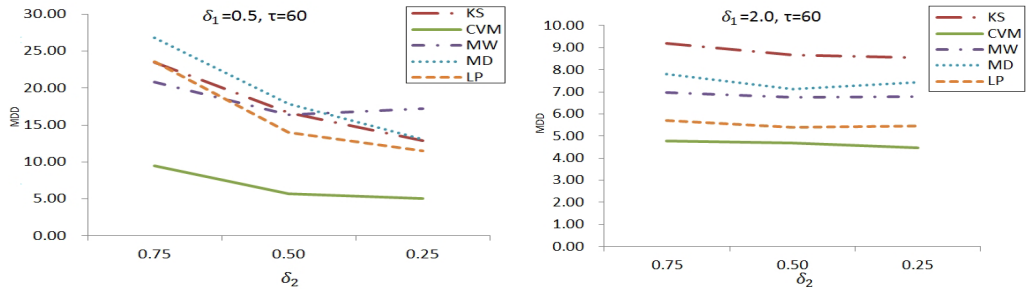
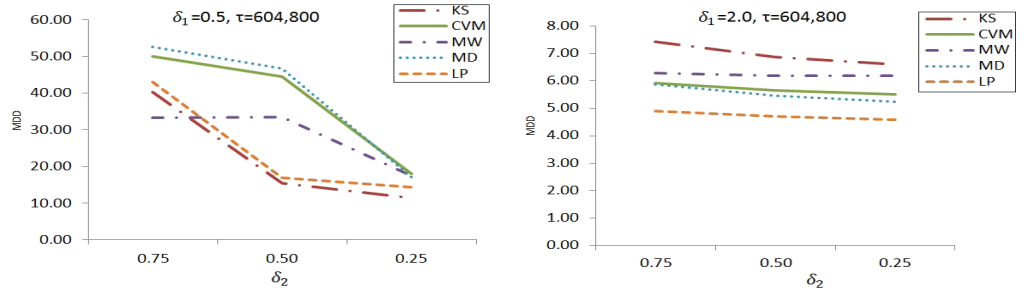


Figure 4.4. The MDD time for the means in Normal distribution with SWD.



(a) $\tau = 60$



(b) $\tau = 604,800$

Figure 4.5. The MDD time for the means and standard deviations in Normal distribution with SWD.

대표하여 τ = 684,000(소표본)과 1일 이상의 시간이 경과한 후 변화점이 발생하는 경우를 대표해서는 τ = 684,000(대표본)을 선택하여 분포 유형별 MDD를 Figure 4.4와 Figure 4.5에 표현하였다.

Figure 4.4는 표준정규분포의 평균의 변화 즉, $H_0^{(1)} : N(0,1)$ vs. $H_1^{I(1)} : N(\delta_1, 1)$에 대하여 다섯 개의 검정방법을 적용하여 MDD를 계산한 후, δ_1의 변화에 따른 MDD값을 소표본 자료와 대표본 자료에서 각각 표현한 것이다. 앞에서 언급한 바와 같이 분포의 변화량이 클수록 MDD값이 작아져서 탐지 성능이 좋아지는 것은 모든 검정방법에서 동일한 현상이다. 소표본 자료에서는 δ_1이 커짐에 따라 전반적으로 완만하게 MDD값이 개선되는 반면, 대표본 자료에서는 δ_1이 0.5에서 1로 변화할 때 급격하게 MDD값이 작아지고 그 이후에서는 완만하게 개선된다. 대표본 자료에서는 일정 수준이상의 평균변화가 발생하여야만 좋은 탐지 성능을 기대할 수 있다. 자료의 크기에 따른 검정방법 별 성능의 차이도 존

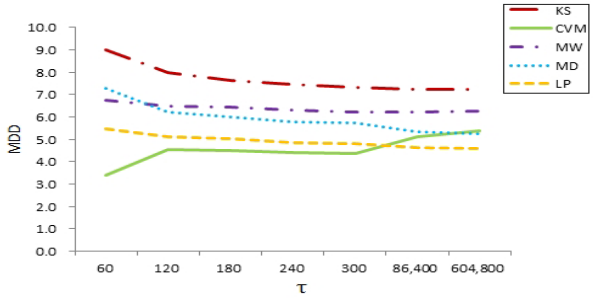


Figure 4.6. The MDD time N(0,1) to U(1,4) with SWD.

재하는 것을 발견할 수 있다. 소표본 자료에서는 δ_1의 변화에 따른 검정방법의 성능차이가 비교적 큰 편이나 대표본 자료에서는 크지 않았다. 또한 소표본 자료에서는 CVM이 다른 검정방법에 비해 월등한 성능을 보여주고 있으며, LP와 MW 그리고 KS와 MD의 순으로 좋은 성능을 보인다. 반면 대표본 자료에서는 그 차이는 크지 않으나 LP, MW, CVM, KS, 그리고 MD의 순으로 좋은 성능을 보이는 것을 탐색할 수 있다.

표준정규분포에서 평균과 표준편차가 동시에 변화할 때 즉, $H_0^{(1)} : N(0,1)$ *vs.* $H_1^{II(1)} : N(\delta_1, \delta_2^2)$에서 평균의 변화가 작은 경우($\delta_1 = 0.5$)와 큰 경우($\delta_1 = 2$)에 대하여 표준편차의 변화에 따른 MDD를 Figure 4.5에 표현하였다. 평균의 변화가 작은 경우에서는 소표본과 대표본 자료 모두에서 표준편차의 차이에 따른 검정방법의 성능의 차이가 큰 반면, 평균의 변화가 큰 경우에서는 표준편차의 차이에 따른 검정방법의 성능차가 거의 발생하지 않는다. 표준편차의 변화가 작은 경우와 큰 경우에 대하여도 평균의 차이에 따른 MDD의 비교에서도 동일한 현상을 발견하였다. 따라서 평균과 표준편차 중 어느 하나가 크게 변화하는 경우라면 탐지 성능은 큰 차이가 없는 것으로 판단한다. 표본의 크기에 따른 검정방법의 탐지 성능의 차이는 평균만 변화하는 경우와 유사하다. 소표본의 경우에서는 CVM이 가장 좋은 탐지 성능을 보이며, LP, MW, MD, KS의 순으로 성능이 좋은 것으로 나타났으며, 대표본에서는 LP가 가장 좋은 탐지 성능을 보이며, MD, CVM, MW, KS의 순서이다.

마지막으로 Figure 4.6은 $H_0^{(1)} : N(0,1)$ *vs.* $H_1^{III} : U(1,4)$에 대하여 변화점의 위치에 따른 MDD를 비교하였다. 대부분의 검정방법에서는 변화점의 위치가 커짐에 따라 즉, 수집되는 자료의 크기가 커짐에 따라 탐지 성능이 천천히 향상되나, CVM의 경우는 성능이 저하되는 현상을 보이고 있는 점이 특이하다. 표본의 크기에 따른 검정방법의 성능차이는 Figure 4.3 및 Figure 4.4와 유사하다. 소표본의 경우는 CVM, LP, MD, MW 그리고 KS의 순으로 탐지 성능이 우수한 반면, 대표본의 경우는 LP, MD, CVM, MW 그리고 KS의 순이다. 소표본인 경우 각 검정방법의 MDD의 차이가 비교적 큰 편이나 대표본의 경우 그 차이가 점차 감소하는 이유는 다음과 같이 정리할 수 있다. SWD를 사용하는 경우의 분포 비교는 $D_t^1 = \{W_{n,t}^C, W_{n,t}^k\}$와 $D_t^2 = \{W_{n,t}^{n-k}\}$이다. 표본수($t$)가 증가하게 되면 D_t^1을 구성하는 표본수($t-n+k$)는 점차 증가하게 되지만 D_t^2를 구성하는 표본수는 항상 $n-k$로서 일정하다. 자료의 개수가 증가함에 따라 표본의 분포는 모집단의 분포와 점차 가까워질 수밖에 없다. 즉 D_t^1에 의해 형성되는 표본분포가 참된 모집단의 분포에 가까울수록 이질적인 자료인 D_t^2에 의해 형성되는 분포와의 차이를 검정하는 검정통계량은 더 정확하게 분포의 차이를 찾아 낼 수 있다. 따라서 표본수가 커짐에 따른 다섯 종류 검정방법에 의한 MDD는 좋은 값으로 일치해 가는 현상이 발생한다.

또한 소표본에서는 경험적 분포함수를 이용한 CVM 방법의 MDD가 가장 좋은 결과를 보이는 반면, 변화점이 자료발생 후 일정기간이 경과한 후에는 순서통계량을 사용한 LP 검정방법의 MDD가 더 좋은

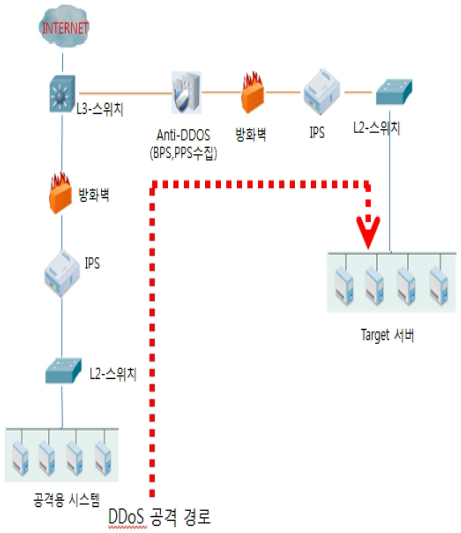
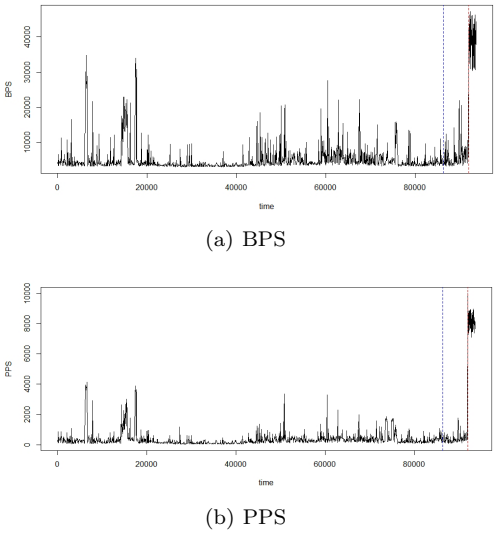


Figure 5.1. Network Diagram for BPS and PPS



(a) BPS

(b) PPS

Figure 5.2. BPS and PPS data

현상을 Figure 4.4, Figure 4.5, Figure 4.6에서 발견할 수 있다. 그 이유는 SWD를 사용하는 경우에 경험적 분포함수를 사용하는 검정방법은 각 자료의 상대도수 ($\hat{F}_{D_t^1}$, $\hat{F}_{D_t^2}$)의 비교이나 순서통계량을 이용한 검정방법은 D_t^2 자료로부터의 상대적인 크기인 순위를 이용하기 때문이라고 판단할 수 있다. 자료수가 증가하게 되면 표본 분포는 모집단 분포에 접근하므로, 표본의 작은 변화가 발생하더라도 그리 크게 반응하지 않으나 소표본에서의 변화량은 크다. 그러나 순서통계량을 이용하는 검정방법은 자료수가 증가하더라도 상대적 크기는 일정하게 유지되는 성질에 따라 성능을 일정하게 유지하기 때문이라고 판단된다.

5. 실증분석

DDoS 공격에 관한 실증 자료를 얻기 위하여 우리는 Figure 5.1과 같은 실제 네트워크에서 직접 DDoS 공격을 수행하였다. Figure 5.1의 점선은 공격용 시스템이 Target 서버까지 데이터를 전송할 때 통과하여야 하는 네트워크 시스템의 경로를 표현해 놓은 것으로서 그 중 Anti-DDoS 시스템은 DDoS 공격을 방어하기 위하여 네트워크의 트래픽의 양이나 개수를 모니터링하기 위한 전용 시스템이다. 우리는 이 시스템을 이용하여 DDoS 공격 이전 24시간 동안 BPS 및 PPS를 동시에 수집하였다. 그 후 Figure 5.1의 공격용 시스템을 이용하여 약 30분 동안 Target 서버로 DDoS 공격을 수행하면서 BPS와 PPS를 수집하였다. 공격용 시스템으로는 DDoS 공격용 프로그램인 'hping'을 설치한 다섯대의 컴퓨터를 사용하였으며, 구체적인 DDoS 공격방법으로써는 Syn flooding, UDP flooding 및 ICMP flooding 공격을 수행하여 총 93,631개의 자료를 수집하였다.

Figure 5.2는 시간이 경과함에 따른 PPS 및 BPS의 값을 표현한 것으로서 첫 번째 수직점선은 SWD 방법 적용 시 사용할 도수분포표를 생성하는 초기자료의 개수를 표시한 것이며, 두 번째 수직 점선은 실제 변화점이다.

DDoS 공격에 의해 상당한 분포의 변화가 발생한 것으로 볼 수 있으며, 특히 BPS 및 PPS의 평균 및 표준편차의 값이 크게 변화한 것을 Table 5.1을 통해서 알 수 있다. 실증자료에 대하여 SWD를 사

Table 5.1. Mean and Standard deviation of BPS and PPS before and after DDoS attack

통계량	BPS(단위: kbps)		CPS(단위: 개)	
	공격이전	공격이후	공격이전	공격이후
평균	514.60	3916.36	355.08	8165.85
표준편차	384.42	337.75	489.25	412.55

Table 5.2. Times detected by all five nonparametric methods with SWD

유의수준	자료유형	τ	$\hat{\tau}$				
			KS	CVM	MW	MD	LP
0.0001	BPS	91,906	91,914	91,909	91907	91908	91907
	BPS	91,906	91,914	91,909	91907	91908	91907
0.00001	PPS	91,906	91,915	91,910	91917	91909	91907
	PPS	91,906	91,915	91,910	91917	91909	91907

용한 다섯 개의 변화점 탐지방법을 적용하였다. SWD를 위한 윈도우자료의 크기는 $n = 20$으로 설정하였으며, 도수분포를 생성하기 위한 초기 데이터의 개수는 공격 발생이전 24시간동안의 자료의 특성을 고려한 도수분포를 생성하기 1일치의 자료(86400)을 사용하였다. 또한 검정을 위한 유의수준은 $\mathrm{ARL}_0 = 10{,}000(\alpha = 0.0001)$과 $100{,}000(\alpha = 0.00001)$을 사용하였으며 결과를 Table 5.2에 나열하였다. $\alpha = 0.0001$에서 각 검정방법 별 탐지시점의 차이는 8초 이내로서 큰 차이가 없었으나 그 중에서도, LP와 MW 방법이 가장 좋은 성능을 보이는 것으로 나타났다. 실증자료에서 $\tau = 91{,}906$이므로 대표본에 해당하며 표준편차의 변화는 작으나 평균의 변화는 크므로 평균 및 표준편차가 함께 변화하는 유형(Type II)이다. 4.2절에서 얻은 Type II 유형에 관한 모의실험에서 분포함수를 사용한 검정방법보다는 순서통계량을 사용한 LP 검정방법이 좋은 탐지성능을 보이는 결과와 일치함을 탐색하였다.

BPS와 PPS 자료간의 탐지 시간의 차이는 발생하지 않았으며, 이는 DDoS 공격으로 인한 분포의 차이가 두 자료 모두에서 모든 검정방법이 귀무가설을 기각하기에 충분할 정도로 크게 나타났기 때문으로 해석된다. $\alpha = 0.00001$에서는 MW의 탐지시점이 약 10초 정도 느리게 나타난 것 외에는 $\alpha = 0.0001$인 경우와 동일하게 나타났다.

6. 결론

DDoS 공격 탐지하기 위한 방법으로써 BPS와 PPS 자료의 분포의 변화점을 탐지하기 위한 다섯 종류의 비모수적 검정방법들을 소개하였으며, 대용량으로 발생하는 BPS 및 PPS 자료에 이들을 적용하는 SWD 방법을 제안하였다.

모의실험을 통해 SWD를 사용한 방법이 이를 사용하지 않은 방법에 비해 탐지 시간이 매우 향상된 방법이며, 탐지 성능 측면에서도 개선된 방법임을 발견하였다. 또한 DDoS 공격에 의한 분포 변화의 유형을 세 가지로 설정하여 각 유형별 최적 방법을 확인한 결과, 변화점의 위치가 자료발생 초기에 존재하는 경우는 모든 변화의 유형에서 경험적 분포함수를 사용하는 CVM 방법이 최적방법인 반면에 변화점의 위치가 자료 관측 후 상당한 시간이 경과한 이후에 존재하는 경우는 순서통계량을 이용한 검정방법인 LP 검정방법이 최적 방법임을 탐색할 수 있었다. 실증적으로 수집한 DDoS 공격 자료를 통해 분석한 결과도 모의실험의 결과와 일치하는 것을 확인할 수 있었다.

이러한 결과를 근거로 하여 특정한 분포를 가정할 수 없으며, 시간이 지남에 따라 관측되는 자료의 양이 대용량인 BPS 및 CPS 자료를 통해 DDoS 공격을 탐지하기 위한 방법으로써 SWD를 사용한 비모수적 검정방법은 매우 적절한 방법이라고 결론내릴 수 있다.

References

Anderson, T. W. (1962). On the distribution of the two-sample Cramer-Von-Mises criterion, *Annals of Mathematical Statistics*, **33**, 1148–1159.

Basseville, M. and Nikoforov, I. V. (1993). *Detection of Abrupt Change Theory and Application*, Prentice Hall, Englewood Clifs, NJ.

Brodsky, B. E. and Darkhovsky, B. S. (1993). *Nonparametric Methods in Change-point Problems*, Kluwer Academic Publishers.

Carl, G., Kesidis, G., Brooks, R. R. and Suresh, R. (2006). Denial-of-service attack-detection techniques, *IEEE Internet Computing*, **10**, 82–89.

Gibbons, J. D. and Chakraborti, S. (2003). *Nonparametric Statistical Inference*, 4th Edition, The university of Alabama.

Gordon, L. and Pollak, M. (1994). An efficient sequential nonparametric scheme for detecting a change in distribution, *Annuls of Statistics*, **22**, 763–804.

Greenwell, R. N. and Finch, S. J. (2004). Randomized rejection procedure for the two-sample Kolmogorov-Smirnov statistic, *Computational Statistics and Data Analysis*, **46**, 257–267.

Karen, S. and Peter, M. (2007). *Guide to Intrusion Detection and Prevention Systems(IDPS)*, Recommendations of the National Institute of Standards and Technology.

Kim, P. K. (1969). On the exact and approximate sampling distribution of the two sample Kolmogorov Smirnov Criterion, *Journal of the American Statistical Association*, **64**, 1625–1637.

Lepage, Y. (1971). A combination of Wilcoxon's and Ansari-Bradley's statistics, *Biometrika*, **58**, 213–217.

Li, L. and Lee, G. H. (2003). DDoS attack detection and wavelets, *Computer Communications and Networks, Proceedings*, **12**, 421–427.

McDonald, D. (1990). A Cusum procedure based on sequential ranks, *Naval Research Logistics*, **37**, 627–646.

Ming, Y. (2011). A nonparametric adaptive CUSUM method and its application in source-end defense against SYN flooding attacks, *Wuhan University Journal of Natural Sciences*, **16**, 414–418.

Ross, G. J. and Adams, N. M. (2012). Two nonparametric control charts for detecting arbitrary distribution changes, *Journal of Quality Technology*, **44**, 102–116.

Ross, G. J., Dimitris, K. and Adams, N. M. (2011). Nonparametric monitoring of data streams for changes in location and scale, *Technometrics*, **53**, 379–389.

Siris, V. A. and Papagalou, F. (2006). Application of anomaly detection algorithms for detecting SYN flooding attacks, *Computer Communications*, **29**, 1433–1442.

Symantec, Inc. (2011). Norton Cyber Crime Report 2011.

Takada, H. H. and Hofmann, U. (2004). Application and analyses of cumulative sum to detect highly distributed denial of service attacks using different attack traffic patterns, *IST INTERMON Newsletter*, **7**, 1–14.

Tartakovsky, A. G., Rozovskii, B. L. and Blazek, R. B. (2006). A novel approach to detection of denial-of-service attacks via adaptive sequential and batch sequential change-point detection methods, *IEEE Transactions on Signal Processing*, **54**, 3372–3382.

Wang, H., Zhang, D. and Shin, K. G. (2004). Change-point monitoring for detection of DoS attacks, *IEEE Transactions on Dependable and Secure Computing*, **1**, 193–208.

비모수적 DDoS 공격 탐지

이종락[a] · 홍종선[a,1]

[a]성균관대학교 통계학과

(2012년 11월 6일 접수, 2012년 12월 5일 수정, 2013년 3월 21일 채택)

요 약

네트워크상에서 분산 서비스 거부(DDoS) 공격 탐지를 위해 수집되는 트래픽 자료(BPS, PPS 등)는 시간 순서대로 발생하는 대용량 자료이다. 대용량 자료에서 공격 탐지를 위한 변화점 탐지 알고리즘은 정확성 뿐 아니라 시간과 공간적인 계산 수행의 효율성이 확보되어야 한다. 본 연구에서는 대용량자료에서 변화점 탐지에 대한 Ross 등(2011)이 연구한 순차적인 Sliding Window and Discretization(SWD) 방법을 확장하였다. 그리고 경험적 분포함수와 순위를 이용한 다섯 종류의 검정방법을 사용하면서 자료의 분포에 대한 가정없이 DDoS 공격을 탐지할 수 있도록 새로운 비모수 모형을 제안한다. 다양한 확률밀도 함수와 이에 대응하는 모평균과 분산을 변화시키면서 모의실험하여 본 연구에서 제안한 비모수적 검정방법을 SWD 방법에 적용하여 모형의 효율성을 탐색하고 토론한다. 그리고 실증 분석을 통해 공격 탐지율 및 공격 탐지의 정확성을 기준으로 성능을 측정하고 비교하였다.

주요용어: 빅데이타, 변화점, 슬라이딩 윈도우, 이산화, 탐지지연시간, 윈도우 데이타.

이 논문은 성균관대학교의 2012학년도 63학술연구비에 의하여 연구되었음.

[1]교신저자: (110-745) 서울시 종로구 성균관로 25-2, 성균관대학교 통계학과, 교수. E-mail: cshong@skku.edu