The Transactions of the Korea Information Processing Society
TKIPS Vol.13, No.4, pp.157~165
eISSN: 3022-7011
https://doi.org/10.3745/TKIPS.2024.13.4.157

Attention Based Collaborative Source-Side DDoS Attack Detection

Hwisoo Kim† · Songheon Jeong†† · Kyungbaek Kim†††

ABSTRACT

The evolution of the Distributed Denial of Service Attack(DDoS Attack) method has increased the difficulty in the detection process. One of the solutions to overcome the problems caused by the limitations of the existing victim-side detection method was the source-side detection technique. However, there was a problem of performance degradation due to network traffic irregularities. In order to solve this problem, research has been conducted to detect attacks using a collaborative network between several nodes based on artificial intelligence. Existing methods have shown limitations, especially in nonlinear traffic environments with high Burstness and jitter. To overcome this problem, this paper presents a collaborative source-side DDoS attack detection technique introduced with an attention mechanism. The proposed method aggregates detection results from multiple sources and assigns weights to each region, and through this, it is possible to effectively detect overall attacks and attacks in specific few areas. In particular, it shows a high detection rate with a low false positive of about 6% and a high detection rate of up to 4.3% in a nonlinear traffic dataset, and it can also confirm improvement in attack detection problems in a small number of regions compared to methods that showed limitations in the existing nonlinear traffic environment.

Keywords : DDos, Collaborative Source-Side Attack Detection, Attention

어텐션 기반 협업형 소스측 분산 서비스 거부 공격 탐지

김 휘 수† · 정 송 헌†† · 김 경 백†††

요 약

분산 서비스 거부 공격(DDoS Attack, Distributed Denial of Service Attack) 수법의 진화는 탐지 과정에서의 어려움을 가중시켰다. 기존 피해자측 탐지 방식의 한계로 인해 발생하는 문제를 극복하기 위한 솔루션 중 하나가 소스측 탐지 기법이었다. 그러나 네트워크 트래픽의 불규칙성으로 인한 성능 저하 문제가 존재하였다. 이 문제를 해결하기 위해 인공지능을 기반으로 한 여러 노드 간의 협업 네트워크를 활용하여 공격을 탐지하려는 연구가 진행되었다. 기존의 방법들은 특히 높은 버스트(Burstness)와 지터(Jitter)의 비선형적 트래픽 환경에서 한계를 보였다. 이러한 문제점을 극복하기 위해 본 논문에서는 어텐션(Attention) 메커니즘을 도입한 협업형 소스측 DDoS 공격 탐지 기법을 제시한다. 제안하는 방식은 여러 소스에서의 탐지 결과를 집계하여 각 지역에 가중치를 할당하며, 이를 통해 전반적인 공격 및 특정 소수 지역에서의 공격을 효과적으로 탐지할 수 있다. 특히, 비선형적인 트래픽 데이터셋에서 약 6% 수치의 낮은 가양성(False Positive)과 최대 4.3% 수치가 향상된 높은 탐지율을 보이며, 기존 비선형적 트래픽 환경에서 한계를 보였던 방법들에 비해 소수 지역의 공격 탐지 문제에 대한 개선도 확인할 수 있다.

키워드 : 디도스, 협업형 소스측 공격 탐지, 어텐션

1. 서 론

현재 국내외 IoT(Internet of Things, 사물인터넷)의 시장 규모가 크게 성장하고 있다. 스마트조명, 홈카메라, 냉난방 시설, 수도, 전기의 원격 제어 및 관리 등 친숙한 스마트홈 뿐만 아니라 헬스케어, 도시인프라, 공장설비 제어 등 다양한 분야로의 적용이 확대되고 있다. 글로벌 연구 기관 마켓앤마켓(Markets&Markets)의 보고서에 따르면, IoT 시장은 연평균 18.8%의 성장률로 2022년의 2,431억 달러에서 2027년에는 5,750억 달러까지 성장할 것으로 분석되었다. 또한, 한국 스마트홈산업협회에 따르면 2018년에는 17조 4,087억 규모였던 국내 IoT 시장이 2021년에는 22조 3,171억으로 증가했으며, 2027년까지 29조 9,375억으로 더욱 성장할 것으로 전망되었다. 이런 빠른 성장을 통해 IoT가 가져오는 편리함과 인

※ 이 논문은 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 지역지능화혁신인재양성사업(IITP-2024-00156287, 50%)과 2022년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임(IITP-2022-0-01203, 50%).
† 준 회 원 : (주)나눔테크 기업부설연구소 연구원
†† 정 회 원 : 전남대학교 정보보안협동과정 석사과정
††† 종신회원 : 전남대학교 인공지능학부/소프트웨어공학과 교수
Manuscript Received : December 26, 2023
First Revision : February 20, 2024
Accepted : February 26, 2024
* Corresponding Author : Kyungbaek Kim(kyungbaekkim@jnu.ac.kr)

기가 인류의 삶의 질을 향상시키고 있다.

그러나 IoT 기기는 성능에 제한적인 면이 있어 데스크톱 수준의 보안을 기대하기 어렵다. 또 한, 인증 문제, 펌웨어 업데이트 부재, 무작위 포트 접근 허용 등과 같은 보안 취약점을 통해 공격자들은 분산 서비스 거부(DDoS) 공격의 대상이 되어 여러 위치에 퍼져 있는 IoT 기기들의 보안 취약점을 악용한다[1]. 특히 2021년 클라우드 보안 회사인 Akamai의 고객 중 일부가 대규모 DDoS 공격에 피해를 입거나 Ransom DDoS 공격을 통해 기업이나 조직에 대한 금전을 요구하는 문제, 보안이 취약한 IoT 기기를 이용하여 대규모의 봇넷(Botnet)을 형성하고 DDoS를 수행하는 사례 등 DDoS 공격은 지속적으로 발전하고 증가하였으며 DDoS 공격 탐지와 대응을 위한 연구가 계속되는 추세이다[2].

DDoS 공격 탐지는 전통적으로 피해자 측에서 수집된 트래픽 볼륨을 정적 임계값을 통해 탐지하는 방식을 사용하였다. 그러나 이 방법은 공격 후에 수동적인 방어만 가능하며, 공격자 추적 및 탐지 지연에 어려움이 있다[12]. 이러한 문제를 해결하기 위해 소스측 공격 탐지 방법이 연구되었지만, 관찰된 트래픽이 상대적으로 적어 정상 트래픽 내에서 공격 트래픽을 분리하기 어려웠다. 따라서 이를 해결하기 위해 적응형 임계 기법[3]이 연구되었으며, 더 높은 성능을 위해 계절적 특성을 감지하는 적응형 임계 방법[4]으로 발전이 이루어졌다.

계절성 트래픽이란 네트워크 트래픽이 특정 주기마다 같은 양상을 띄는 것을 의미하며 Fig. 1과 같이 일정한 패턴을 보인다. 결과적으로 정상 트래픽 볼륨이 장기간 동일한 때마다 기록될 경우 트래픽 계절성 모델링이 가능하며, 이전의 통상적인 트래픽을 기반으로 정상 트래픽 속에서 공격 트래픽을 더 효과적으로 유추할 수 있다.

딥러닝의 활성화 증대에 따라 딥러닝을 기반한 네트워크 트래픽 예측 방법이 연구되었으나 해당 기법은 선형적인 특성을 가진 네트워크에서는 효과적이지만[5], 높은 지터(Jitter)[6]에 따른 무작위성이 크게 영향을 미치는 비선형적인 특성의 네트워크에서는 성능이 저하되는 경향이 있다. 또한 대규모의 동시다발적인 DDoS 공격이 시간대가 다른 여러 지역에서 발생할 경우 단일 사이트에 대한 소스측 탐지 방법은 피해

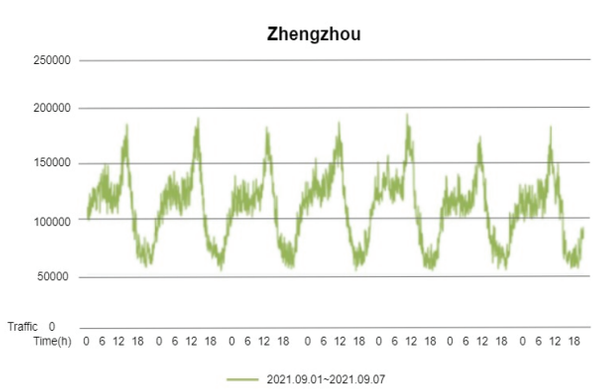

Fig. 1. Traffic Seasonality in Zhengzhou, China

자측 네트워크에 비해 소량의 트래픽만이 탐지되고 여러 지역이 아닌 단일 지역만을 관찰한다는 제약으로 인해 공격 트래픽을 탐지하지 못할 수 있다.

이러한 소스측 공격 탐지의 성능 저하 문제를 극복하기 위한 해결책으로, 협업형 소스측 공격 탐지 기법이 제시되었다[7]. 이 기법은 네트워크 트래픽의 예상치 못한 버스트성(Burstness)에 따른 가양성(이하 False Positive)을 줄이는 데 큰 효과를 보인다[8]. 또한, LSTM(Long Short-Term Memory)을 기반으로 각 소스측 탐지 모듈에서 집계된 탐지 결과와 트래픽 변화율, 시간 지수 등의 정보를 수집하고 최종 공격 탐지 결과를 공유해 협업 사이트들의 적응 임계값을 조정하는 기법이 연구되었으며, 실제 DNS(Domain Name System) 트래픽 데이터를 이용하여 성능을 평가하였다[9]. 그러나 해당 기법은 여러 소스측 탐지 모듈로부터 전체적인 공격이 입력으로 들어온 경우를 최종적인 공격으로 판단하여 소수 지역에 대한 공격을 감지하지 못한다. 이는 위험도가 높은 소수 지역에서만 공격이 발생한 경우에 탐지를 하지 못할 수 있음을 의미한다.

위와 같은 문제를 해결하기 위해 본 논문에서는 어텐션(Attention) 모델을 적용한 협업형 소스측 DDoS 공격 탐지를 제안하였다. 제안된 기법은 어텐션(Attention) 메커니즘의 특정 부분에 초점을 맞추어 개발되었다. 이를 통해 전체적인 공격뿐만 아니라 각 지역별 위험도를 구분하여, 공격 위험도가 높은 소수 지역에서의 공격을 효과적으로 탐지할 수 있게 되었다.

2. 관련 연구

2.1 소스측 공격 탐지

기존 피해자측 공격 탐지 기법의 한계가 뚜렷해지고 피해자측보다 더 일찍 공격 트래픽을 탐지할 필요성이 증대됨으로 인해 소스측 공격 탐지 기법이 연구되었다. 소스측 공격 탐지 시스템은 피해자로 전송되는 DoS 공격 트래픽을 탐지하기 위해 네트워크의 게이트웨이에 공격 탐지 모듈을 배치한다. 이 모듈은 DNS 샘플러, NTP 샘플러와 같은 샘플러를 활용하여 SDN을 통해 네트워크 트래픽을 수집한다. 수집된 트래픽 볼륨은 적응형 임계값을 통해 분석되며, 임계값을 초과하는 트래픽 볼륨은 공격 트래픽으로 간주되어 알림이 발생한다. 동시에 이전의 로그 기록을 활용하여 다음 시간대의 정상 트래픽 볼륨을 추정하고, 이를 기반으로 마진값을 고려하여 적응형 임계값을 동적으로 조절한다. Fig. 2는 소스측 공격 탐지 알고리즘이다.

2.2 협업형 소스측 공격 탐지

기존의 단일 지역 기반 소스측 공격 탐지 기법은 탐지 모듈이 위치한 지역의 트래픽 특성에 따라 성능에 제약이 있었다. 특히, 시차가 다른 여러 지역에서 동시에 대규모 공격이 발생

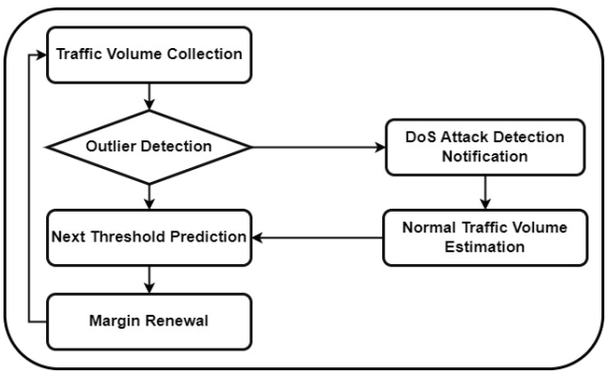

Fig. 2. Source-Side Attack Detection

할 경우, 단일 소스측 네트워크의 기법으로는 이러한 공격을 탐지하는 데 어려움이 있었다.

위 문제를 해결하기 위해 연구된 협업 소스측 DDoS 공격 탐지 기법은 다양한 시간대에 위치한 각 소스측 공격 탐지 모듈에서의 정보를 통합한다. 이 정보들을 종합하여 가중치를 고려한 후, 이를 공유하여 최종적으로 공격 여부를 판단하는 기법이며 Fig. 3은 협업 소스측 DDoS 공격 탐지 기법에 대한 개요를 나타내고 있다. 공유된 탐지 결과에는 각 소스측의 적응 임계값(Adaptive Threshold)에 마진이 포함된 통계적 가중치(Statistical Weight)와 각 소스측의 탐지 결과 등이 해당된다.

이렇게 각 소스측의 공격 탐지 모듈들이 통계적 가중치를 공유함으로써, 적응 임계값이 가진 False Positive 증가 문제나 탐지율 감소와 같은 부정적인 효과를 줄일 수 있다. 그러나 네트워크 트래픽의 패턴과 이에 해당하는 적응 임계값의 최적 마진 또한 계속 변하게 됨으로 최적의 성능을 위해서는 통계적 가중치의 재계산이 진행되어야 한다. 또 한, 불규칙한 사용 패턴을 보이는 네트워크 사용자들에서 비선형성이 관찰되는 상황에서 다수의 소스측 탐지 모듈이 배치된 네트워크 탐지 결과를 집계할 경우 협업형 공격 탐지의 성능이 저하될 수 있다.

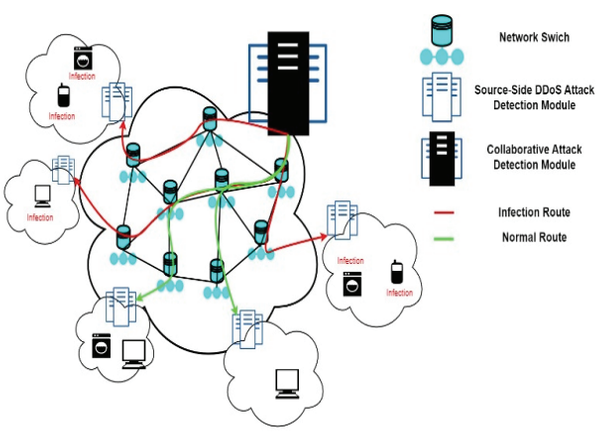

Fig. 3. Collaborative Source-Side Attack Detection

2.3 통계적 가중치 기반 협업형 소스측 공격 탐지

협업형 소스측 공격 탐지 모듈의 성능을 향상시키기 위해 소스측 DoS 공격 탐지 모듈의 성능을 통계적으로 나타내야하며, 특정 시점에서의 공격 탐지 확률과 오탐 확률을 계산한 값이 필요하다. 이를 위해 수집된 탐지 결과 SR_z와 k 사이트의 z^{th} 시간 창에서 통계적 가중치 w_z^k를 이용해 가중 산술 평균 WA_z를 계산한다. 이는 Equation (1)과 같다.

$$WA_z = \frac{\sum_{k \in l}^{N} w_z^k * SR_z^k}{\sum_{k \in l}^{N} w_z^k} \quad (1)$$

해당 가중 산술 평균 WA_z가 임계값 th보다 크면 최종 탐지 결과 FR_z는 Equation (2)와 같이 최종 탐지로 결정된다.

$$FR_z = \begin{cases} 1 \text{ if } WA_z \geq th \\ 0 \text{ if } WA_z < th \end{cases} \quad (2)$$

통계적 가중치 기반의 협업형 소스측 공격 탐지 기법은 소스측 탐지 모듈의 성능에 영향을 받는다. 소스측 공격 탐지 모듈에서 오탐 및 과탐이 자주 관찰될 경우 협업 탐지의 성능도 마찬가지로 저하된다. 특히 소스측의 네트워크에서 잦은 버스트(Burst)와 지터(Jitter)가 관찰되는 등 비선형적인 네트워크 특징이 큰 경우에 통계적 가중치 기반의 협업 탐지 성능이 더욱 저하된다.

2.4 LSTM 기반 협업형 소스측 공격 탐지

통계적 가중치 기반 협업형 소스측 공격 탐지 방법의 한계로부터 발생하는 문제를 극복하기 위해 LSTM을 활용한 협업형 소스측 공격 탐지 기법이 제시되었다. LSTM 기반 협업형 소스측 공격 탐지 기법에서는 이전 연구의 단점을 보완하기 위해 각 소스측의 협업 성능과 트래픽 패턴을 고려하였다. 이 연구에서 트래픽 패턴은 소스측 네트워크에서 트래픽이 시시각각 변하는 특성과 함께 선형적인 트래픽에서 통계적 가중치와 적응 임계값의 성능이 높게 나오는 반면, 높은 지터(Jitter), 버스트(Burst) 등과 같은 비선형적인 트래픽에서는 이들의 성능이 낮아질 수 있다는 점이 관찰된다.

이에 대응하여 공격 탐지 이후에는 시계열 딥러닝을 활용하여 네트워크 트래픽을 예측하는 방법을 도입하여 성능을 향상시키고자 하였다. 뿐만 아니라, 단일 사이트의 한계를 극복하기 위해 각 협력자(Collaborator, 소스측) 간의 탐지 결과를 효과적으로 공유하는 방식을 도입하였다. 이 기법의 프레임워크는 소스측 공격 탐지 모듈, 종단간의 연결 상태 모니터링과 오류 복구를 담당하는 이벤트 핸들러, 협업자 리스트를 집계

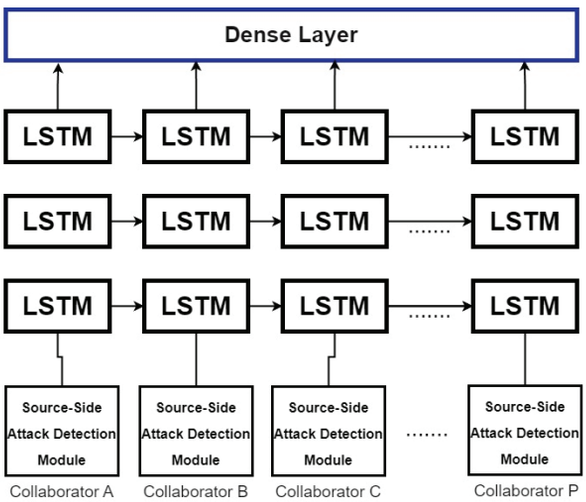

Fig. 4. LSTM Based Collaborative Source-Side Attack Detection Model

하여 최종 공격 결과를 결정하는 신뢰 관리 모듈, 그리고 각 소스측 탐지 모듈로부터 나온 결과를 종합하여 최종적인 공격 탐지 결과를 도출하는 협업형 소스측 공격 탐지 모듈 등으로 구성되어 있으며 Fig. 4의 LSTM 기반 협업형 공격 탐지 모델로 구현되어 있다.

LSTM 기반 공격 탐지 기법은 각 소스측 공격 탐지 모듈이 위치한 지역들 중에서 전체적인 공격이 발생했을 때 탐지하는 특성을 가진다. 예를 들어, 16개의 지역 중 대다수의 지역에서 동시에 공격이 발생해야 최종적으로 공격으로 간주하는 방식을 채택하고 있으나 이러한 접근 방식은 소수의 지역에서 발생하는 지역적 공격에 대한 탐지가 어려운 한계를 가지고 있다.

이전의 LSTM 기반 공격 탐지 기법은 각 지역별 통계적 가중치를 계산하지만, 이는 특정 시점에서 각 협업자의 탐지 및 오탐 확률을 계산하고, 이를 기반으로 신뢰 관리 모듈로부터 협업 탐지에 포함할 협업자 리스트를 만드는 데 사용된다. 즉, 이러한 접근 방식은 협업자(지역) 그 자체에 직접적으로 가중치를 부여하는 것이 아니라, 각 협업자의 독립된 탐지 결과를 종합하여 최종적인 결정을 내리는 방식이다. 이로 인해 공격이 자주 발생하는 높은 위험도를 가진 특정 지역에서 실제 DoS 공격이 발생했을 때, 해당 지역을 효과적으로 탐지하지 못하는 문제가 발생할 수 있다.

3. Attention 기반 협업형 소스측 DDoS 공격 탐지

위에서 제시된 단점을 극복하기 위해서 지역 자체의 특성을 고려하여 가중치를 부여하는 방법을 도입할 필요가 있다. 만약 소수 지역에서 공격이 발생하였을 경우 가중치가 높은 지역은 단일 지역에서 공격이 발생하더라도 이를 최종 공격으로 판단할 수 있는 근거가 될 수 있다. 반대로 가중치가 낮은 (평시 위험도가 낮은) 단일 지역에서 공격이 발생하더라도 이는 큰 위협으로 판단되지 않아 최종 공격이 아닌 것으로 판단함을 의미한다.

이에 특정 부분을 어텐션(Attention)하는 메커니즘을 사용하여 Fig. 5와 같이 각 협업자들에게 지역별 위험도를 나타내는 가중치를 부여하여 소수 지역으로부터의 공격을 판단하는 근거로 사용한다. 학습 간 현재 시점과 이전 시점의 탐지 결과들을 재참조해가며 각 지역들의 가중치를 계산한다. 어텐션(Attention) 적용 시 공격을 예측하는 매 시점마다 전체 지역을 재참고하게 된다. 이때 각 지역에 대해 동일 비율이 아닌 공격 예측 시점에서 더 연관성이 있는 지역을 주의하게 된다.

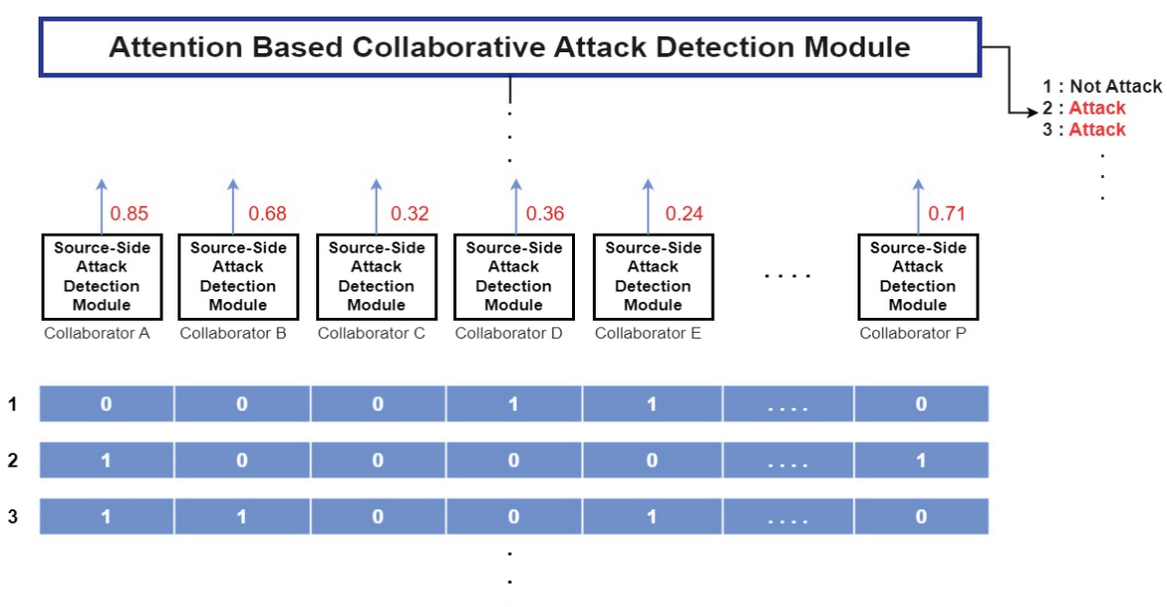

Fig. 5. Detection of Attacks in a Small Number of Areas by Region Weight

따라서 위험도가 높은 지역에 대해 더 높은 가중치를 부여할 수 있고 이전 기법에서 제기된 지역적 공격에 대한 탐지가 가능하며 기울기 소실이 발생하지 않아 성능적인 측면에서도 개선이 된다.

해당 모델에서는 LSTM이 아닌 Attention Layer를 적용하였다. 1개의 Attention Layer와 Dense Layer로 구성되어 있으며 레이어 노드는 하나이다. 입력 벡터로 각 소스측 탐지 모듈로부터 트래픽 볼륨 변화율, 공격 탐지 결과, 통계적 가중치를 전달 받으며 매 시점마다 들어온 입력들을 재참조하며 학습을 진행한다. 결과적으로 위험도가 높은 지역이 계산되고 해당 지역들로부터의 공격이 발생하는 경우에도 최종 공격으로 판별하게 된다.

어텐션(Attention)을 식으로 표현하면 Equation (3)과 같다.

$$Attention(Q,K,V) = Attention\ value \tag{3}$$

Q(Querry)는 소스측으로부터 탐지된 결과에 대한 t시점에서의 hidden state이다. K(Key)와 V(Value)는 탐지결과들 중 학습을 위해 임의로 공격이라고 지정한 것의 hidden state이며 Q를 K값들과 유사도를 구하고 Key값과 맵핑되어 있는 Value에 반영한다.

$$CI_z^i = (SR_z^i,\ ch_z^i,\ t_z^i,\ w_z^i) \tag{4}$$

Equation (4)는 입력벡터로써 협력자 i의 z^{th} 시간 창에서 관찰된 공격탐지율SR_z^i, 트래픽 볼륨 변화량ch_z^i, 시간지수 t_z^i, 통계적 가중치w_z^i로 구성된다. Fig. 6에서 확인할 수 있듯

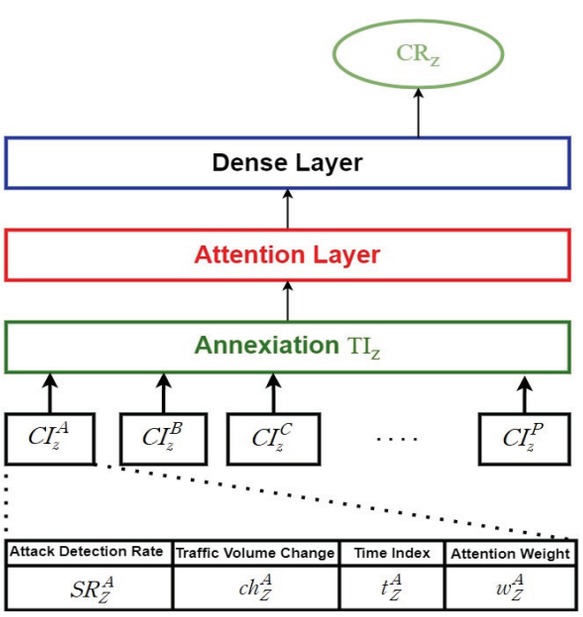

Fig. 6. Attention Model Learning

16개의 협력자로 구성되어 있어 협력자 A부터 P까지의 CI_z^A, CI_z^B, CI_z^C, ⋯, CI_z^P를 받아 학습하며 해당 모델은 정상(0)과 공격(1)의 협업 결과 CR_z를 반환한다.

$$score(s_t, h_i) = s_t h_i \tag{5}$$

어텐션 점수(attention score)는 Equation (5)와 같으며, 현재 t시점에서 Key값의 hidden state s_t가 Query의 hidden state값 h_i와 얼마나 유사한가 판단하는 값이다. 이를 구하기 위해 s_t를 전치하고 hidden state와 내적(dot product)을 한다.

$$e^t = [s_t h_1, \cdots, s_t h_n] \tag{6}$$

$$ad_t = softmax(e^t) \tag{7}$$

Equation (6)은 s_t와 query의 모든 hidden state의 어텐션 점수 모음 e^t를 구하는 수식이다. e^t에 소프트맥스를 적용해 모든 값의 합이 1이 되는 분포를 얻는다. 이 각각의 값을 어텐션 가중치(Attention Weight)라고 하며, 해당 분포 ad_t를 얻기 위해 Equation (7)을 정의한다.

$$C_t = \sum_{i=s}^{N} ad_{ti} h_i \tag{8}$$

Equation (8)은 어텐션 값(Attention Value) C_t식을 나타낸다. 최종 어텐션 값을 얻기 위해 각 가중치들을 모두 곱한 후 더하는 Weighted Sum을 한다.

어텐션 값을 s_t와 결합(concatenate)한 후 한 개의 벡터를 생성한다. 이를 tanh함수를 통해 출력층 연산을 위한 새로운 어텐션 벡터(Attention Vector) a_t를 얻는다. 이는 Equation (9)와 같이 표현할 수 있으며 W_c는 학습 가능한 가중치 행렬이다.

$$a_t = f(C_t, s_t) = \tanh(W_c[C_t; s_t]) \tag{9}$$

구한 a_t를 통해 여러 번의 학습을 수행하여 지역별 위험도에 대한 가중치를 갱신하며 소수 지역 공격에 대해 더 정확하게 탐지할 확률이 높아진다. 이와 같이 계산된 협업 결과 CR_z를 정적 임계값 th와 비교하여 이보다 크면 최종적인 DDoS 공격 결과인 FR_z로 판단한다. 이는 다음 Equation (10)으로 정의된다.

$$FR_z = \begin{cases} 1\ if\ \ CR_z \geq th \\ 0\ \ if\ \ CR_z < th \end{cases} \tag{10}$$

4. 성능 평가

제안하는 어텐션(Attention) 기반 협업 공격 탐지 기법의 성능을 평가하기 위해 임의의 별도 생성한 데이터가 아닌 실제 DNS(Domain Name System) 데이터를 수집하였고 이전의 기법에 적용한 경우와 제안한 새로운 기법에 적용한 경우의 결과를 비교하였다. 또 한, 학습 편차 발생을 고려하여 각각 4회씩 측정을 진행하였다.

실험을 진행하기 전에 몇 가지 기본적인 가정을 설정하였다. 전 세계적으로 각 지역마다 소스측 공격 탐지 모듈이 존재하며, 이들의 데이터를 중앙에서 집중적으로 관리하는 서비스 공급자가 필요하다는 것이다. 실제 세계에서는 이러한 역할을 수행하는 주체로는 각 국가의 정부기관이나 주요 통신사들이 될 것으로 예상된다. 한 국가 내에서도 다양한 통신사들 간의 협력이 필요한 시나리오가 있을 수 있으며, 이렇게 다양한 국가들 간의 DDoS 공격 탐지 데이터를 통합하는 것은 정책적, 제도적 측면에서 복잡한 문제로 여겨져 다양한 논의와 절차를 거쳐야 할 중요한 이슈로 간주된다. 이에 따라 여러 소스측에서 공격을 탐지한 결과를 받아 이를 협업 탐지 모듈 단에서 최종적인 공격을 판별하기 위해 관련된 제반 사항들이 모두 갖춰져있는 상태라고 가정한다. 테스트 환경은 Fig. 7과 같이 로컬 네트워크 상에서 도커(Docker)를 활용하여 가상 서버 16개를 설정하으며, 이 가상 서버들은 개별적으로 실제 소스측 공격 탐지 모듈로 작동한다. 또 한, 협업형 공격 탐지 모듈은 WSL(Windows Subsystem for Linux) 환경에서 실행된다.

제안한 어텐션(Attention) 기반 협업 공격 탐지 모듈의 성능의 평가, 검증을 위해 국제인터넷주소관리기구(Internet Corporation for Assigned Names and Nubmers, ICANN)에서 실제 수집한 DNS 트래픽데이터를 사용하였으며, 사용된 DNS 트래픽데이터는 서로 다른 시간대에 위치한 4개의 지

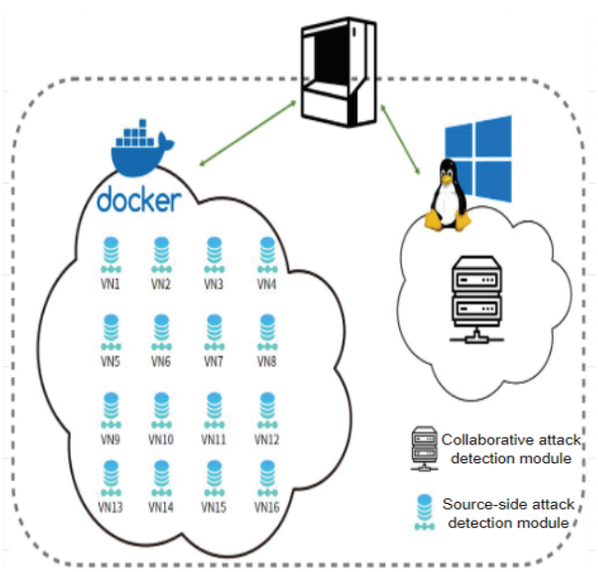

Fig. 7. Overview of Test Environment

Table 1. Test Area

Area	City 1	City 2	City 3	City 4
Northeast Asia	Beijing	Incheon	Shanhai	Zhangzhou
E.U	Paris. Orly	Geneva	Leeds Bradford	Florence
Eastern U.S	Atlanta	Chicago	Reston	Wilmington
Western U.S	Anchorage	LA	Portland	San jose

역과 4개의 도시, 총 16개의 사이트에서 2021년 9월 1일부터 2021년 11월 30일까지 총 3개월 동안 수집된 데이터이다. 무작위 지역이 아닌 현실에서 DDoS 공격이 자주 발생 가능한 환경으로 이루어진 지역을 선정하여 테스트의 타당성을 부여하였다. 16개의 도시를 선택하는 기준은 여러 가지 요소에 기반하였다. 먼저, 이들 도시 간에는 일정 이상의 시차를 갖고 있어야 했다. 또한, 해당 도시의 인구 수는 상당히 많아야 하며, 동시에 그 도시에서는 매일 일정량 이상의 트래픽 볼륨이 기록되는 지역으로 선별하였으며, 이는 Table 1의 목록과 같다.

시간대가 서로 다른 16개의 소스측 공격 탐지 모듈에서 트래픽 특성을 인식해 적응형 임계를 적용하기 위해 일정 이상의 시차를 가진 Northeast Asia, E.U, Eastern U.S, Western U.S 지역의 ICANN에서 제공하는 3개월간 최소 시간 단위인 5분 간격의 트래픽을 수집하였다. 수집된 3개월 분량의 트래픽 데이터는 지역별 총 26208개이며 이 중 2/3는 통계적 가중치를 계산하기 위해 사용되며 나머지 1/3 데이터 중 75%는 모델 학습을 위해 사용된다. 남은 25%의 데이터는 검증 테스트를 위해 사용되며 이 중 1/13을 공격 트래픽으로 무작위 지정하여 진행된다.

학습 시 오류 및 오차의 가능성을 배제하기 위해 각 모델별 최소 10회의 학습을 진행한 후 좋은 결과를 추출하였다. 테스트를 진행하기 위해 제안한 어텐션(Attention) 기반의 기법과 기존 LSTM 기법 외에 LSTM-Attention 모델에 대한 성능도 별도로 측정하였다. Fig. 8에 따르면 제안한 어텐션(Attention) 기반 기법이 기존 LSTM 기반 기법에 비해 더 높은 성능을 나타내는 것을 확인할 수 있다. 탐지율 부분에서 어텐션(Attention) 기법은 평균 97.6%의 성능을 보이고 LSTM 기반 기법은 최대 96%, 최저 93.3%의 비교적 낮은 성능을 나타내고 있음을 확인 할 수 있다.

LSTM-Attention 모델의 경우 성능이 가장 좋을 것으로 기대하였으나 LSTM으로부터 받은 Hidden state에 의해 LSTM보다는 성능이 높지만, 단일 어텐션(Attention) 모델에 비해 성능이 부족한 모습을 확인할 수 있다. 또한 LSTM 기반 탐지 기법의 경우 매 학습마다 성능이 급변하는 등의 이유로 제안한 어텐션(Attention) 기법보다 학습을 더 많이 진행하였다. 이에 대비해 어텐션(Attention) 기반 기법은 학습 회차에 따른 성능 변화량이 크지 않아 보다 안정적이라고 볼 수 있다.

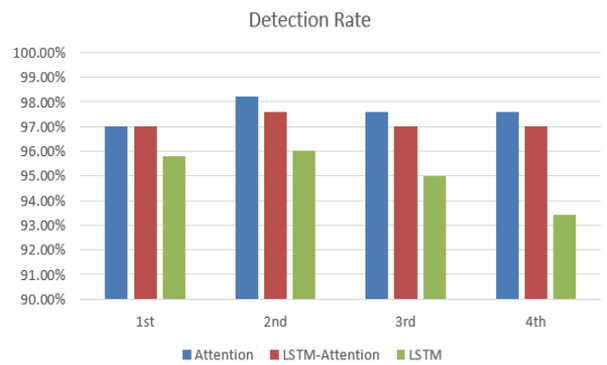

Fig. 8. Comparison of Detection Rates of Attention-based Techniques

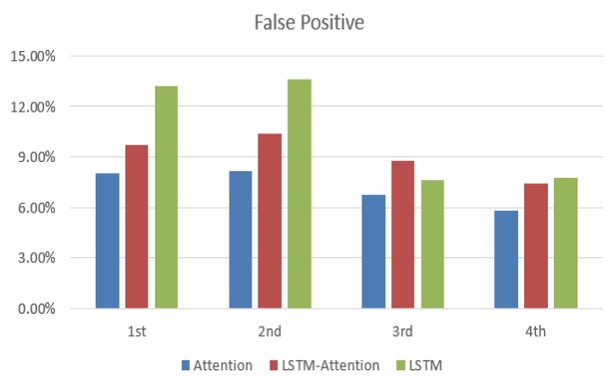

Fig. 9. Comparison of False Positive of Attention-based techniques

False Positive는 공격이 아닌 것을 공격이라고 판정한 수치이며 Fig. 9에서 확인할 수 있듯이 제안한 어텐션(Attention) 기법은 평균 7%대를 보여주는 반면, LSTM 기반 모델은 최대 13%로 높은 수치를 보여준다. 이는 수집한 DNS 데이터셋의 높은 지터(Jitter)를 가진 데이터에 의한 것으로 곧 탐지율 저하와 False Positive가 증가하는 원인이 되었다.

Fig. 10과 같이 수집한 데이터의 몇몇 지역은 트래픽이 순간적으로 높게 솟구치는 마이크로버스트(MicroBurst)가 많이 발생한다. 마이크로버스트(MicroBurst)는 짧은 순간 밀리초 수준으로 발생하나 이러한 짧은 순간에 발생한 트래픽 급변으로 인해 패킷 대기열이 발생하고 트래픽 패킷을 일정하게 보내지 못하는 스위칭 지연이 발생하게 되며, 이로 인해 지터(Jitter)가 발생한다.

어텐션(Attention) 기반 기법의 경우 LSTM 기반 기법보다 위와 같은 현상에서 더 자유로워 높은 지터(Jitter) 상황에서도 유연하게 높은 성능을 낼 수 있음을 보였다. F1 Score의 경우 낮은 탐지율과 높은 False Positive를 보여준 LSTM 기반 모델이 가장 낮게 나왔고 이와 반대의 성향을 보이는 어텐션(Attention) 기반 모델이 가장 높은 점수를 나타냈음을 Fig. 11에서 확인할 수 있다.

앞서 기존의 공격 탐지 모델의 한계를 극복하기 위해 소수 지역 탐지 여부에 대한 평가를 진행하였다. 이를 위해 16개의

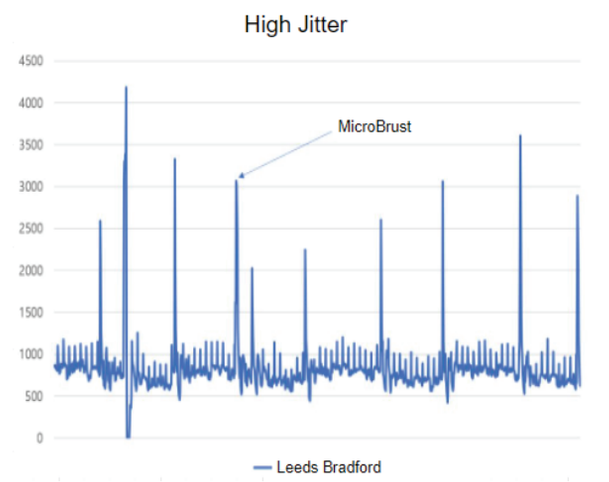

Fig. 10. Areas Showing High Jitter

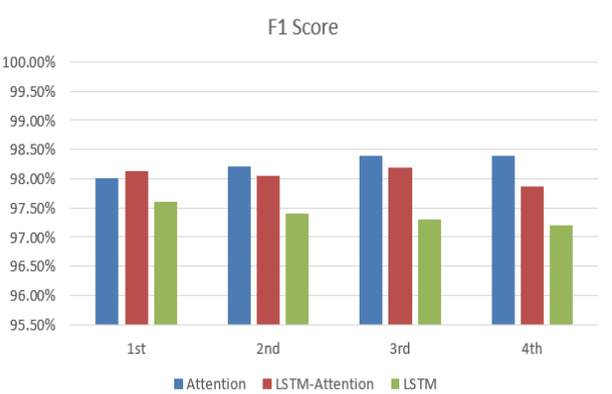

Fig. 11. Comparison of F1 Score of Attention-based Techniques

지역을 대상으로 실험을 진행하였으며, 각 실험에서는 특정 조건을 충족하는 지역을 공격 지역으로 설정하였다. 이러한 조건은 해당 지역의 인구 수, 트래픽 양, 사회적 중요성 등을 종합적으로 고려하였다. 예를 들어, Beijing, Shanghai, LA와 같이 고유의 특성을 가진 한 지역을 선정하거나, Paris. Orly+Leeds Bradford, Beijing+Shanghai, Shanghai+LA와 같이 두 지역을 동시에 선정하는 등의 조합을 고려하였고 동시에 4개의 지역을 평가하였다.

평가는 DNS 데이터셋의 테스트 데이터 2186개 중1/3에 해당하는 168개의 데이터를 공격으로 정의하였다. 각 지역에 동일한 횟수의 공격을 했을 때, 각 지역별 탐지를 확인하기 위해 이 중에서 소수 지역을 공격하는 경우도 일정 개수 포함하도록 했다. 그러나 3개 이상의 지역을 동시에 평가하는 경우, 테스트 데이터에 속한 공격 트래픽이 부족하여 다른 지역의 트래픽을 추가하여 실험을 진행하였다. 예를 들어, Zhangzhou와 Leeds Bradford 지역으로 이루어진 지역은 트래픽이 부족하여 10개를 채우지 못하므로 다른 지역과 트래픽 혼합을 진행하였다.

실험 결과 지역적 공격에 대한 탐지 불가에 대해 개선되었음을 Table 2에서 확인할 수 있다. 단일 지역에서부터 총 4개의 지역까지 공격이 발생하더라도 모델이 효과적으로 탐지하

Table 2. Whether Small Area Attacks are Detected

Area	Attacks	Detection
Beijing	10	10
Shanghai	10	10
Zhanzhou	10	9
Leeds Bradford	10	9
LA	10	10
Reston	10	9
Beijing, Shanghai	10	10
Beijing, Zhanzhou	10	9
Zhanzhou, Leeds Bradford	10	8
Beijing, Shanghai, Zhanzhou	10	9
Reston, LA, Chicago	10	9
Beijing, Incheon, Shanghai, Zhanzhou	10	10
Beijing, Chicago, Reston, LA	10	10

Table 3. Jitter and Burst by Region

Area	Jitter	Burst
Beijing	15.49	6.5
Incheon	5.72	5.65
Shanghai	7.61	5.49
Zhanzhou	8.45	5.85
Florence	11.01	4.3
Geneva	7.85	4.14
Leeds Bradfort	146.06	3.76
Paris. Orly	24.22	5.76
Atlanta	8.1	5.34
Chicago	11.11	5.07
Reston	8.06	3.96
Wilmington	10.3	5.87
Anchorage	5.97	5.39
LA	9.25	4.74
Portland	12.75	5.13
San jose	7.85	4.14

는 모습을 볼 수 있으며, 그 이상의 지역 수에 대한 실험 또한 경과가 동일할 것으로 예상된다. 100%의 탐지를 달성하지 못한 부분에 대해서는 모델 자체의 전체적인 탐지율에 관련한 것으로 판단된다.

추가적으로 지터(Jitter)가 높은 지역에 대한 소수 지역 공격 탐지율을 테스트하였다. 이는 일반적인 상황에서는 탐지가 잘 되더라도 네트워크에서 지터(Jitter)와 같은 비선형성이 큰 경우에는 성능이 저하될 수 있기 때문에 중요한 평가 요소이다. Table 3에서 확인할 수 있듯이, 사용한 데이터셋 중 Paris. Orly와 Leeds Bradford 지역은 지터(Jitter)가 매우 높은 것으로 확인할 수 있다. 특히, Leeds Bradford 지역에서는 타 지역 대비 Fig. 10에서 확인할 수 있듯이 다량의 마이크로버스트

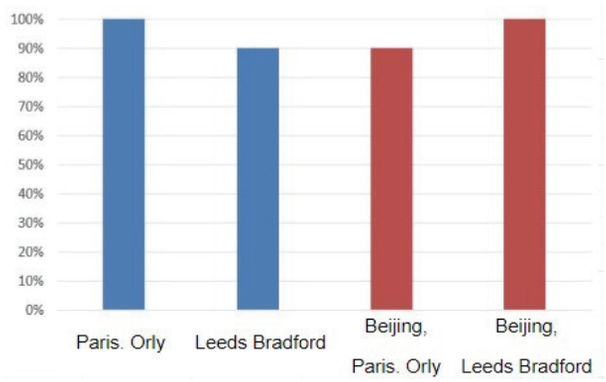

Fig. 12. High Jitter Area Attack Detection

(MicroBurst)가 나타났다. 해당 그래프는 데이터셋의 극히 일부분을 나타낸 것이지만, 다른 지역에 비해 독보적으로 높은 지터(Jitter)를 보여주었다.

평가 결과에서는 높은 지터(Jitter) 상황에서도 공격 지역을 단일 지역 또는 두 지역으로 설정하여 테스트하였을 때, 모델이 잘 탐지 해냈다는 것을 Fig. 12를 통해 확인 할 수 있다. 따라서 기존의 기법은 지터(Jitter)가 높은 상황에서 성능이 저하되는 문제가 있었지만, 제안된 기법은 지터(Jitter)와 같은 트래픽의 비선형성으로 인해 발생하는 성능 저하에도 효과적으로 대처 가능함을 확인할 수 있다.

5. 결 론

이 논문에서는 단일 지역의 소스측 공격탐지 기법으로 확인하기 어려운 분산 서비스 거부 공격을 탐지하기 위해, 어텐션(Attention) 기반의 협업형 소스측 DDoS 공격 탐지 방법을 제안하였다. 연구는 모의 데이터가 아닌 실제 DNS 네트워크 트래픽을 사용하여 효과를 평가하였으며, 결과로써 지역적 공격을 효과적으로 탐지하는 동시에 전체적인 탐지율을 최대 4.3% 상승시키고 False Positive를 약 6% 감소시키는 개선을 확인하였다. 특히, 데이터셋에 포함된 지터(Jitter)가 강한 지역에서도 이전 대비 확실한 성능 개선이 이루어져 전반적인 공격 탐지 능력이 향상되는 것을 확인하였다. 이러한 결과는 DDoS 공격의 위협이 계속해서 증대되는 현 상황에서, 본 연구에서 제안한 방법이 실제 환경에서 유용하게 활용될 수 있을 것으로 기대한다.

References

[1] Y. R. Siwakoti, M. Bhurtel, D. B. Rawat, A. Oest, and R. C. Johnson, “Advances in IoT security: Vulnerabilities, enabled criminal services, attacks, and countermeasures,” *IEEE Internet of Things Journal*, Vol.10, Iss.13, 2023.

[2] Y. Al-Hadhrami and F. K. Hussain, "DDoS attacks in IoT networks: a comprehensive systematic literature review," *World Wide Web*, Article, 06 Jan. 2021 Vol.24, pp.971-1001, 2021.

[3] R. Alkanhel, E. S. M. El-kenawy, D. L. Elsheweikh, A. A. Abdelhamid, A. Ibrahim and D. S. Khafaga, "Metaheuristic optimization of time series models for predicting networks traffic," *Computers, Materials & Continua*, Vol.75, No.1, pp.427-442, 2023, https://doi.org/10.32604/cmc.2023.032885

[4] Q. Li, X. Wu, Z. Cao, and J. Ling, "Anomaly detection of iot traffic based on LSTM and attention mechanism," *ICMLC '23: Proceedings of the 2023 15th International Conference on Machine Learning and Computing*, pp.457-463, 2023, https://doi.org/10.1145/3587716.3587792

[5] M. Alizadeh, M. T. Beheshti, A. Ramezani and H. Saadatinezhad, "Network traffic forecasting based on fixed telecommunication data using deep learning," *2020 6th Iranian Conference on Signal Processing and Intelligent Systems (ICSPIS)*, 23-24 Dec. 2020.

[6] A. Feldmann et al., "The lockdown effect: Implications of the COVID-19 pandemic on internet traffic," *Proceedings of the ACM Internet Measurement Conference*, pp.1-18, 2020.

[7] R. S. Tambe, H. Dand, and M. D. Salunke, "Role of machine learning ensemble in DDoS intrusion detection," *2023 IEEE 5th International Conference on Cybernetics, Cognition and Machine Learning Applications (ICCCMLA)*, 2023.

[8] S. Yeom, and K. Kim, "Improving performance of collaborative source-side DDoS attack detection," *2020 21st Asia-Pacific Network Operations and Management Symposium (APNOMS)*, pp.239-242, IEEE, 2020.

[9] S. Yeom, C. Choi, and K. Kim, "LSTM-based collaborative source-side DDoS attack detection," *IEEE Access*, Vol.10, pp.44033-44045, 2022.

[10] S. Yeom, C. Choi, and K. Kim, "Source-side DoS attack detection with LSTM and seasonality embedding," *Proceedings of the 36th Annual ACM Symposium on Applied Computing*, pp.1130-1137, 2021.

[11] T. Kawazoe and N. Fukuta, "On implementing a simulation environment for a cooperative multi-agent learning approach to mitigate DRDoS attacks," *International Joint Conference on Artificial Intelligence, IJCAI 2022: Recent Advances in Agent-Based Negotiation: Applications and Competition Challenges*, pp.15-29, 2022.

[12] K. Hwisoo "A study on attention based collaborative source-side DDoS attack detection," Chonnam National University Master's Thesis, 2023.

김 휘 수

https://orcid.org/0009-0003-0562-3362
e-mail : ryan95288@gmail.com
2021년 조선대학교 컴퓨터공학과(학사)
2023년 전남대학교 정보보안협동과정(석사)
2023년 ~ 현 재 (주)나눔테크 기업부설연구소 연구원
관심분야 : DDoS, 네트워크 보안, 실시간 데이터 분석 처리, 정보보안 등

정 송 헌

https://orcid.org/0009-0007-5589-9147
e-mail : thdgjs0514@jnu.ac.kr
2022년 광주대학교 사이버보안경찰학과 (학사)
2023년 전남대학교 정보보안협동과정 석사과정
관심분야 : 정보보안(Personal Information), 클라우드, 모빌리티, 블록체인 등

김 경 백

https://orcid.org/0000-0001-9985-3051
e-mail : kyungbaekkim@jnu.ac.kr
1999년 한국과학기술원 전기공학 및 컴퓨터공학(학사)
2001년 한국과학기술원 전기공학 및 컴퓨터공학(석사)
2007년 한국과학기술원 전기공학 및 컴퓨터공학(박사)
2007년 ~ 2008년 Network and Distributed Systems Group, Computer Science, in University of California Irvine
2008년 ~ 2012년 Information Systems Group, Computer Science, University of California Irvine
2012년 ~ 2016년 전남대학교 전자컴퓨터공학과 조교수
2016년 ~ 2021년 전남대학교 전자컴퓨터공학과 부교수
2021년 ~ 현 재 전남대학교 인공지능학부/소프트웨어공학과 교수
관심분야 : 지능형 분산시스템, SDN/NFV, 빅데이터 플랫폼, 인공지능, 블록체인, 소셜네트워크 등