https://doi.org/10.33778/kcsa.2020.20.4.071

트래픽 유통계획 기반 사이버전 훈련데이터셋 생성방법 설계 및 구현★

김 용 현*, 안 명 길*

요 약

사이버전 훈련 시스템에 현실감 있는 트래픽을 제공하기 위해서는 사전에 트래픽 유통계획 작성과 정상/위협 데이터셋을 이용한 훈련데이터셋 생성이 필요하다. 본 논문은 사이버전 훈련 시스템에 실제 환경과 같은 배경 트래픽을 제공하기 위한 트래픽 유통계획 저작과 훈련데이터셋을 생성하는 방법의 설계와 구현 결과를 제시한다. 트래픽 유통계획은 트래픽을 유통할 훈련 환경의 네트워크 토폴로지와 실제 및 모의환경에서 수집한 트래픽 속성 정보를 이용하여 저작하는 방법을 제안한다. 트래픽 유통계획에 따라 훈련데이터셋을 생성하는 방법은 단위트래픽을 이용하는 방법과 프로토콜의 비율을 이용하는 혼합트래픽 양상 방법을 제안한다. 구현한 도구를 이용하여 트래픽 유통계획을 저작하고, 유통계획에 따른 훈련데이터셋 생성결과를 확인하였다.

Design and Implementation of Cyber Warfare Training Data Set Generation Method based on Traffic Distribution Plan

Kim Yong Hyun*, Ahn Myung Kil*

ABSTRACT

In order to provide realistic traffic to the cyber warfare training system, it is necessary to prepare a traffic distribution plan in advance and to create a training data set using normal/threat data sets. This paper presents the design and implementation results of a method for creating a traffic distribution plan and a training data set to provide background traffic like a real environment to a cyber warfare training system. We propose a method of a traffic distribution plan by using the network topology of the training environment to distribute traffic and the traffic attribute information collected in real and simulated environments. We propose a method of generating a training data set according to a traffic distribution plan using a unit traffic and a mixed traffic method using the ratio of the protocol. Using the implemented tool, a traffic distribution plan was created, and the training data set creation result according to the distribution plan was confirmed.

Key words : Traffic Distribution Plan, Training Data Set, Cyber Warfare Training System, Traffic Generator

접수일(2020년 8월 31일), 수정일(1차: 2020년 10월 13일), 게재확정일(2020년 10월 22일)

★ 본 논문은 민·군 기술협력사업의 지원으로 수행된 연구임 (UM7312RD3)

* 국방과학연구소

1. 서 론

국방을 포함하여 국가 기간시설에 대한 사이버테러로 인한 피해가 날로 커지고 있으며, 민간분야에서는 금전적 목적의 사이버공격의 심각성이 날로 증가하고 있다. 사이버전이 현실화됨에 따라 사이버전을 위한 기술 및 전문 인력에 대한 수요가 급격히 증가하고 있다. 전문 인력 양성을 위하여 민간뿐 만아니라 군에서도 사이버전 훈련장을 구축하고 운영 중이다. 하지만 사건사례 중심의 콘텐츠를 구축하여 실습하는 형태가 보통이며, 실제와 같은 훈련환경 구축은 제한적이다. 또한 사이버전 기술 검증을 위한 공용의 기준 데이터셋이 없어 기술간 비교평가가 어려운 상황이다.

다양한 시나리오에 맞춰 훈련을 수행하고 사이버보안 기술의 검증 및 기술간 비교검증을 위해 사이버전 정상/위협 데이터셋 구축도구와 훈련/검증용 데이터셋 구축이 필요하다. 훈련/검증용 데이터셋을 통해 훈련 및 검증 담당자가 적시적소에 필요한 시나리오 개발이 가능하고, 실제 환경에서 수집된 트래픽을 기반으로 정상 행위를 발생시켜 현실감 있는 훈련 및 검증 환경 제공이 가능하다.

실제와 같은 사이버전 훈련 및 검증 환경을 구축하기 위해 국방과학연구소는 실가상 환경에서 정상/위협 데이터들을 수집 분석하여 실제 환경과 유사하게 트래픽을 자동 생성하는 도구 개발에 대한 연구를 수행하였다. 사이버전 훈련 시스템에 트래픽을 제공하기 위해서는 트래픽을 유통하기 전에 트래픽 유통계획 작성과 정상/위협 데이터셋을 이용한 훈련데이터셋 생성이 필요하다. 본 논문에서는 트래픽 유통계획 저작과 훈련데이터셋을 생성하는 방법을 설계하고, 구현하고자 한다.

본 논문의 구성은 다음과 같다. 2장에서는 관련연구로 기존의 사이버전 훈련시스템과 트래픽발생기를 기술한다. 3장은 제안하는 트래픽 유통계획 저작과 훈련데이터셋 생성 방법을 상세히 설명한다. 4장은 제안한 방법의 구현결과와 실제 저작된 트래픽 유통계획과 생성된 훈련데이터셋 결과를 제시하고, 5장에서 결론을 맺는다.

2. 관련 연구

2.1 사이버전 훈련 시스템

실전적 훈련을 위한 사이버전 훈련 시스템은 크게 구성 모의(constructive simulation)와 실가상 환경 모의(Live-Virtual simulation) 방식으로 나눌 수 있다[1]. 구성 모의 방식은 모의 사용자가 모의 체계를 사용한다. 실가상 환경 모의 방식은 실제의 사용자가 실체계와 상호작용하는 것이다[2]. 구성 모의의 경우, 실제 구축된 복잡한 시스템을 추상화 모델링(abstract modeling)을 통해 표현한 후, 수립한 방책의 효과도(measure of effectiveness)를 검증하거나 방책의 실행 순서를 훈련하는데 사용한다. 구성 모의 훈련 시스템의 경우, 추상화된 사이버 위협의 효과를 수치상으로 모의하기 때문에 훈련 대상에게 실제와 동일한 경험을 주기 어렵다. 보다 효과적인 훈련을 위해 실제 물리 장비를 연결하여 사이버 위협의 영향이 어떤 식으로 영향을 미치는지 훈련생에게 전달하는 방식을 혼용하기도 하나 사이버 위협의 부가 효과(side effect)를 재현할 수는 없다[3]. 이 방식의 시스템 구축 예로는 DARPA(Defense Advanced Research Project Agency)에서 개발한 LARIAT(Lincoln Adaptable Real-time Information Assurance Testbed), 미 Scalable Network Technologies사가 개발한 사이버 방어 훈련 시스템인 NDTrainer가 있다[2,4].

두 번째 방식인 실가상 환경 모의 시스템은 현실적인 사이버 훈련을 위해 실제의 사이버 환경을 가상화 기술을 이용하여 동적으로 생성한 유사한 테스트베드 환경에서 실제 시스템 보안 담당자가 사이버 위협에 대해 수립된 방책의 기법 및 절차를 훈련할 수 있도록 지원한다. 이러한 방식의 사이버 위협에 대한 방어 훈련은 일반적으로 공격을 담당하는 레드팀, 방어를 담당하는 블루팀, 훈련에 대한 모니터링을 수행하는 화이트팀으로 구성된다. 시스템 운용은 특정 위협이 적용된 상태의 테스트베드를 제공하여 블루팀으로 하여금 이를 분석하고 조치를 취하게 하거나 훈련 내용에 의해 미리 정해진 위협을 레드팀이 수동으로 발생시켜 이에 대한 훈련자의 대응 능력을 화이트팀이 평가하는 형태로 이루어진다. 대표적이 사례는 미국 DARPA가 개발한 국가 사이버전 시험장(NCR, National Cyber Range)과 이스라엘의 사이버 보안 에뮬레이션 훈련 센터인 사이버짐, KISA의 시큐리티짐(Security-Gym)

등이 있다[2,4].

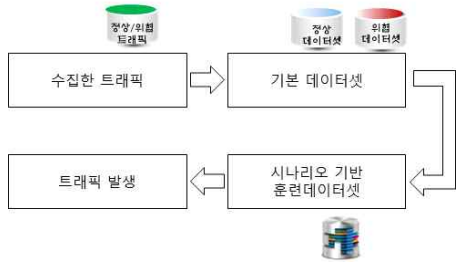
2.2 트래픽 생성기

트래픽 생성기는 일반적으로 시스템의 네트워크 부하 시험과 사이버 위협 대처 능력을 보기 위한 목적으로 개발되며 하드웨어와 소프트웨어 형태로 존재한다. 하드웨어 형태를 대표하는 IXIA사나 SPIRENT사는 기존의 IDS/IPS 등의 Inline 정보보호 장비에 대한 성능계측 수준에서 훈련환경에 적합하도록 시나리오 기반의 위협 트래픽 발생, 가상화 기반 기술 등 새로운 기능을 개발하여 제품화하고 있다[3]. 또한 네트워크 대역폭이 증가함에 따라 트래픽 발생장비의 대역폭도 계속 증가하는 방향으로 발전하고 있다.

특정한 용도에 관계없이 트래픽 생성기의 주요 목표는 최종 사용자의 현실적인 행위를 모방하는 다양한 트래픽 유형 및 특성을 생성하는 것이다. Vishwanath의 연구[5]에 따르면, 네트워크 트래픽은 확률적 생성(stochastic generation), 실제 네트워크 트래픽 복제(replication of production network traffic), 테스트 네트워크의 응용 프로그램에 대한 인스트럭션(instructions) 리스트 사용과 같은 세 가지 방법으로 생성될 수 있다. Botta et. al.[6]은 네트워크 트래픽 생성기가 작동하는 계층에 따라 네트워크 트래픽 생성기를 세 가지로 구분하였다. 응용 프로그램이 생성하는 네트워크 트래픽의 측면에서 특정 네트워크 응용의 행위를 에뮬레이션하는 응용 프로그램 수준 트래픽 생성기, 현실적인 트래픽의 복제가 플로우 레벨(예 : 전송 패킷 수 및 전송 바이트 수, 플로우 지속 시간)에서만 요청되는 경우에 사용되는 플로우 수준 트래픽 생성기, 패킷의 출발 간 시간(Inter Departure Time, IDT)과 패킷 크기(Packet Size, PS)를 기반으로 하는 네트워크 트래픽 생성기를 참조하는 패킷 수준 트래픽 생성기이다. 현재 트래픽생성기의 대부분은 패킷 생성기에 해당한다.

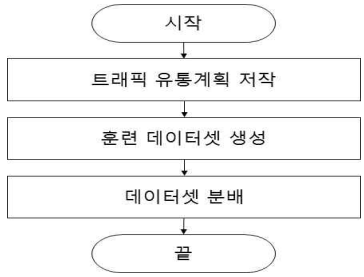
3. 트래픽 유통계획 저작 및 훈련데이터셋 생성 방법 설계

사이버전 훈련 시스템에 사용하는 배경 트래픽을 만들기 위해서는 (그림 1)과 같은 순서의 데이터셋 구축 과정이 필요하다. 실제 및 모의 네트워크 환경에서 수집한 트래픽, 트래픽의 속성을 추출한 기본 데이터셋, 훈련 시나리오에 맞춰 변환한 시나리오 기반 훈련 데이터셋이 필요하다.



(그림 1) 데이터셋 구축 순서

본 논문은 사이버전 훈련 시스템의 네트워크 토폴로지 정보를 이용한 트래픽 유통계획 수립 및 트래픽 생성 방법에 관한 것이다. 트래픽 유통계획을 위해서는 사전에 실/모의환경에서 수집하여 구축한 기본 데이터셋이 구축되어야 한다. 사이버전 훈련 시스템에 시나리오 기반으로 트래픽을 발생시키기 위해서는 (그림 2)와 같이 메타데이타 형태로 구축된 기본 데이터셋과 사이버전 훈련 시스템의 네트워크 토폴로지 정보를 이용하여 트래픽 유통계획을 저작하고, IP, domain name, 사용자 정보 등 트래픽 내에서 여러 변동 가능한 요소를 수정 요소로써 반영하여 훈련데이터셋을 생성해야 한다. 생성된 훈련데이터셋은 트래픽을 발생시키는 장비에 분배하여 사이버전 훈련 시스템에 트래픽으로 제공된다.

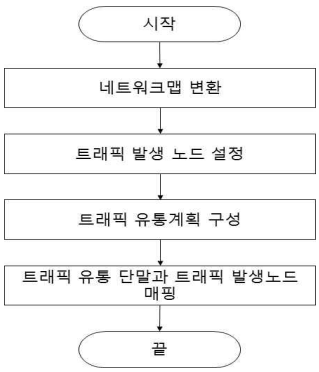


(그림 2) 트래픽 발생을 위한 준비과정

본 장에서는 트래픽 유통계획 저작과 훈련데이터셋 생성방법에 대한 설계내용을 기술한다.

3.1 트래픽 유통계획 저작 방법

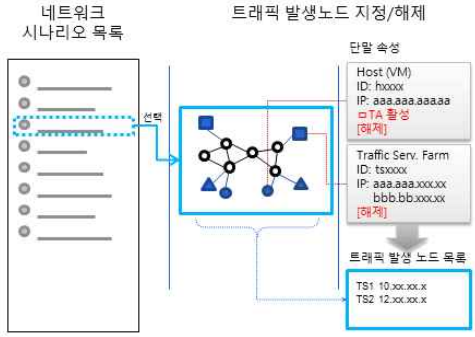
트래픽 유통계획 저작은 트래픽을 유통하게 될 훈련환경의 네트워크 토폴로지와 네트워크에 유통할 트래픽을 활용한다. 시스템 관점에서 트래픽 유통계획을 저작하는 순서는 (그림 3)과 같다. 트래픽 유통계획 저작을 위해 사이버전 훈련 시스템을 지정하고, 네트워크 토폴로지를 불러온다. 네트워크 토폴로지의 노드로부터 트래픽을 생성하고 수신하는 트래픽 서버그룹 및 트랙픽 사용자 그룹을 생성하고, 트래픽 에이전트를 선택한다. 그런 다음 선택한 네트워크 토폴로지에서 사용할 트래픽을 기본 데이터셋 DB로부터 선택한다. 트래픽 발생 양상에 따라 단위 트래픽 템플릿이나 혼합 트래픽을 선택하고, 네트워크 토폴로지에서 식별한 트래픽 발생 노드와 트래픽 템플릿의 트래픽 유통단말을 매핑한다.



(그림 3) 트래픽 유통계획 저작 순서

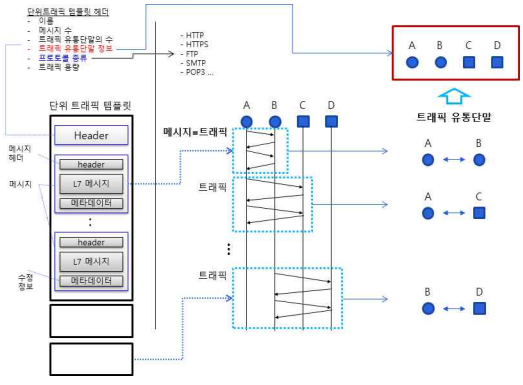
네트워크 토폴로지에서 트래픽 발생 노드를 설정하는 방법은 (그림 4)와 같다. 먼저 훈련에 사용할 네트워크 토폴로지를 네트워크 토폴로지 목록에서 선택하고, 도시된 네트워크 맵에서 트래픽을 발생하거나 사용하는 단말을 설정한다. 트래픽 발생 노드는 트래픽 발생기, 트래픽 에이전트로 구분하며, 지정된 노드는 트래픽 발생 노드로 저장하여 관리한다. 선택된 트래픽 발생 노드는 네트워크 토폴로지에 사용하는 트래픽의 유통단말과의 매핑에 사용된다.

네트워크 토폴로지에서 사용할 트래픽 양상선택은 단위트래픽 템플릿을 이용하는 방법과 혼합트래픽을 이용하는 방법이 있다. 단위트래픽 템플릿을 이용하는



(그림 4) 트래픽 발생 노드 설정

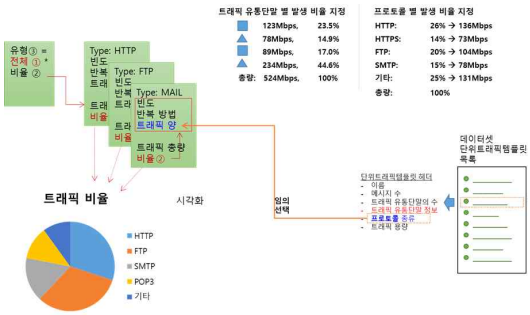
방법은 (그림 5)와 같이 기본 데이터셋 DB에서 사용할 단위트래픽 템플릿을 선택한다. 단위 트래픽 템플릿은 헤더와 다수의 메시지로 구성되며, 각 메시지는 메시지헤더, L7 메시지, 메타데이터로 구성이 된다. 단위 트래픽 템플릿의 헤더에는 이름, 메시지 수, 트래픽 유통단말의 수, 트래픽 유통단말 정보, 프로토콜 종류, 트래픽 용량이 포함되어 있다. 헤더 내의 트래픽 유통단말 정보를 통해 트래픽 유통단말을 파악한다. 또한 메시지 수와 트래픽 종류 정보를 통해 (그림 5)와 같이 트래픽 유통과정을 재현할 수 있다. 트래픽 유통계획은 트래픽 발생방법을 설정함으로써 저작이 완료되고, 훈련데이터셋 생성 과정을 통해 실체화된다.



(그림 5) 트래픽 유통계획에 단위트래픽 템플릿을 이용하는 방법

혼합트래픽 양상은 단위 트래픽을 이용하여 실제 세션을 반영한 배경 트래픽을 생성하는 방법이다. 혼

합트래픽은 배경트래픽으로써 실제 특정 트래픽 세션이 중요한 것이 아닌 대역폭을 채우는 개념으로 실제 트래픽을 유통시킬 수 있는 방법에 더 초점을 맞춘 트래픽 발생 방식이다. 혼합트래픽 양상을 이용하는 방법은 (그림 6)과 같이 전체 혼합 트래픽의 발생량을 설정하고, 프로토콜 종류와 프로토콜 비율을 설정하면 저작은 완료되고, 혼합트래픽 시뮬레이터를 구동함으로써 트래픽 유통계획은 실체화된다. 혼합 트래픽 시뮬레이터는 혼합트래픽 양상 설정에 맞게 트래픽을 자동으로 채워주는 것으로, 데이터셋에 존재하는 단위 트래픽을 프로토콜 별로 임의 선택하여 원하는 대역폭 만큼을 채워서 트래픽을 구성한다.

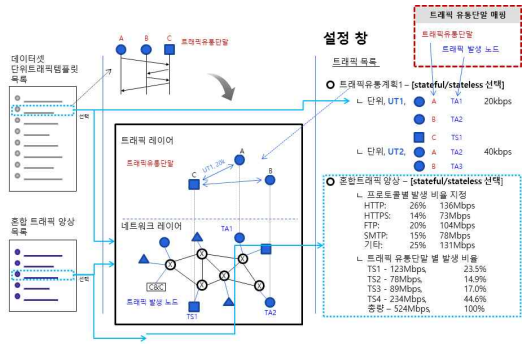


(그림 6) 트래픽 유통계획에 사용하는 트래픽 중 혼합트래픽 양상을 이용하는 방법

트래픽 유통계획 저작의 마지막 단계는 어떤 트래픽을 네트워크의 어느 단말에서 발생시키고 어느 단말로 흘러가게 할 것인가를 결정하는 것으로, (그림 7)과 같이 단위 트래픽 별로 지정된 트래픽 유통단말을 네트워크 토폴로지의 트래픽 발생 노드와 매핑하여 주는 과정이다. 하나의 트래픽 발생 노드가 여러 트래픽을 발생시킬 수 있으므로, 다수의 트래픽 유통단말과 매핑 될 수 있다. 한편 혼합 트래픽 양상은 트래픽 발생기와 매핑이 되며, 가용한 트래픽 발생기 포트로 설정해 준다.

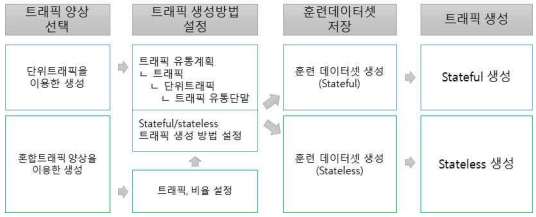
3.2 훈련데이터셋 생성 방법

트래픽 유통계획에 따라 트래픽을 생성하는 방법은 (그림 8)과 같이 단위트래픽을 이용하는 방법과 프로토콜의 비율을 이용하는 혼합트래픽 양상 방법이 있



(그림 7) 트래픽 유통단말과 트래픽 발생 노드 매핑 과정

다. 각각은 트래픽 양상 선택, 트래픽 생성방법 설정, 훈련데이터셋 저장, 트래픽생성 단계로 진행한다. 선택한 방법에 따라 기본 데이터셋 DB에서 사용할 단위트래픽을 선택하고 이를 DB에 저장한 후 훈련데이터셋 생성 과정을 통해 PCAP 파일로 저장한다. 트래픽 생성은 단위 트래픽 템플릿을 이용하여 순서에 맞게 주고 받을 수 있는 트래픽인 Stateful 방식과 백그라운드 트래픽으로써, 트래픽발생장치 포트간에만 유통되는 트래픽인 Stateless 방식이 있다.

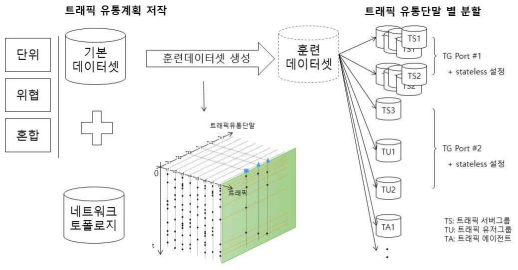


(그림 8) 트래픽 유통계획에서 훈련데이터셋 생성 방법

트래픽 템플릿 형태로 저장되어 있는 트래픽 유통계획은 훈련데이터셋 생성 과정을 통해 패킷 형태로 저장된다. 훈련데이터셋은 단위 트래픽 단위로 PCAP 파일로 저장되며 고유한 Group ID를 부여한다. 각 트래픽 유통단말은 하나의 단위트래픽에 대한 훈련데이터셋을 나누어 갖게 되며, 이를 통해 이번에 송신하여야 하는 패킷과 이번에 수신하여야 하는 패킷의 정보를 확인할 수 있으며, 다음 패킷을 송신할 수 있다.

훈련데이터셋 생성은 하나의 단위 트래픽에 대하여 L4-L7 트래픽을 발생시키기 위한 순서를 정의하고 있는 트래픽을 생성해 주는 역할을 한다. 훈련데이터셋 생성에서는 트래픽 유통계획의 수정 요소를 포함하는 실제 유통될 트래픽을 생성하여야 하며, L4, L3(, L2) 헤더를 포함하는 실제 패킷의 유통을 위하여 패킷을 준비하는 기능을 제공한다. 따라서 훈련 데이터셋은 각 트래픽 유통 단말로 배포되어야 하며, stateful 트래픽의 경우 보낼 부분과 받을 부분을 모두 가지고 있어야 하고, stateless 트래픽의 경우 받을 부분을 제외하고 보낼 부분만 트래픽 유통단말로 배포한다. 훈련데이터셋은 여러 트래픽이 하나의 파일로 묶여 배포되는 것이 아닌, 순서에 연관된 트래픽 별로 분할되어 파일의 형태로 배포되어야 다수의 트래픽에 대한 발생을 효과적으로 제어할 수 있다.

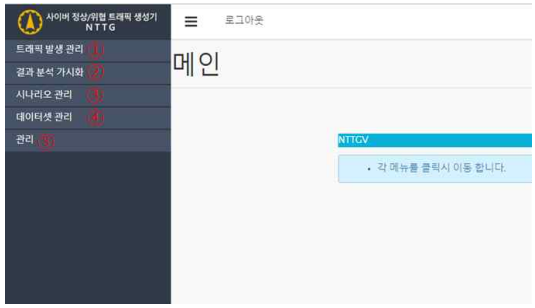
저작된 트래픽 유통계획을 기반으로 훈련데이터셋을 생성하고 이를 분할하는 과정을 다시 정리한 것이 (그림 9)이다. (그림 9)의 과정을 거쳐 생성된 훈련데이터셋은 PCAP 파일로 저장되며, 제어채널을 이용하여 트래픽유통계획과 함께 각 트래픽 유통단말로 전달되어 트래픽을 발생한다.



(그림 9) 훈련데이터셋 생성 및 분할 방법

4. 트래픽 유통계획 기반 훈련데이터셋 생성 방법 구현결과

사이버 정상/위협 트래픽 발생기(NTTG, Normal & Threat Traffic Generation and Interface)는 실제 네트워크 환경에서 수집한 데이터셋과 시뮬레이션 기반의 의도된 데이터셋을 활용하여 다양한 훈련 시스템의 환경에 맞는 훈련데이터셋을 생성하고 이에 의한 트래픽 유통 및 유통 결과 분석 기능을 제공하는 소프트웨어이다. 국방과학연구소에서 개발한 NTTG는 사이버 훈련 시스템과 일체감을 높이기 위해 API 기반 트래픽 유통계획 저작 기능과 ① 사이트 별로 가상환경 초기화, 훈련 데이터셋 배포 및 트래픽 유통 시작, 중지 등의 제어를 통해 트래픽 발생을 관리하는 트래픽 발생 관리 기능과 ② 발생한 트래픽의 결과를 분석하여 가시화 제공하는 결과 분석 가시화 기능, 그리고 ③ 트래픽을 유통시키는 대상 환경인 훈련용 시스템의 네트워크 토폴로지 및 트래픽 유통계획을 관리하는 시나리오 관리기능, ④ 정상/위협 데이터셋 관리를 위한 데이터셋 관리 기능, ⑤ 소프트웨어 사용을 위한 계정 및 조직을 관리하는 관리기능으로 구성된다.



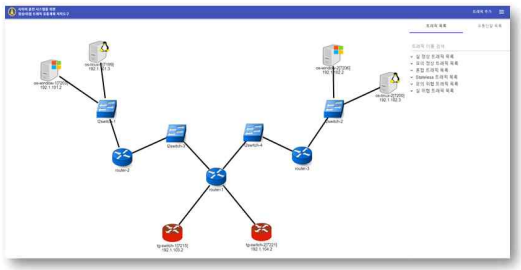
(그림 10) 사이버 정상/위협 트래픽 생성기 구현결과

본 논문에서 기술한 트래픽 유통계획 저작 및 훈련데이터셋 생성 방법에 대한 설계는 NTTG의 일부 기능으로 구현하였다. 구현 도구를 활용하여 실제 트래픽 유통계획 저작과 훈련데이터셋을 생성하고, 생성된 훈련데이터셋을 이용하여 트래픽을 발생시킴으로써 트래픽 유통계획 저작의 정확도를 확인하고자 한다.

4.1 트래픽 유통계획 저작 결과

본 절에서는 사이버전 훈련시스템에서 사용할 훈련 시나리오를 기반으로 트래픽 유통계획을 저작하고, 훈련데이터셋을 생성한 결과를 제시한다. 훈련 시나리오에서 사용하는 네트워크 토폴로지는 (그림 11)과 같이 윈도우 시스템 2대, 리눅스 시스템 2대, 트래픽 발생기 2대, 스위치 4대, 라우터 3대로 구성된 소규모 네트워크이다. 이 네트워크에 유통할 트래픽은 단위트래픽 16

개와 11개 프로토콜로 구성된 혼합 트래픽을 사용한다.



(그림 11) 훈련 데이터셋 PCAP의 패킷 확인

트래픽 유통계획을 저작하기 위해서 사이버전 훈련 시스템의 네트워크 토폴로지 목록에서 하나를 선택하여 NTTG의 트래픽 생성/분석 도구의 형식으로 변환한다. 변환된 네트워크 토폴로지 정보는 데이터베이스 테이블에 저장되며, <표 1>과 같이 추가로 트래픽 발생 노드를 설정하였다.

<표 1> 트래픽 발생노드 설정

노드명	노드 id	역할	트래픽 발생노드
윈도우1	719	TrafficAgent	os-window-1
윈도우2	720	TrafficAgent	os-window-2
리눅스1	717	TrafficAgent	os-linux-1
리눅스2	718	TrafficAgent	os-linux-2
TG1	724	ServerFarm	tg-switch-1
TG2	725	ServerFarm	tg-switch-2

네트워크 토폴로지에 유통시킬 단위 트래픽은 16개를 사용하였다. 실제 네트워크 환경에서 수집한 단위 트래픽이며, 사용한 프로토콜은 HTTP 5개, SMTP 2개, POP3 2개, FTP 2개, SMB 5개이다. 각 단위 트래픽의 트래픽 유통단말은 2대이다. 혼합 트래픽 양상은 <표 2>와 같이 11가지 프로토콜을 선정하였다. 11가지 프로토콜은 모의환경에서 수집한 데이터셋이며, 각각에 대해 크기, 시작시간, 종료시간까지 설정하였다. 혼합 트래픽의 트래픽 유통단말도 2대이다.

트래픽별 트래픽 유통단말과 네트워크 토폴로지의 트래픽 발생 노드 간 매핑결과는 <표 3>과 같다. 트래픽 발생노드는 트래픽에 중복되어 설정되어 있음을 확인할 수 있다.

<표 2> 혼합 트래픽 정보

트래픽 양상	혼합 트래픽 프로토콜	크기 (kbps)	시작시간(s)	종료시간(s)
혼합 트래픽	Game	500	2	4
	HTTP		4	6
	Database		6	8
	Print		8	10
	Login		10	12
	SMB		12	14
	Voice		14	16
	FTP		16	18
	SMTP		18	20
	POP3		20	22
	HTTP	1000	10	13

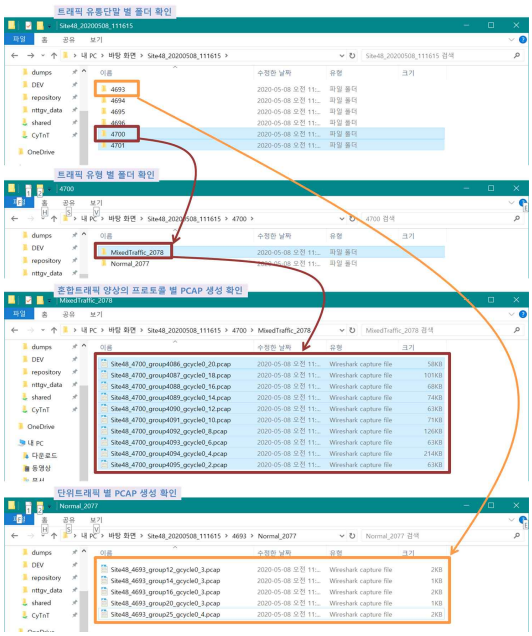
<표 3> 트래픽 별 트래픽 유통단말-트래픽 발생 노드 매핑 결과

트래픽	트래픽 발생 노드(src)	트래픽 발생 노드(dst)
단위-1	os-window-1	os-window-2
단위-2	os-window-1	os-linux-1
단위-3	os-window-1	os-linux-2
단위-4	os-window-2	os-linux-1
단위-5	os-window-2	os-linux-2
단위-6	os-linux-1	os-linux-2
단위-7	tg-switch-1	tg-switch-2
단위-8	tg-switch-1	os-window-1
단위-9	os-window-2	tg-switch-1
단위-10	tg-switch-1	os-linux-1
단위-11	os-linux-2	tg-switch-1
단위-12	tg-switch-2	tg-switch-1
단위-13	tg-switch-2	os-window-1
단위-14	os-window-2	tg-switch-2
단위-15	tg-switch-2	os-linux-1
단위-16	os-linux-2	tg-switch-2
혼합	tg-switch-1	tg-switch-2

4.2 훈련데이터셋 생성 결과

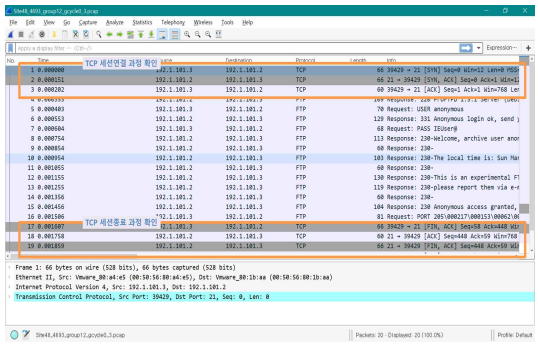
훈련데이터셋은 트래픽 생성/분석 매니저의 제어패널에서 [훈련 데이터셋 생성] 버튼을 클릭하여 생성한다. 훈련 데이터셋 생성이 완료되면, [FTP 접속 도구]에서 [PCAP 생성 폴더: nttgv_data/tmp_pcap_files]에 접근하여, 훈련 데이터셋 생성 시간에 대한 사이트의 [훈련 데이터셋 root 폴더] (폴더명: Site##_yyMMdd_hhmmss 형식)가 생성된다. 생성한 훈련데이터셋의 root 폴더에 [트래픽 유통단말 별 폴

더] (폴더명: 트래픽 유통단말의 노드 번호)가 생성된다. 훈련데이터셋 root 폴더 하위의 각 트래픽 유통단말의 폴더에는 트래픽 유통단말에 지정된 트래픽 유통계획에 대한 [트래픽 별 폴더] (폴더명: Normal_###, Mixed_### 등)가 생성된다. 트래픽 유통단말에 대한 단위트래픽 별 폴더에는 단위트래픽 별 [훈련 데이터셋 PCAP 파일]이 트래픽 유통계획에 맞게 생성되어 있다. 혼합트래픽에 대한 PCAP은 트래픽 발생기 노드에 배정되므로, 트래픽 발생기 폴더 하위에 혼합 트래픽 양상 폴더 (Mixed_###) 내에 혼합트래픽이 단위 트래픽 (프로토콜) 별로 PCAP 파일로 생성된다. (그림 12)에 트래픽 유통단말별로 단위트래픽 별 PCAP 파일 생성과 혼합트래픽 양상의 프로토콜별로 PCAP 파일이 생성됨을 확인할 수 있다.



(그림 12) 훈련 데이터셋 PCAP 생성 확인

Stateful 트래픽에 대하여, 해당 트래픽 폴더 (Normal_###)에 생성된 단위트래픽 PCAP 파일을 Wireshark로 열어, TCP Handshaking 과정 (SYN, SYN/ACK, ACK)과 종료과정 (FIN, FIN/ACK, FIN, FIN/ACK)이 단위 트래픽의 앞/뒤에 각각 포함되어 있음을 (그림 13)을 통해 확인할 수 있다.



(그림 13) 훈련 데이터셋 PCAP의 패킷 확인 (단위트래픽)

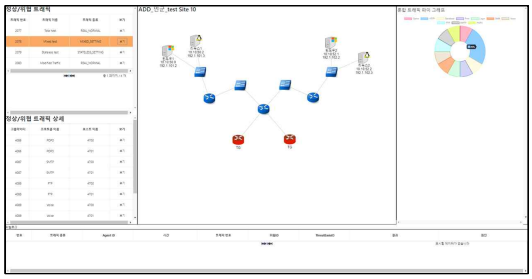
4.3 트래픽 발생 결과

생성한 훈련데이터셋을 이용하여 사이버전 훈련 시스템에 실제 트래픽을 발생시킨 결과를 분석하여 트래픽 유통계획 저작의 정확도를 확인하였다.

2077번 트래픽 번호에 대한 단위 트래픽들 중 하나 (이름 : unit 1)에 대하여 보기 버튼을 클릭하여 트래픽 유통경로가 가시화 된 것을 확인한다. 가시화 하고자 하는 호스트(윈도우1, ip:192.1.101.2)를 클릭하여 훈련 데이터셋과 실제 발생한 트래픽을 수집한 결과가 가시화 된 것을 확인할 수 있다. (그림 14)는 stateful 트래픽 의 유통결과를 가시화 화면을 나타낸 것이다.

(그림 14) Stateful 트래픽 유통결과 가시화

2078번 트래픽 번호에 따른 프로토콜별 발생 비율에 대한 그래프가 가시화 된 것을 확인한다. (그림 15)와 같이 혼합 트래픽의 비율을 파이 그래프 형태로 제시되는 결과를 통해 저작된 것과 동일함을 확인할 수 있다.



(그림 15) 혼합 트래픽 발생 비율 파이 그래프

5. 결 론

국방과학연구소는 현실감 있는 사이버전 훈련 환경을 구축하기 위해 훈련환경에 필요한 정상/위협 데이터들을 수집 분석하여 실제 환경과 유사하게 트래픽을 자동 생성하는 도구 개발에 대한 연구를 진행하였다. 본 논문에서는 사이버전 훈련 시스템에 실제 환경과 같은 배경 트래픽을 제공하기 위하여 트래픽 유통계획 저작과 훈련데이터셋을 생성하는 방법을 설계하고, 구현하였다. 트래픽 유통계획은 트래픽을 유통할 훈련 환경의 네트워크 토폴로지와 실제와 모의환경에서 수집한 트래픽 속성 정보를 이용하여 저작하는 방법을 제안하였다. 트래픽 유통계획에 따라 트래픽을 생성하는 방법은 단위트래픽을 이용하는 방법과 프로토콜의 비율을 이용하는 혼합트래픽 양상 방법을 제안하였다. 본 논문에서 제안한 방법은 실제 도구로 구현하였으며, 구현결과를 통해 유통계획에 따른 훈련데이터셋 생성 결과를 확인하였다.

본 논문에서 설계한 트래픽 유통계획 기반 훈련데이터셋 생성 방법은 사이버전 훈련 시스템의 네트워크 토폴로지를 직접 활용하여 기 구축한 데이터셋을 훈련 시나리오 상황에 맞춰 다양하게 적용할 수 있는 기반을 제공한다. 향후 대규모 네트워크 토폴로지를 포함하여 다양한 네트워크 환경에서의 훈련데이터셋을 생성하고, 실제 트래픽 발생결과를 분석하여 정확도를 높이는 연구를 진행하고자 한다.

참고문헌

[1] 안명길, 김용현. “사이버전사의 훈련을 위한 시스템 구축 방안 연구”, 한국정보보호학회 논문지, Vol.26, No.2, pp.533-540, 2016.

[2] 조완수, “사이버전 훈련을 위한 사이버 레인지 기술 동향”, 국방과학기술플러스, Vol. 234, 2016.

[3] 홍수연, 김광수, 김태규, “사이버전 훈련을 위한 상태 저장 트래픽 발생 Architecture 설계 및 구현”, 한국군사과학기술학회지, 제23권, 제3호, pp. 267-276, 2020.

[4] 조완수, “사이버전 훈련도구 분석”, 국방과학연구소, ADDR-412-160879, 2016.

[5] Vishwanath, K., “Realistic and responsive network traffic generation”, ACM SIGCOMM 2006, Vol. 36, No. 4, pp. 111 - 122, 2006.

[6] Botta, A., Dainotti, A., Pescape, A., “A tool for the generation of realistic network workload for emerging networking scenarios”, Computer Networks, Vol. 56, No. 15, pp. 3531-3547, 2012.

〔저자소개〕

김 용 현 (Yong Hyun Kim)
1993년 2월 광운대학교 전자공학과 학사
1995년 2월 광운대학교 전자공학과 석사
2013년 2월 광운대학교 전자통신공학과 박사
1995년~현재 국방과학연구소 책임연구원
email : yonghyunkim@add.re.kr

안 명 길 (Myung Kil Ahn)
1997년 2월 충남대학교 정보통신공학과 학사
2003년 2월 서강대학교 컴퓨터공학과 석사
2017년 중앙대학교 전자공학과 박사 수료
2006년~현재 국방과학연구소 책임연구원
email : happyahn@add.re.kr