국민대학교
소프트웨어융합대학
소프트웨어학부

캡스톤 디자인 I
종합설계 프로젝트

프로젝트 명	생성형 침입방지 기술 기반 보안플랫폼
팀 명	34조
문서 제목	중간보고서

Version	1.0
Date	2024-04-01

팀원	엄석현 (팀장)
	김태경
	김태윤
	박준서
지도교수	윤명근 교수

CONFIDENTIALITY/SECURITY WARNING

문서 정보 / 수정 내역

Filename	중간보고서-생성형 침입 방지 기술 기반 보안 플랫폼.docx
원안작성자	김태경, 김태윤, 박준서, 엄석현
수정작업자	김태경, 김태윤, 박준서, 엄석현

수정날짜	대표수정자	Revision	추가/수정 항목	내 용
2023-03-27	엄석현	1.0	최초 작성	

목 차

1 프로젝트 목표

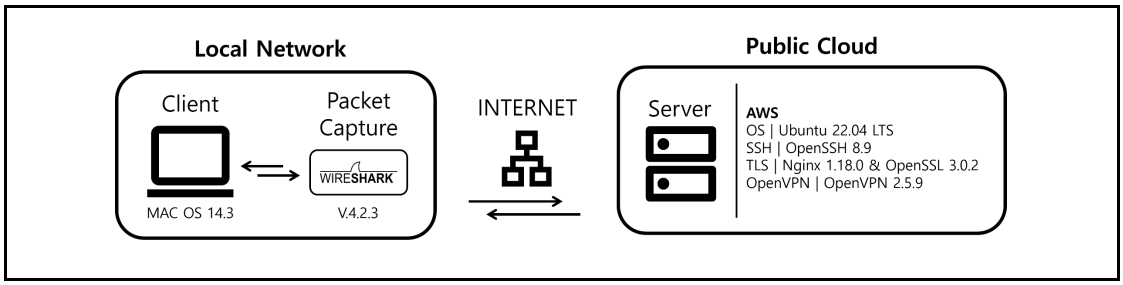
생성형 침입 방지 기술 (GIPS: Generative Intrusion Prevention on data Stream)[1] 알고리즘을 기반으로 자동화된 데이터 분석 및 패턴 인식 알고리즘을 개선합니다. 실시간으로 스트리밍 되는 데이터의 패턴을 식별해, 추출한 공통 시그니처를 Yara Rule 형태로 자동 생성해 주는 기능을 구현하여 다양한 분석 툴에서 활용할 수 있도록 합니다. 또한 다양한 데이터 소스로부터 손쉽게 분석이 가능하도록 시각화하여 보여주는 Web 플랫폼을 구성하여 증가하고 있는 제로데이 공격과 같은 보안 위협을 식별할 수 있는 플랫폼을 구축하고자 합니다. 나아가, 국내 최고 수준인 보안 전문 기업/기관인 (주)시큐아이, (주)윈스, (주)누리랩, KISTI(한국과학기술정보연구원) 중 1개 기업 /기관과 협력하여 실제 산업 현장에서 필요로 하는 기술을 연계하여 개발하는 것을 목표로 하고 있습니다.

2 수행 내용 및 중간결과

2.1 계획서 상의 연구내용

Wireshark와 같은 패킷 분석 도구에서는 서비스가 사용중인 프로토콜을 식별할 때 포트 번호를 이용합니다. 그렇기에 기존의 Well-Known Port에서 다른 Port 번호로 변경하여 사용할 경우 분석도구에서는 실제 프로토콜을 식별할 수 없습니다. 이러한 상황에서는 패킷의 Payload를 직접 확인하여 프로토콜을 식별해야 합니다. 그러나 최근에는 보안을 강화하기 위해 TLS와 같은 암호화 프로토콜을 사용하여 패킷의 Payload를 암호화 하여 전송하기 때문에 Payload를 확인하는 것이 불가능합니다. 따라서 패킷의 헤더 정보나 패킷의 플로우 정보 들을 이용해서 암호화 트래픽 환경에서 서비스 프로토콜을 식별하는 연구 진행을 목표로 삼았습니다.

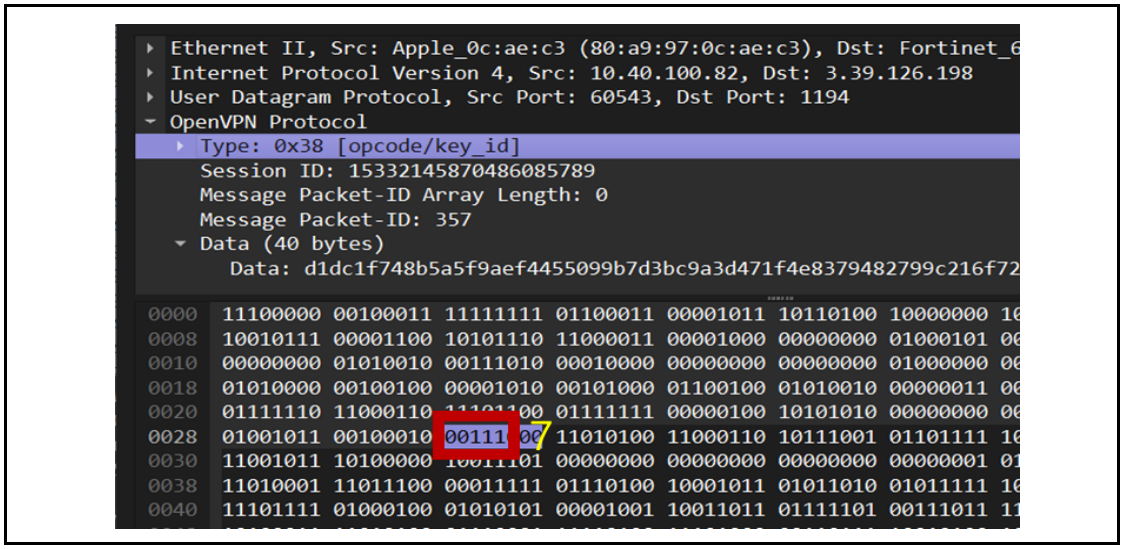
2.2 수행내용



<실험 환경>

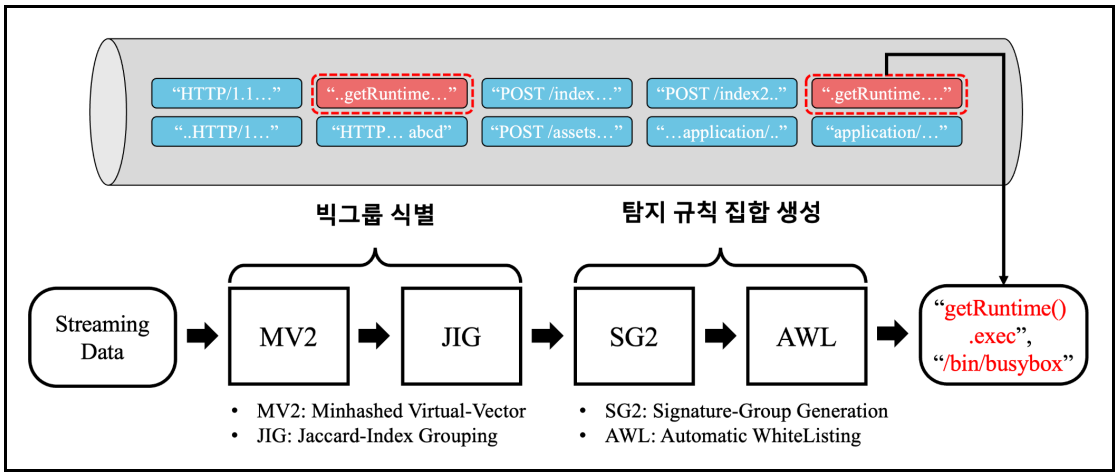
퍼블릭 클라우드(AWS) 기반 테스트 베드 환경을 구축하여 통신 실험을 진행하였습니다. WireShark를 이용해 SSH, TLS, OpenVPN 프로토콜을 사용하는 통신을 분석하였고, 각 프로토콜의 Well-Known Port를 사

용했을때와 Custom Port로 변경하였을 때 Wireshark에서 프로토콜이 식별 가능한지 비교 분석하였습니다.



<OpenVPN 프로토콜 식별 예시>

OpenVPN의 경우 Payload가 암호화 되어있더라도, 헤더의 OP Code 필드에서 Handshake를 위해서 일정 규칙이 정해져있기에 규칙적인 값이 발생하여 OpenVPN임을 식별할 수 있음을 확인하였습니다.



<GIPS 기술 전체 구조도>

생성형 침입 방지 기술 (GIPS:Generative Intrusion Prevention on data Stream) 논문을 기반으로 하여 빅데이터로 부터 반복되는 패턴을 식별할 수 있는 알고리즘을 구현하였습니다. 구현한 알고리즘을 검증하기 위해 악성과 양성의 사물인터넷(IoT) 패킷이 모두 있는 CIC IoT 데이터셋과 공개 웹 서버 로그 데이터 셋을 사용하여 작성한 알고리즘이 논문의 결과와 비교하여 유의미한 효과를 나타내는지에 대하여 테스트하였습니다.

또한 기존 알고리즘 성능을 개선하기 위해 MV2 모듈에 적용되어있는 기존 청킹 AE 알고리즘[2] (CDC)을 변경해보았습니다. CDC 알고리즘은 많은 장점을 가지고 있지만, 형태소(의미) 단위로 청킹이 되지 않는 한계점이 있어 이를 극복해보고자 Protocol-Aware하게 분석할 수 있는 모듈을 시범 개발해보았습니다. 다양한 기법을 직접 구현하여 프로토콜 별 성능을 비교하며 최적의 기법을 찾고자 합니다.

3 수정된 연구내용 및 추진 방향

3.1 수정사항

암호화된 트래픽에서 어플리케이션 식별에서 생성형 침입 방지 기술로 주제 변경으로 인해 도메인 지식을 새롭게 추가해야하기 때문에 지도 교수님과 의견을 나눈 후 논문 리딩이나 발표 영상들을 활용하여 빠르게 도메인 지식을 습득할 것입니다.

기존 정보보호 연구실에서 가지고 있는 침입 방지 기술인 생성형 침입 방지 기술 (GIPS: Generative Intrusion Prevention on data Stream) 알고리즘을 기반으로 자동화된 데이터 분석 및 패턴 인식 알고리즘을 개선(구현)하여 실시간으로 스트리밍 되는 데이터의 패턴을 식별해, 추출한 공통 시그니처를 Yara Rule 형태로 자동 생성해주는 기능을 구현하여 다양한 분석 툴에서 활용할 수 있도록 합니다. 그렇기 때문에 기술을 이해하는데 필요한 확률적 알고리즘과 정규분포 등의 필요한 스터디를 개발과 함께 진행할 계획입니다.

4 향후 추진계획

4.1 향후 계획의 세부 내용

(주)시큐아이, (주)윈스, (주)누리랩, KISTI(한국과학기술정보연구원) 중에서 연구를 같이 진행할 기업이 확정 된 후에는 기업의 데이터셋을 이용하여 현재 기술 평가 후 부족한 부분이 있다면 개선할 것입니다.

침입 방지 기술로 뽑아낸 시그니처들이 정말로 유의미하다면 시그니처들의 임베딩 통해서 벡터로 만들고 크로마DB같은 vector DB와 연계하는 것을 목표로 두고 있습니다.

암호화 되지 않은 데이터들에 대해서는 눈으로 보기에도 시그니처의 의미를 파악할 수 있는 경우도 있기에 이것을 이용하여 연구 진행을 하다가 유의미한 시그니처가 잘 나오면 암호화된 트래픽에서 앞선 연구 기법을 이용하여 연구를 진행할 것입니다.

5 고충 및 건의사항

- 논문을 분석해가면서 코드를 만든다는 경험이 처음이라 어려움을 겪었습니다.
- 주제와 맞지 않거나 불필요한 값이 많아 데이터 셋을 선정하는데 오랜 시간이 걸렸습니다.
- 보안 분야에서 벡터DB를 사용해 진행된 연구가 없어 어려운 점이 있습니다.

6 각주

1. Seo, HyungBin, and MyungKeun Yoon. "Generative intrusion detection and prevention on data stream." 32nd USENIX Security Symposium (USENIX Security 23). 2023.

2. Zhang, Yucheng, et al. "AE: An asymmetric extremum content defined chunking algorithm for fast and bandwidth-efficient data deduplication." 2015 IEEE Conference on Computer Communications (INFOCOM). IEEE, 2015.