Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization

(새로운 침입 탐지 데이터 세트 및 침입 트래픽 특성 생성을 향하여)

Iman Sharafaldin, Arash Habibi Lashkari and Ali A. Ghorbani
CIC(Canadian Institute for Cybersecurity), University of New Brunswick(UNB), 캐나다

키워드: 침입 탐지, IDS 데이터세트, DoS, 웹 공격, 침입, 무차별 공격.

초록
이 논문은 실제 기준을 충족하고 공개적으로 사용할 수 있는 양성 및 7개의 공통 공격 네트워크 흐름을 포함하는 신뢰할 수 있는 데이터 세트를 생성한다. 결과적으로 특정 공격 범주를 탐지하기 위한 최상의 기능 집합을 나타내기 위해 포괄적인 네트워크 트래픽 기능 및 기계 학습 알고리즘 집합의 성능을 평가한다.

1. 소개
대부분의 경우 트래픽 다양성과 공격의 다양성이 부족하고 익명화가 되어 있어 데이터 세트가 현재 추세를 반영하지 않고 있다. 많은 연구자들은 제안된 기술을 테스트하고 평가하기 위해 포괄적이고 유효한 데이터 세트를 찾는 데 어려움을 겪고 있으며 적합한 데이터 세트를 사용 것 자체가 중요한 도전 과제이다.
지난 연구와 제안된 평가 프레임 작업에 따르면, 공격 다양성, 익명성, 사용 가능한 프로토콜, 완전한 캡처, 완전한 상호 작용, 완전한 네트워크 구성, 완전한 트래픽, 기능 세트, 이질성, 라벨링 및 메타데이터와 같은 11가지 특성은 완전하고 유효한 IDS 데이터 세트에 매우 중요하다.(Gharib et al., 2016).

우리의 기여:
이 논문에서 우리의 기여는 두가지 이다. 먼저, CICIDS2017이라는 새로운 IDS 데이터 세트를 생성한다. 이 데이터 세트는 일반적인 업데이트 공격으로 11가지 필수 기준을 모두 포괄한다. 데이터 세트는 완전히 레이블이 지정되고 CICFlowMeter 소프트웨어를 사용하여 모든 양성 및 침입 흐름에 대해 80개 이상의 네트워크 트래픽 기능이 추출 및 계산된다.(Habibi Lashkari et al., 2017). 두 번째로 이 논문는 생성된 데이터 세트를 분석하여 다양한 공격을 탐지하기 위한 최상의 기능 세트를 선택하고 데이터 세트를 평가하기 위해 7개의 공통 머신 러닝 알고리즘을 실행했다.

2. 사용 가능한 데이터 세트
1998년 이후 공개적으로 사용 가능한 11개의 IDS 데이터 세트를 분석하고 평가하여 포괄적이고 신뢰할 수 있는 데이터 세트에 대한 실제 필요성을 반영하는 부족함과 문제를 보여 준다.

- DARPA(Lincoln Laboratory 1998-99): 실제 네트워크 트래픽을 나타내지 않으며 오탐지 부재와 같은 불규칙성을 포함한다. 또한 공격 유형과 네트워크 인프라 측면에서 최신 네트워크에서 IDS를 효과적으로 평가하기에는 데이터 세트가 구식입니다. 또한 실제 공격 데이터 기록이 부족하다(McHugh, 2000)(Brown et al., 2009).
- KDD'99(University of California, Irvine 1998-99): 많은 수의 중복 레코드가 있으며 왜곡된 테스트 결과로 이어진 데이터 손상으로 가득 차 있다(Tavallaee et al., 2009). NSL-KDD는 KDD의 일부 단점을 해결하기 위해 KDD(Tavallaee et al., 2009)를 사용하여 만들어졌다(McHugh, 2000).
- DEFCON(The Shmoo Group, 2000-2002: 2000년에 생성된 DEFCON-8 데이터 세트에는 포트 스캔 및 버퍼 오버플로 공격이 포함되어 있는 반면, 2002년에 생성된 DEFCON-10 데이터 세트에는 포트 스캔 및 스윕, 불량 패킷, 관리 이 데이터 세트에서 "Capture the Flag(CTF)" 경쟁 중에 생성된 트래픽은 일반적인 백그라운드 트래픽이 아닌 침입 트래픽으로 주로 구성되어 있기 때문에 실제 네트워크 트래픽과 다르다. 이 데이터 세트는 경고 상관 관계 기술을 평가하는 데 사용된다(Nehinbe, 2010)(Group, 2000)
- CAIDA(Center of Applied Internet Data Analysis 2002-2016): 여러 결점으로 인해 효과적인 벤치마킹 데이터 세트가 아니다. ysis(CAIDA), 2016)(Proebstel, 2008)(Ali Shiravi 및 Ghorbani, 2012)에서 자세한 내용을 참조
- LBNL(Lawrence Berkeley National Laboratory 및 ICSI 2004-2005): 페이로드가 없으며 개별 IP를 식별할 수 있는 정보를 제거하기 위해 과도한 익명화로 인해 어려움을 겪는다(Nechaev et al., 2004).
- CDX(United States Military Academy 2009): IDS 경고 규칙을 테스트하는 데 사용할 수 있지만 트래픽 다양성 및 양이 부족하다(Sangster et al., 2009).

- Kyoto(Kyoto University 2009): 이 데이터 세트는 허니팟을 통해 생성되었으므로 수동 레이블 지정 및 익명화 프로세스는 없지만 허니팟을 향한 공격만 관찰할 수 있기 때문에 네트워크 트래픽에 대한 보기가 제한된다. (Song et al. al., 2011) (M. Sato, 2012) (R. Chitrakar, 2012).
- Twente(University of Twente 2009): auth/ident, ICMP 및 IRC 트래픽과 같이 완전히 무해하거나 악의적이지 않은 일부 동일 네트워크 트래픽이 있다. 또한 이 데이터 세트에는 일부 알 수 없고 상관 관계가 없는 경고 트래픽이 포함되어 있다. 레이블이 지정되어 있고 더 현실적이지만 공격의 양과 다양성이 부족하다(Sperotto et al., 2009).
- UMASS(University of Massachusetts 2011): 트래픽 및 공격의 다양성 부족으로 인해 IDS 및 IPS 기술을 테스트하는 데 유용하지 않다(Swagatika Prusty and Liberatore, 2011).
- ISCX2012(뉴브런즈윅 대학교 2012): 오늘날 네트워크 트래픽의 거의 70%가 HTTPS이고 이 데이터 세트에 HTTPS 추적이 없기 때문에 새로운 네트워크 프로토콜을 나타내지 않다. 또한 시뮬레이션된 공격의 분포는 실제 통계를 기반으로 하지 않다(Ali Shiravi 및 Ghorbani, 2012).
- ADFA(University of New South Wales 2013): 공격의 다양성 및 다양한 공격의 부족 외에도 이 데이터 세트의 일부 공격 동작은 정상 동작과 잘 분리되지 않다(Xie and Hu, 2013)(Xie et al., 2014).

3. 실험

포괄적인 테스트베드를 만들기 위해 우리는 Attack-Network와 Victim-Network라는 두 가지 네트워크를 설계하고 구현했다.
Victim-Network: 방화벽, 라우터, 스위치 및 대부분의 일반 운영 체제와 각 PC에서 무해한 동작을 제공하는 에이전트를 갖춘 높은 수준의 보안 인프라이다.
Attack-Network: 공격 시나리오를 실행하기 위해 공용 IP와 서로 다른 필수 운영 체제가 있는 PC 세트와 라우터 및 스위치로 설계된 완전히 분리된 인프라이다. 다음은 인프라, 무해한 프로필 에이전트 및 공격 시나리오에 대한 설명이다.

3.1 테스트베드 아키텍처

Figure 1에서 볼 수 있듯이 테스트베드 인프라는 완전히 분리된 두 개의 네트워크, 즉 Victim-Network와 Attack-Network로 나뉘다. 이전 데이터 세트와 달리 Victim-Network에서는 Windows, Linux 및 Macintosh라는 공통 세 가지 운영 체제의 서로 다른 버전과 함께 라우터, 방화벽, 스위치를 포함한 공통적으로 필요한 장비를 다루었다. Table 1은 설치된 운영 체제 및 관련 공용 및 개인 IP와 함께 서버, 워크스테이션 및 방화벽 목록을 보여준다.
Attack-Network에는 라우터 1개, 스위치 1개, Kali 및 Windows 8.1 운영 체제가 설치된 PC 4대가 포함된다. 피해자 네트워크는 3개의 서버, 1개의 방화벽, 2개의 스위치, DC(도메인 컨트롤러) 및 활성 디렉터리로 상호 연결된 10개의 PC로 구성된다. 또한 Victim-Network의 메인 스위치에 있는 한 포트는 미러 포트로 구성되어 네트워크로의 모든 송수신 트래픽을 완벽하게 캡처한다.

Table 1: Victim-Network Operating Systems and IPs.

	Machine	OS	IPs
Victim-Network	Servers	Win Server 2016 (DC and DNS)	192.168.10.3
		Ubuntu 16 (Web Server)	192.168.10.50-205.174.165.68
		Ubuntu 12	192.168.10.51-205.174.165.66
	PCs	Ubuntu 14.4 (32, 64)	192.168.10.19-192168.10.17
		Ubuntu 16.4 (32-64)	192.168.10.16-192.168.10.12
		Win 7Pro	192.168.10.9
		Win 8.1-64	192.168.10.5
		Win Vista	192.168.10.8
		Win 10 (Pro 32-64)	192.168.10.14-192.168.10.15
		Mac	192.168.10.25
	Firewall	Fortinet	
Attackers	PCs	Kali	205.174.165.73
		win 8.1	205.174.165.69
		Win 8.1	205.174.165.70
		Win 8.1	205.174.165.71

3.2 양성 프로필 에이전트

사실적인 배경 트래픽을 생성하는 것은 이 작업의 최우선 순위 중 하나이다. 이 데이터 세트의 경우 제안된 B-Profile 시스템(Sharafaldin et al., 2017)을 사용했다. 이 시스템은 인간 상호 작용의 추상적인 동작을 프로파일링하고 자연스러운 양성 백그라운드 트래픽을 생성하는 역할을 한다. 이 데이터 세트에 대한 B-프로파일은 HTTP, HTTPS, FTP, SSH 및 이메일 프로토콜을 기반으로 25명의 사용자의 추상 동작을 추출한다.

처음에는 기계 학습 및 통계 분석 기술로 사용자가 생성한 네트워크 이벤트를 캡슐화하려고 한다. 캡슐화된 기능은 프로토콜의 패킷 크기 분포, 흐름당 패킷 수, 페이로드의 특정 패턴, 페이로드 크기 및 프로토콜의 요청 시간 분포이다. 그런 다음 사용자로부터 B-Profile을 도출한 후 Java로 개발된 에이전트를 사용하여 실제 양성 이벤트를 생성하고 미리 정의된 5개의 프로토콜에 대해 Victim-Network에서 B-Profile을 동시에 수행한다.

3.3 공격 프로필 및 시나리오

이 데이터 세트에서는 최신 업데이트된 공통 공격 계열 목록을 기반으로 6개의 공격 프로필을 생성하고 관련 도구 및 코드를 사용하여 실행한다.

- 무차별 암호 대입 공격(Brute Force Attack): 이것은 비밀번호 크래킹 뿐만 아니라 웹 응용 프로그램에서 숨겨진 페이지와 콘텐츠를 발견하는 데 사용할 수 있는 가장 인기 있는 공격 중 하나이다. 이것은 기본적으로 시행착오(hit and try) 공격으로, 희생자가 성공할 때까지 반복적으로 시도하는 공격이다.
- Heatbleed 공격: 이것은 보통 TLS(Transport Layer Security) 프로토콜의 널리 사용되는 OpenSSL 암호 라이브러리에서 발생하는 버그로부터 기인한다. 일반적으로 취약한 대상(보통 서버)에게 작은 페이로드와 큰 길이 필드를 가진 비정상적인 하트비트 요청을 보내어 피해자의 응답을 유도하는 방식으로 악용된다.
- Botnet: Botnet 소유자가 다양한 작업을 수행하기 위해 사용하는 여러 인터넷 연결 장치입니다. 데이터를 훔치고, 스팸을 보내고, 공격자가 장치 및 해당 연결에 액세스할 수 있도록 하는 데 사용할 수 있다.
- DoS 공격: 공격자는 시스템이나 네트워크 리소스를 일시적으로 사용할 수 없게 만든다. 일반적으로 시스템에 과부하를 걸고 적법한 요청의 일부 또는 전부가 이행되는 것을 방지하기 위해 대상 시스템이나 리소스에 불필요한 요청을 플러딩함으로써 수행된다.
- DDoS 공격: 일반적으로 여러 시스템이 피해자의 대역폭 또는 리소스를 초과할 때 발생한다. 이러한 공격은 다수의 손상된 시스템(예: botnet)이 대규모 네트워크 트래픽을 생성하여 대상 시스템을 범람시킨 결과인 경우가 많다.
- 웹 공격: 이 공격 유형은 개인과 조직이 현재 보안을 중요하게 여기기 때문에 매일 발생한다. 우리는 SQL Injection을 사용한다. 이는 공격자가 SQL 명령어들의 문자열을 생성한 다음, 데이터베이스에 강제로 정보를 반환하도록 이용하는 기법이다. Cross-Site Scripting (XSS)은 개발자들이 코드를 제대로 테스트하지 않아 스크립트 삽입 가능성을 찾지 못할 때 발생한다. 그리고 HTTP를 통한 브루트 포스는 관리자의 비밀번호를 찾기 위해 비밀번호 목록을 시도하는 기법이다.
- 침입 공격: 내부로부터의 네트워크 침입은 일반적으로 Adobe Acrobat Reader와 같은 취약한 소프트웨어를 악용한다. 악용에 성공한 후 피해자의 컴퓨터에서 백도어가 실행되고 Nmap을 사용한 IP 스윕, 전체 포트 스캔 및 서비스 열거와 같은 피해자의 네트워크에 대한 다양한 공격을 수행할 수 있다.

4 데이터 세트

구현된 공격은 Brute Force FTP, Brute Force SSH, DoS, Heartbleed, Web Attack, Infiltration, Botnet, DDoS가 있다. 섹션 3에서 설명한 공격 시나리오를 기반으로 각 공격을 실행하기 위해 우리는 가장 훌륭하고 공개적으로 사용 가능한 도구 중 하나를 사용하거나 Python으로 개발했다. (데이터 세트는 http://www.unb.ca/cic/datasets/IDS2017.html에서 공개적으로 사용 가능)

- 무차별 대입 공격(화요일 오전-오후): 이 시나리오에서 공격자는 Kali Linux를 사용하고 피해자는 웹 서버로서 Ubuntu 16.04 시스템입니다. 공격자는 아침에 FTP-Patator를 실행하고 오후에 SSH-Patator를 실행한다.
- DoS 공격(수요일 오전): Hulk, GoldenEye, Slowloris, Slowhttptest를 사용한다. 이 시나리오에서 공격자는 Kali Linux이고 피해자는 Apache 웹 서버가 있는 Ubuntu 16.04 시스템이다.
- DoS 공격(수요일 오후): 이 시나리오에서는 Ubuntu 12.04에서 OpenSSL의 취약한 버전인 OpenSSL 버전 1.0.1f를 컴파일 및 설치한 다음 Heartleech를 사용하여 서버에서 웹 서버 프로세스의 메모리 덤프를 검색했다.
- 웹 공격(목요일 오전): DVWA(Damn Vulnerable Web App)를 사용한다. Selenium 프레임워크로 자동화된 코드를 개발했다. 공격자는 Kali Linux이고 피해자는 웹 서버인 Ubuntu 16.04 시스템이다.
- 침투 공격(목요일 오후): Metasploit을 사용한다. 첫 번째 단계에서 피해자가 Dropbox를 통해 Windows 기기에 감염된 파일을 다운로드하거나 맥킨토시 기기에 감염된 USB 플래시 메모리로부터 다운로드하면, 공격자는 두 번째 단계로 진행하기 위해 Nmap 및 포트 스캔을 피해자의 전체 네트워크에 대해 실행한다. 공격자는 Kali Linux이며 피해자는 Victim-Network의 Windows, Ubuntu 및 Macintosh 시스템이다.

- Botnet 공격(금요일 오전): 이 데이터 세트에서는 원격 셸, 파일 업로드/다운로드, 스크린샷 캡처 및 키 로깅을 제공할 수 있는 Python 기반 Botnet인 Ares를 사용했다. 공격자는 Kali Linux이며 피해자는 Vista, 7, 8.1, 10(32비트) 및 10(64비트)의 5가지 Windows OS 이다.
- DDoS 공격 및 PortScan(금요일 오후): UDP, TCP 또는 HTTP 요청을 피해 서버로 보내는 LOIC를 사용한다. 공격자는 Windows 8.1 시스템 그룹이고 피해자는 웹 서버인 Ubuntu 16 시스템이다. 또한 sS, sT, sF, sX, sN, sP, sV, sU, sO, sA, sW, sR, sL 및 B와 같은 기본 NMap 스위치에 의해 모든 Windows 시스템에 대해 Portscan 공격을 실행한다.

5 분석

이 애플리케이션의 흐름 레이블에는 SourceIP, SourcePort, DestinationIP, DestinationPort 및 프로토콜이 포함된다. 그런 다음 섹션 4에 설명된 일일 공격 일정에 따라 매일 생성된 흐름에 레이블을 지정했다. 추출된 80개의 특징은 CICFlowMeter 웹 페이지(CICFlowMeter, 2017)에서 정의되고 설명되었다.

두 번째 단계에서는 80개의 추출된 기능에서 각 공격을 탐지하기 위한 최상의 기능 세트를 찾기 위해 scikit-learn의 RandomForestRegressor 클래스를 사용했다(Pedregosa et al., 2011). 먼저 전체 데이터 세트에서 각 기능의 중요도를 계산한 다음 각 클래스에 대한 각 특징 분할의 평균 표준화 평균 값에 해당 특징 중요도 값을 곱하여 최종 결과를 얻는다. Table 3은 가장 잘 선택된 기능 목록과 각 섹션의 해당 가중치를 보여준다.

Table 3: Feature Selection.

Label	Feature	Weight
Benign	B.Packet Len Min	0.0479
	Subflow F.Bytes	0.0007
	Total Len F.Packets	0.0004
	F.Packet Len Mean	0.0002
DoS GoldenEye	B.Packet Len Std	0.1585
	Flow IAT Min	0.0317
	Fwd IAT Min	0.0257
	Flow IAT Mean	0.0214
Heartbleed	B.Packet Len Std	0.2028
	Subflow F.Bytes	0.1367
	Flow Duration	0.0991
	Total Len F.Packets	0.0903
DoS Hulk	B.Packet Len Std	0.2028
	B.Packet Len Std	0.1277
	Flow Duration	0.0437
	Flow IAT Std	0.0227
DoS Slowhttp	Flow Duration	0.0443
	Active Min	0.0228
	Active Mean	0.0219
	Flow IAT Std	0.0200
DoS slowloris	Flow Duration	0.0431
	F.IAT Min	0.0378
	B.IAT Mean	0.0300
	F.IAT Mean	0.0265
SSH-Patator	Init Win F.Bytes	0.0079
	Subflow F.Bytes	0.0052
	Total Len F.Packets	0.0034
	ACK Flag Count	0.0007
FTP-Patator	Init Win F.Bytes	0.0077
	F.PSH Flags	0.0062
	SYN Flag Count	0.0061
	F.Packets/s	0.0014
Web Attack	Init Win F.Bytes	0.0200
	Subflow F.Bytes	0.0145
	Init Win B.Bytes	0.0129
	Total Len F.Packets	0.0096
Infiltration	Subflow F.Bytes	4.3012
	Total Len F.Packets	2.8427
	Flow Duration	0.0657
	Active Mean	0.0227
Bot	Subflow F.Bytes	0.0239
	Total Len F.Packets	0.0158
	F.Packet Len Mean	0.0025
	B.Packets/s	0.0021
PortScan	Init Win F.Bytes	0.0083
	B.Packets/s	0.0032
	PSH Flag Count	0.0009
DDoS	B.Packet Len Std	0.1728
	Avg Packet Size	0.0162
	Flow Duration	0.0137
	Flow IAT Std	0.0086

평가 메트릭

- 정밀도(Pr) 또는 양의 예측 값
- Recall(Rc) 또는 민감도
- F-측정(F1)

Table 4는 KNN, Random Forest (RF), ID3, Adaboost, MLP, Naive-Bayes (NB), Quadratic Discriminant Analysis (QDA)에 대한 평가 메트릭의 가중 평균을 기반으로 한 성능 평가 결과를 보여준다. 이 결과들은 생성된 데이터 세트로부터 얻어졌다. 실행 시간과 평가 지표를 고려할 때 RF는 가장 짧은 실행 시간과 가장 높은 정확도를 가진 최고의 알고리즘이다.

Table 4: The Performance Examination Results.

Algorithm	Pr	Rc	F1	Execution (Sec.)
KNN	0.96	0.96	0.96	1908.23
RF	0.98	0.97	0.97	74.39
ID3	0.98	0.98	0.98	235.02
Adaboost	0.77	0.84	0.77	1126.24
MLP	0.77	0.83	0.76	575.73
Naive-Bayes	0.88	0.04	0.04	14.77
QDA	0.97	0.88	0.92	18.79

2016년에 발행된 최근 데이터셋 평가 프레임워크(Gharib et al., 2016)에 따르면, 각 데이터셋마다 11가지 기준을 만족시키는 것이 필요하다. 이전의 IDS(침입 탐지 시스템)용 데이터셋 중 어느 하나도 모든 기준을 만족시키지 못했다. 이제 생성된 데이터 세트에서 각 평가 기준을 어떻게 다루었는지에 대해 논의할 것이다.

- 완전한 네트워크 구성: 모뎀, 방화벽, 스위치, 라우터 및 Windows, Ubuntu 및 Macintosh와 같은 다양한 운영 체제의 존재를 포함하는 완전한 네트워크 토폴로지를 가진다.
- 전체 트래픽: Victim-Network에 사용자 프로파일링 에이전트와 12개의 서로 다른 시스템을 보유하고 Attack-Network에서 실제 공격을 수행한다.
- Labeled Dataset: 섹션 4 및 표 2는 매일 양성 및 공격 레이블을 보여준다. 또한 공격 타이밍에 대한 세부 정보는 데이터 세트 문서에 게시된다.
- 완전한 상호 작용: 그림 1에서 볼 수 있듯이 두 개의 서로 다른 네트워크와 인터넷 통신을 통해 내부 LAN 내부와 내부 LAN 사이를 모두 처리했다.
- 전체 캡처: 태핑 시스템과 같은 미러 포트를 사용하므로 모든 트래픽이 캡처되어 스토리지 서버에 기록된다.
- 사용 가능한 프로토콜: HTTP, HTTPS, FTP, SSH 및 이메일 프로토콜과 같은 모든 공통 사용 가능한 프로토콜을 제공한다.
- 공격 다양성: 이 데이터 세트에서 이미 다룬 웹 기반, 무차별 공격, DoS, DDoS, 침투, 하트블리드, 봇 및 스캔과 같은 2016년 McAfee 보고서를 기반으로 하는 가장 일반적인 공격을 포함한다.
- 이기종: 공격이 실행되는 동안 메인 스위치에서 네트워크 트래픽을 캡처하고 모든 피해자 시스템의 메모리 덤프와 시스템 호출을 캡처한다.
- 특징 세트: 생성된 네트워크 트래픽에서 80개 이상의 네트워크 흐름 기능을 추출하고 네트워크 흐름 데이터 세트를 CSV 파일로 제공한다.
- 메타 데이터: 섹션 4의 데이터 세트에 대해 완전히 설명하고 그림 1과 함께 테이블 1 및 2의 데이터 세트에 대한 세부 정보를 제공합니다. 또한 공격 시간 일정, 로그 목록 및 메모리 덤프 프로세스를 포함한 자세한 내용은 최종 문서에서 설명한다. 데이터 세트에 첨부될 것이다.

마지막으로 Table 5는 사용 가능한 11개의 데이터 세트와 생성된 데이터 간의 비교를 보여준다.

Table 5: Comparison between generated dataset and public datasets based on last IDS dataset evaluation framework.

	Network	Traffic	Label.	Interact.	Captu.	Protocols					Attack Diversity							~~Ano.~~	Heter.	Features	Meta.
						http	https	SSH	FTP	Email	Browser	Bforce	DoS	Scan	Bdoor	DNS	Other				
DARPA	YES	NO	YES	YES	YES	YES	NO	YES	YES	YES	NO	YES	YES	YES	NO	NO	YES	NO	NO	NO	YES
KDD'99	YES	NO	YES	YES	YES	YES	NO	YES	YES	YES	NO	YES	YES	YES	NO	NO	YES	NO	NO	YES	YES
DEFCON	NO	NO	NO	YES	YES	YES	NO	YES	NO	NO	NO	NO	NO	YES	YES	NO	YES	-	NO	NO	NO
CAIDAs	YES	YES	NO	NO	NO	-	-	-	-	-	NO	NO	YES	YES	NO	YES	YES	YES	NO	NO	YES
LBNL	YES	YES	NO	NO	NO	YES	NO	YES	NO	NO	-	-	-	YES	-	-	-	YES	NO	NO	NO
CDX	NO	NO	NO	YES	YES	YES	NO	YES	YES	YES	NO	NO	YES	YES	NO	YES	-	-	NO	NO	NO
KYOTO	YES	NO	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	NO	NO	YES	YES
TWENTE	YES	YES	YES	YES	YES	YES	NO	YES	YES	NO	NO	YES	NO	YES	NO	NO	YES	-	-	NO	YES
UMASS	YES	NO	YES	NO	YES	YES	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	YES	-	-	NO	NO
ISCX2012	YES	NO	YES	YES	YES	YES	NO	YES	YES	YES	YES	YES	YES	YES	YES	NO	YES	NO	YES	NO	YES
ADFA2013	YES	YES	YES	YES	YES	YES	NO	YES	YES	YES	YES	YES	NO	NO	YES	NO	YES	NO	-	NO	YES
CICIDS2017	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES

6. 결론

신뢰할 수 있고 공개적으로 사용 가능한 IDS 평가 데이터 세트를 보유하는 것은 이 도메인의 연구원 및 생산자의 근본적인 관심사 중 하나이다.

실제 기준을 충족하고 공개적으로 사용할 수 있는 7개의 업데이트된 공통 공격 계열을 포함하는 새로운 IDS 데이터 세트를 생성한다(http://www.unb.ca/cic/datasets/IDS2017.html). 미래에는 PC 수를 늘리고 더 최신 공격을 수행하고 싶다.