CIC

Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization

Iman Sharafaldin, Arash Habibi Lashkari, and Ali A. Ghorbani

Canadian Institute for Cybersecurity (CIC), University of New Brunswick (UNB)

UNB

ABSTRACT

With exponential growth in the size of computer networks and developed applications, the significant increasing of the potential damage that can be caused by launching attacks is becoming obvious. Meanwhile, Intrusion Detection Systems (IDSs) and Intrusion Prevention Systems (IPSs) are one of the most important defense tools against the sophisticated and ever-growing network attacks. Due to the lack of adequate dataset, anomaly-based approaches in intrusion detection systems are suffering from accurate deployment, analysis and evaluation. This paper produces a reliable dataset that contains benign and seven common attack network flows, which meets real world criteria and is publicly available. Consequently, the paper evaluates the performance of a comprehensive set of network traffic features and machine learning algorithms to indicate the best set of features for detecting the certain attack categories.

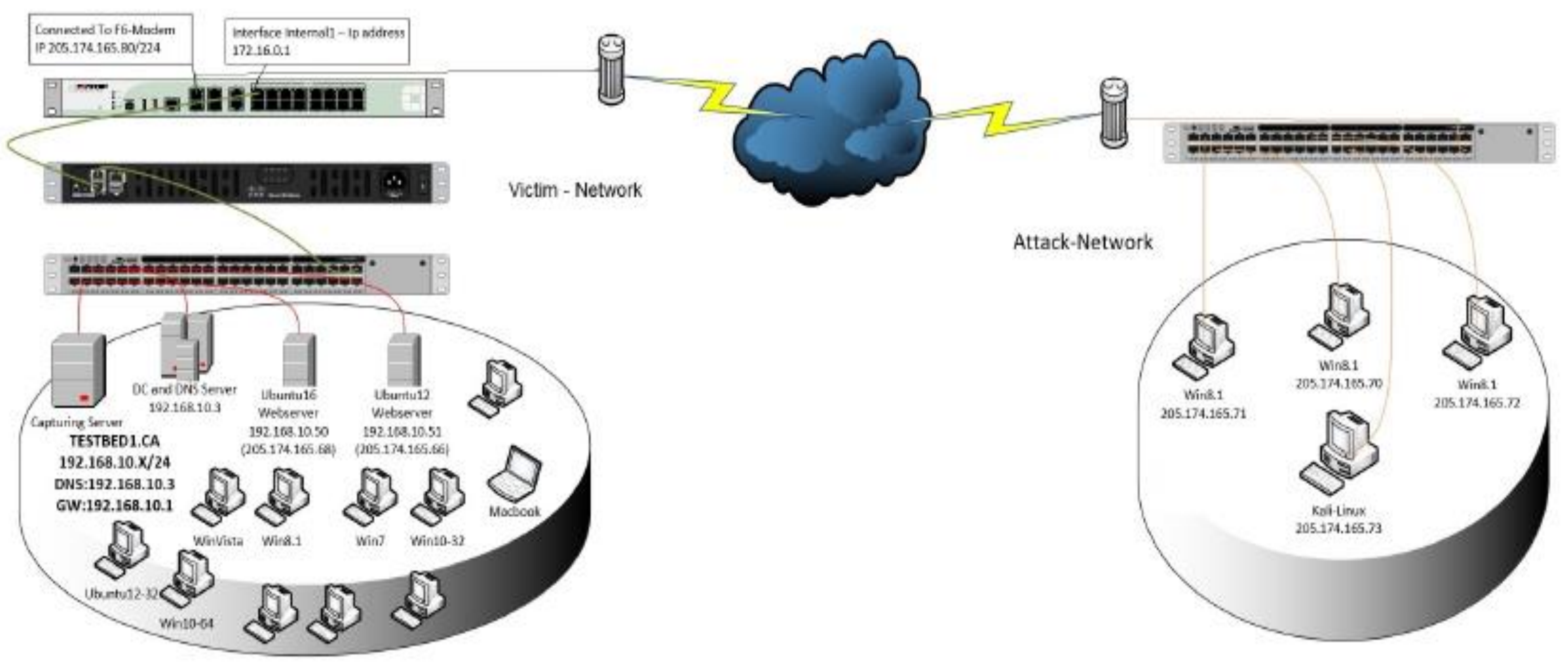
Comparison between generated dataset and public datasets based on last IDS dataset evaluation framework

	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

List of CICIDS2017 Attacks

Days	Labels
Monday	Benign
Tuesday	BForce,SFTP and SSH
Wednes.	DoS and Hearbleed Attacks slowloris, Slowhttptest, Hulk and GoldenEye
Thurs.	Web and Infiltration Attacks Web BForce, XSS and Sql Inject. Infiltration Dropbox Download and Cool disk
Friday	DDoS LOIT, Botnet ARES, PortScans (sS,sT,sF,sX,sN,sP,sV,sU, sO,sA,sW,sR,sL and B)

Testbed Architecture



Top Features Per Attack Category

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

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

Conclusion and Future Works:

- In this paper, we have monitored the state-of-the-art in the IDS dataset generation and evaluation by analyzing the eleven publicly available datasets. Then we generate a new IDS dataset includes seven common updated family of attacks that meet real worlds criteria and is publicly available (http://www.unb.ca/cic/datasets/IDS2017.html).