SUPPLEMENTARY MATERIAL: "ADVERSARIAL NETWORK TRAFFIC: TOWARDS EVALUATING THE ROBUSTNESS OF DEEP LEARNING-BASED NETWORK TRAFFIC CLASSIFICATION"

Amir Mahdi Sadeghzadeh
Department of Computer Enginnering
Sharif University of Technology
amsadeghzadeh@ce.sharif.edu

Rasool Jalili
Department of Computer Enginnering
Sharif University of Technology
jalili@sharif.edu

Saeed Shiravi
Department of Computer Enginnering
Sharif University of Technology
shiravi@ce.sharif.edu

A Architecture and hyperparameters of classifiers

The architecture and hyperparameters of packet classifiers, flow content classifiers, and flow time series classifiers are depicted in figures 7, 8, and 9, respectively.

Table 7: The architecture (A) and hyperparameters (B) of 1D-CNN packet classifiers PC-HP, and PC-P.

(A)

#	Layer	Parameters
1	1D-Convolution	Number of filters: 256, Filter size: 4, Strides: 2
2	Batch normalization	-
3	Relu	-
4	1D-Convolution	Number of filters: 128, Filter size: 8, Stride: 2
5	Batch Normalization	-
6	Relu	-
7	1D-Max Polling	Filter size: 2, Strides: 1
8	Dense	Size: 256
9	Batch normalization	-
10	Relu	-
11	Dropout	Rate: 0.1
12	Dense	Size: 128
13	Batch normalization	-
14	Relu	-
15	Dropout	Rate: 0.1
16	Dense	Size: 6
17	Softmax	-

(B)

Hyperparameter	Value
Optimizer	Adamax
Learning rate	0.01
Batch size	128
Epoch number	100
Validation metric	Accuracy
DL framework	Tensorflow 1.15 & Keras 2.2.5

Table 8: The architecture (A) and hyperparameters (B) of 1D-CNN flow content classifiers FCC-HP, and FCC-P.

(A)

#	Layer	Parameters	#	Layer	Parameters
1	1D-convolution	Number of filters: 32 Filter size: 4 Strides: 1	17	Dense	Size: 512
2	Elu	Alpha: 1.0	18	Batch normalization	-
3	Batch normalization	-	19	Relu	-
4	1D-Max pooling	Filter size: 8 Strides: 2	20	Dropout	Rate: 0.05
5	1D-convolution	Number of filters: 64 Filter size: 4 Strides: 1	21	Dense	Size: 512
6	Elu	Alpha: 1.0	22	Batch normalization	-
7	Batch normalization	-	23	Relu	-
8	1D-Max pooling	Filter size: 8 Strides: 2	24	Dropout	Rate: 0.05
9	1D-convolution	Number of filters: 128 Filter size: 8 Strides: 1	25	Dense	Size: 6
10	Elu	Alpha: 1.0	26	Softmax	-
11	Batch normalization	-			
12	1D-Max pooling	Filter size: 8 Strides: 2			
13	1D-convolution	Number of filters: 256 Filter size: 8 Strides: 1			
14	Elu	Alpha: 1.0			
15	Batch normalization	-			
16	1D-Max pooling	Filter size: 8 Strides: 2			

(B)

Hyperparameter	Value
Optimizer	Adamax
Learning rate	0.0005
Batch size	64
Epoch number	1000
Validation metric	Accuracy
DL framework	Tensorflow 1.15 & Keras 2.2.5

Table 9: The architecture (A) and hyperparameters (B) of 1D-CNN flow time series classifiers FTSC-PS, and FTSC-IAT.

(A)

#	Layer	Parameters	#	Layer	Parameters
1	1D-Convolution	Number of filters: 128, Filter size: 16, Strides: 1	22	1D-Convolution	Number of filters: 256, Filter size: 4, Strides: 1
2	Elu	Alpha: 1.0	23	Elu	Alpha: 1.0
3	Batch normalization	-	24	Batch normalization	-
4	1D-Convolution	Number of filters: 128, Filter size: 16, Stride: 1	25	1D-Convolution	Number of filters: 256, Filter size: 4, Strides: 1
5	Elu	Alpha: 1.0	26	Elu	Alpha: 1.0
6	Batch Normalization	-	27	Batch Normalization	-
7	1D-Max Polling	Filter size: 4, Strides: 2	28	1D-Max Polling	Filter size: 4, Strides: 2
8	1D-Convolution	Number of filters: 128, Filter size: 8, Strides: 1	29	Dense	Size: 1024
9	Elu	Alpha: 1.0	30	Batch normalization	-
10	Batch normalization	-	31	Relu	-
11	1D-Convolution	Number of filters: 128, Filter size: 8, Strides: 1	32	Dropout	Rate: 0.05
12	Elu	Alpha: 1.0	33	Dense	Size: 512
13	Batch Normalization	-	34	Batch normalization	-
14	1D-Max Polling	Filter size: 4, Strides: 2	35	Relu	-
15	1D-Convolution	Number of filters: 128, Filter size: 4, Strides: 1	36	Dropout	Rate: 0.05
16	Elu	Alpha: 1.0	37	Dense	Size: 512
17	Batch normalization	-	38	Batch normalization	-
18	1D-Convolution	Number of filters: 128, Filter size: 4, Strides: 1	39	Relu	-
19	Elu	Alpha: 1.0	40	Dropout	Rate: 0.05
20	Batch Normalization	-	41	Dense	Size: 6
21	1D-Max Polling	Filter size: 4, Strides: 2	42	Softmax	-

(B)

Hyperparameter	Value
Optimizer	Adamax
Learning rate	0.0003
Batch size	64
Epoch number	1000
Validation metric	Accuracy
DL framework	Tensorflow 1.15 & Keras 2.2.5

B Dataset cleaning and normalization

We used Snort[1] for flow management and feature extraction. The ISCXVPN2016 dataset requires a preprocessing phase to clean some background network traffic, such as DNS and NETBIOS. We only chose flows with at least three packets and 1000 bytes of payload to be in the dataset, and due to background flows are very small (less than ten packets), we put larger flows with higher probability in train and test sets than validation set to make train and test sets cleaner. The normalization process of the flow time series dataset is discussed in more detail in the following.

The domain of packet size is between the minimum header size and $MaxPktSize$. The packets size is first normalized by standard normalization (Equation 1), and later the minimum of normalized data is added to them, which makes normalized data positive. Finally, the domain of normalized data is limited between zero and one by dividing normalized data in the maximum of them.

$$Normalizzed\ Packet\ Size = \frac{PacketSize - \mu_{PktSize}}{Std_{PktSize}} \tag{1}$$

where $\mu_{PktSize}$, and $Std_{PktSize}$ indicate mean and standard deviation of packets size, respectively. The domain of inter-arrival time is between zero microseconds and the longest flow timeout. The flow timeout is configurable, and in our system, the longest flow timeout is 180 seconds. Log normalization (Equation 2) is used to normalize inter-arrival time data.

$$Normalizzed\ InterArrivalTime = 2 \times \log_{IAT_{max}}^{(InterArrivalTime+1\mu Sec)} \tag{2}$$

where IAT_{max} is the maximum of inter-arrival times data.

C Number of packets in flow time series and flow byte sequence

We have two criteria to choose the number of packets in flow byte sequence and flow time series: (i) the classifier should have high performance on it, and (ii) number of packets should be as few as possible to enforce online policies on network traffic. Based on experiments, we chose ten packets of each flow to be in flow byte sequence, and 100 packets of each flow to be in flow time series. Table 10 shows the overall accuracy of flow content classifiers FCC-HP, and FCC-P, and flow time series classifiers FTSC-PS, and FTSC-IAT over various numbers of packets in flow byte sequence, and flow time series, respectively.

Table 10: (A) The overall accuracy of flow content classifiers for different number of packets in the byte sequence of each flow. (B) The overall accuracy of flow time series classifiers for different number of packets in the time series of each flow.

(A)

Num of Pkts in flow byte sequence of each flow	Overall Accuracy(%)	
	FCC-HP	FCC-P
7	79.25	78.50
10	81.52	80.6
12	80.38	79.61
15	81.50	79.72
17	79.95	78.08
20	80.31	78.26

(B)

Num of Pkts in time series of each flow	Overall Accuracy(%)	
	FTSC-PS	FTSC-IAT
10	71.94	72.5
25	74.73	74.09
50	74.26	75.02
75	75.38	76.16
100	76.21	76.51
250	75.56	76.79
500	75.66	75.77
750	76.12	75.11
1000	74.73	74.95

D Overall Accuracy

The overall accuracy of packet classifiers, flow content classifiers, and flow time series classifiers are depicted in figures 4, 5, and 6, respectively.

[1] https://www.snort.org/

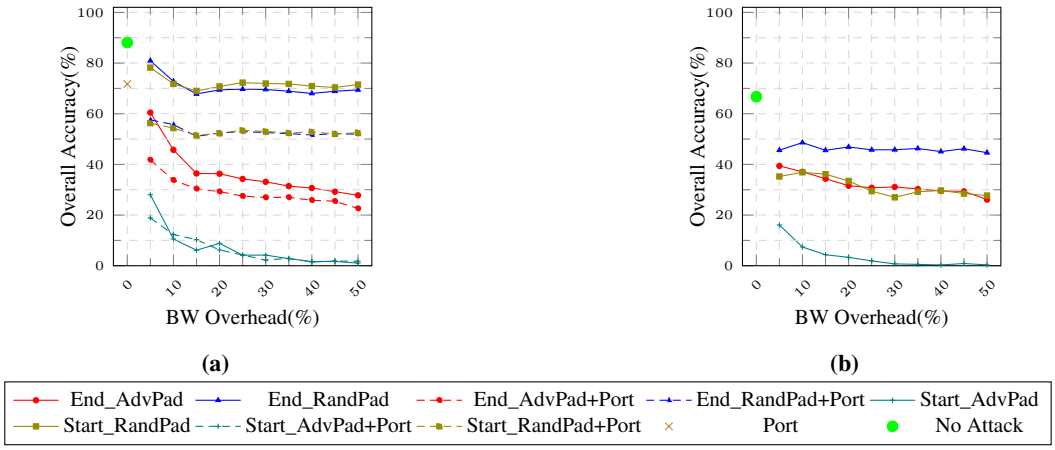

Figure 4: The overall accuracy of PC-HP (a) and PC-P (b) under different attacks over various sizes of BandWidth (BW) overhead. The legends show various kinds of attacks that have been applied.

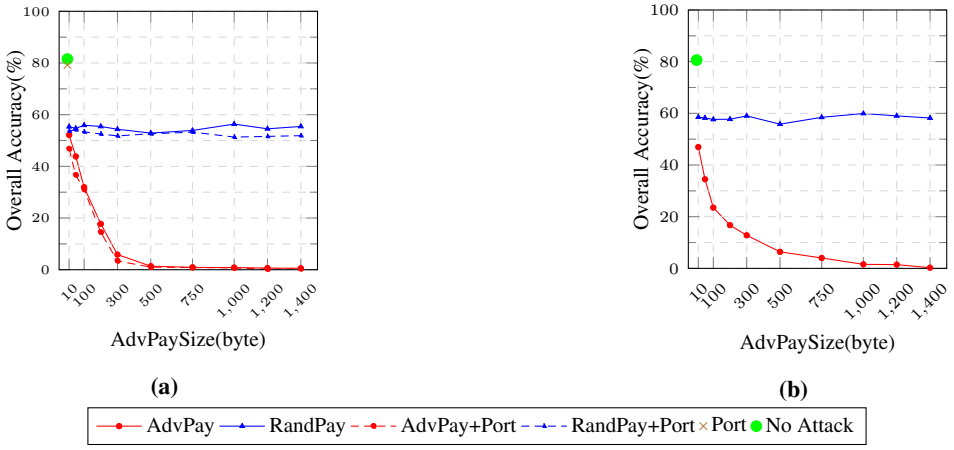

Figure 5: The overall accuracy of FCC-HP (a) and FCC-P (b) under different attacks over various sizes of adversarial payload. The legends show various kinds of attacks that have been applied.

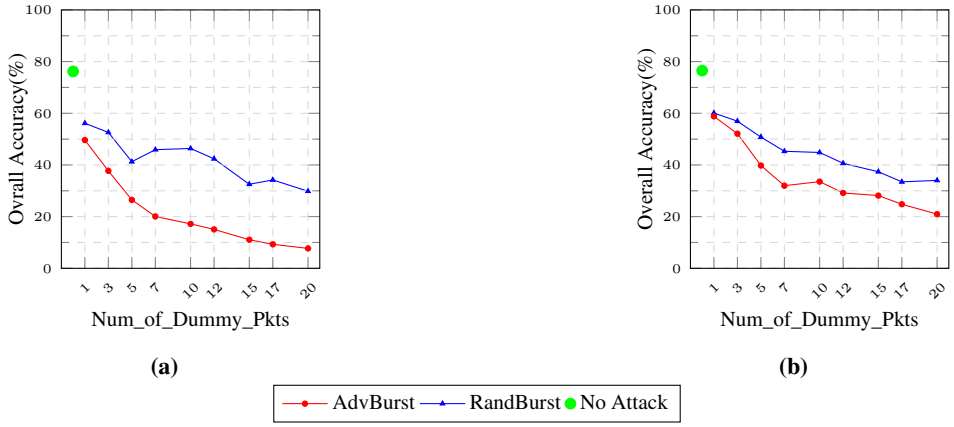

Figure 6: The overall accuracy of FTSC-PS (a) and FTSC-IAT (b) under different attacks over various number of dummy packets. The legends show various kinds of attacks that have been applied.

E Precision and F-score of packet classifiers under various attacks

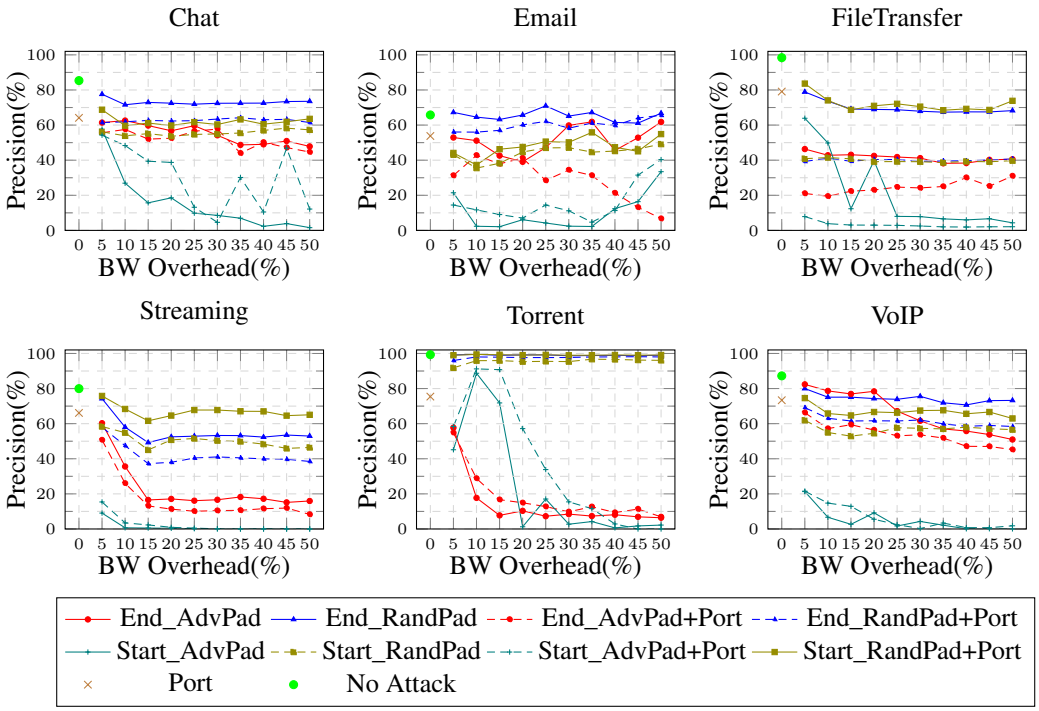

Figure 7: The precision of PC-HP under different attacks over various sizes of BandWidth (BW) overhead. The legends show the kinds of attacks that have been applied to PC-HP.

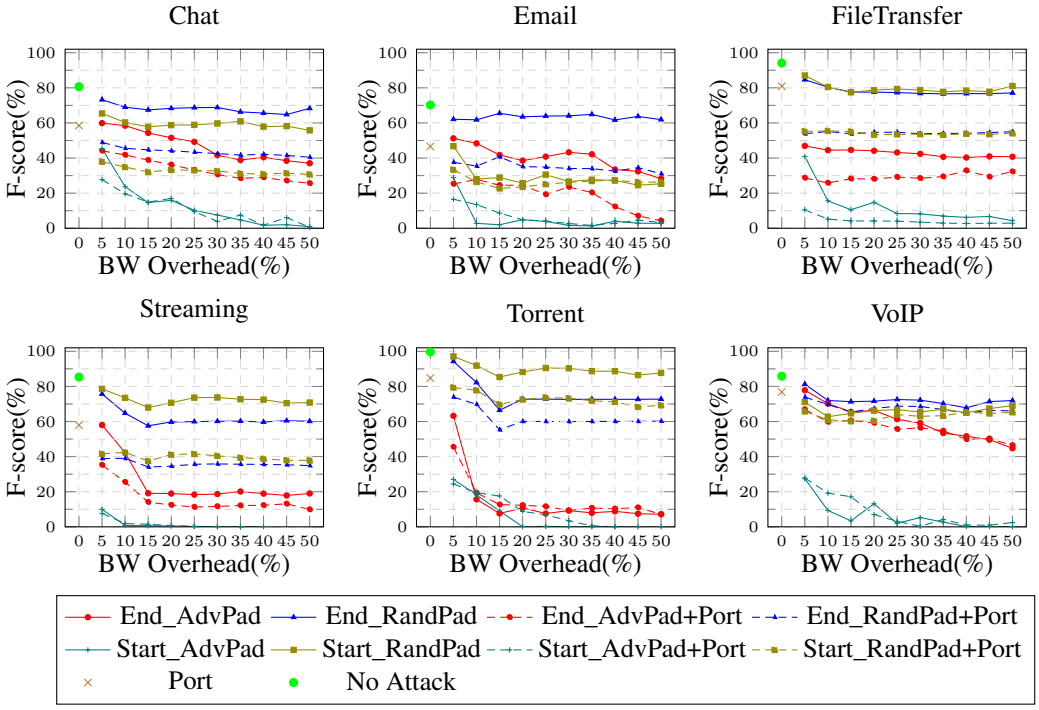

Figure 8: The F-score of PC-HP under different attacks over various sizes of BandWidth (BW) overhead. The legends show the kinds of attacks that have been applied to PC-HP.

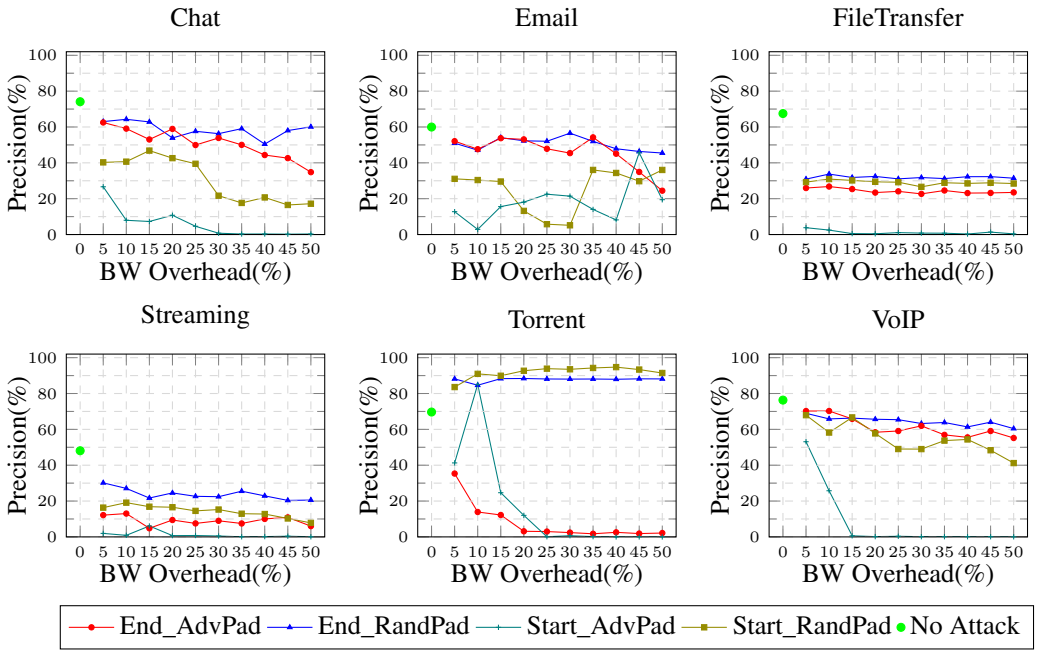

Figure 9: The precision of PC-P under different attacks over various sizes of BandWidth (BW) overhead. The legends show the kinds of attacks that have been applied to PC-P.

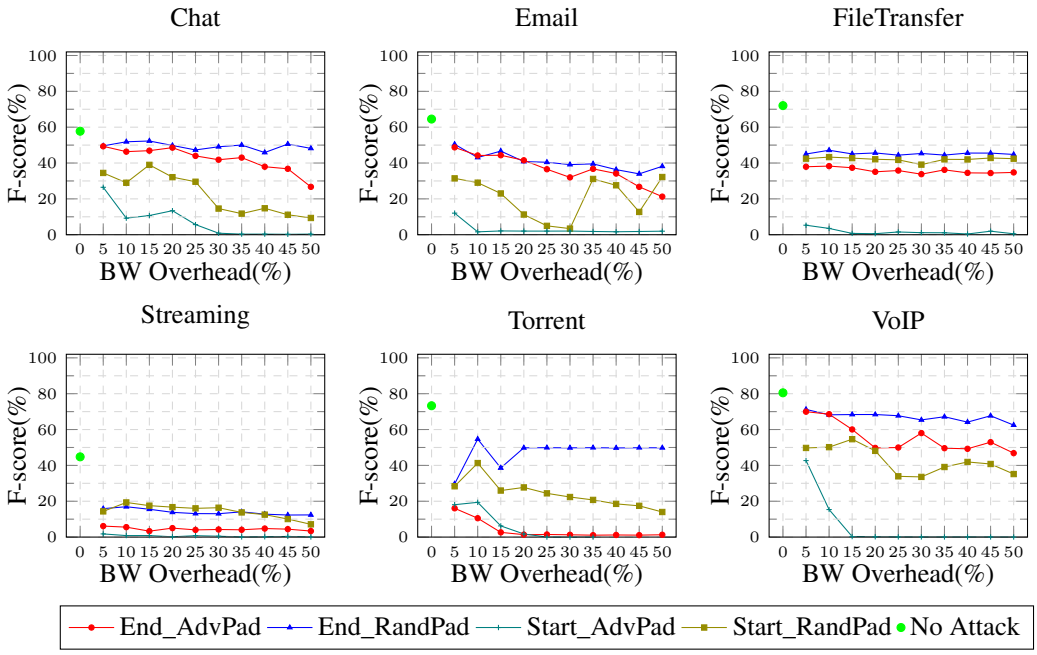

Figure 10: The F-score of PC-P under different attacks over various sizes of BandWidth (BW) overhead. The legends show the kinds of attacks that have been applied to PC-P.

F Precision and F-score of flow content classifiers under various attacks

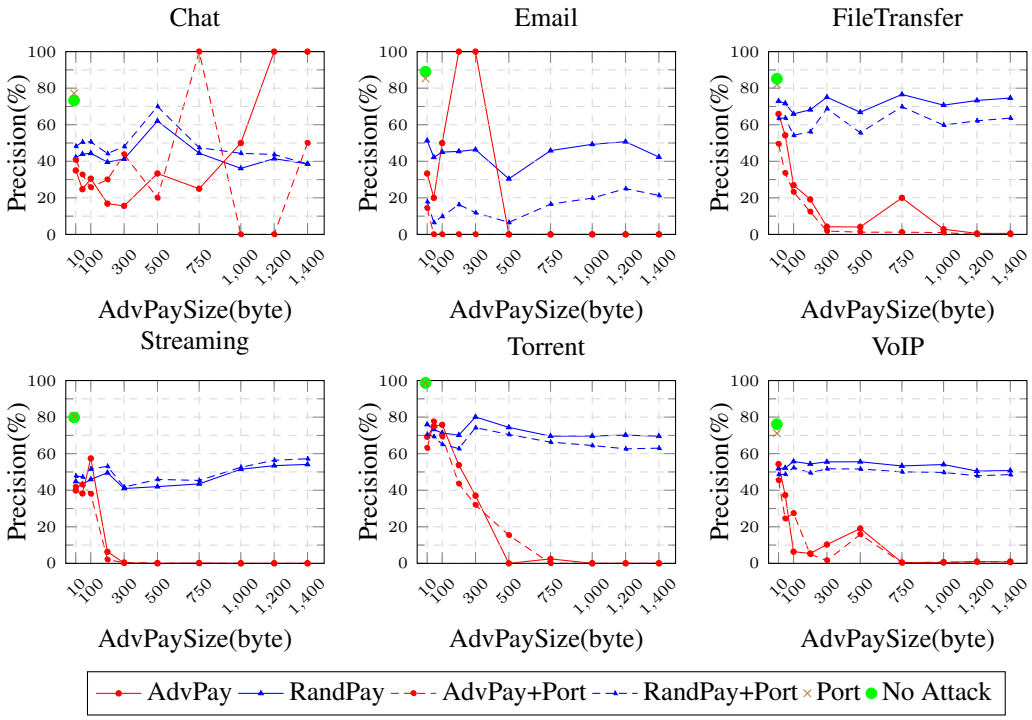

Figure 11: The precision of FCC-HP under different attacks over various sizes of adversarial payload. The legends show the kinds of attacks that have been applied to FCC-HP.

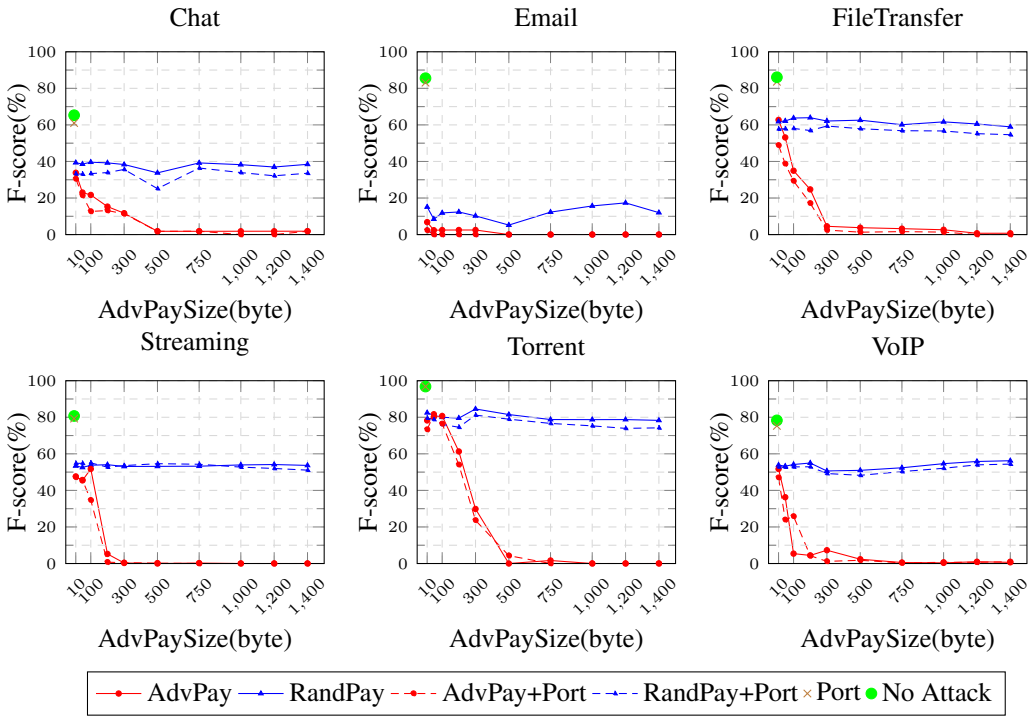

Figure 12: The F-score of FCC-HP under different attacks over various sizes of adversarial payload. The legends show the kinds of attacks that have been applied to FCC-HP.

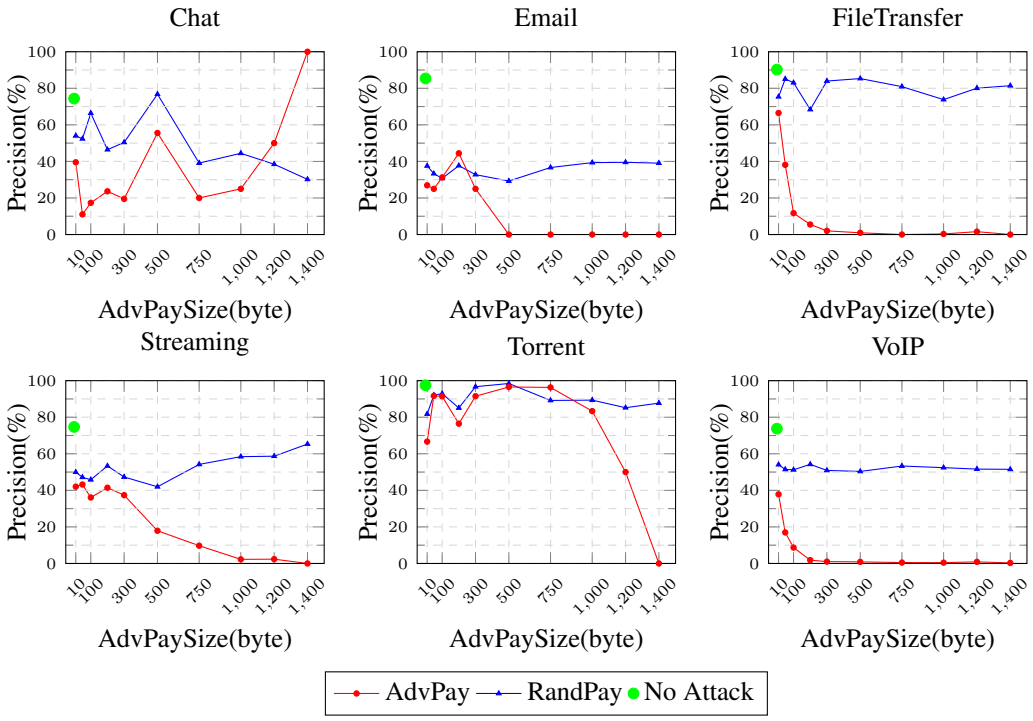

Figure 13: The precision of FCC-P under different attacks over various sizes of adversarial payload. The legends show the kinds of attacks that have been applied to FCC-P.

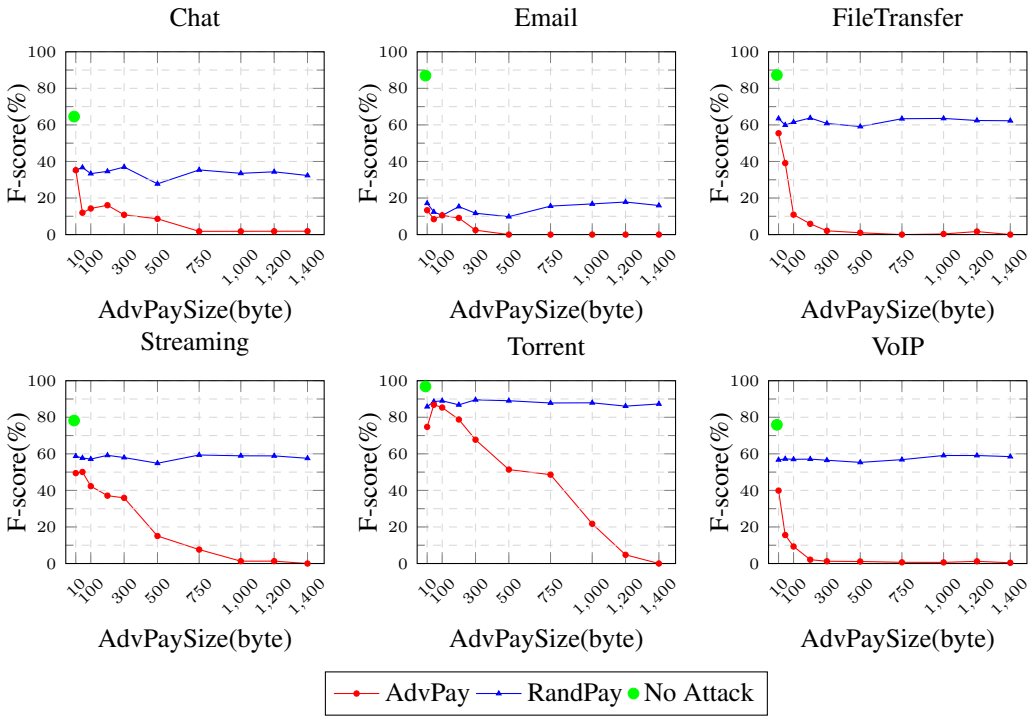

Figure 14: The F-score of FCC-P under different attacks over various sizes of adversarial payload. The legends show the kinds of attacks that have been applied to FCC-P.

G Precision and F-score of flow time series classifiers under various attacks

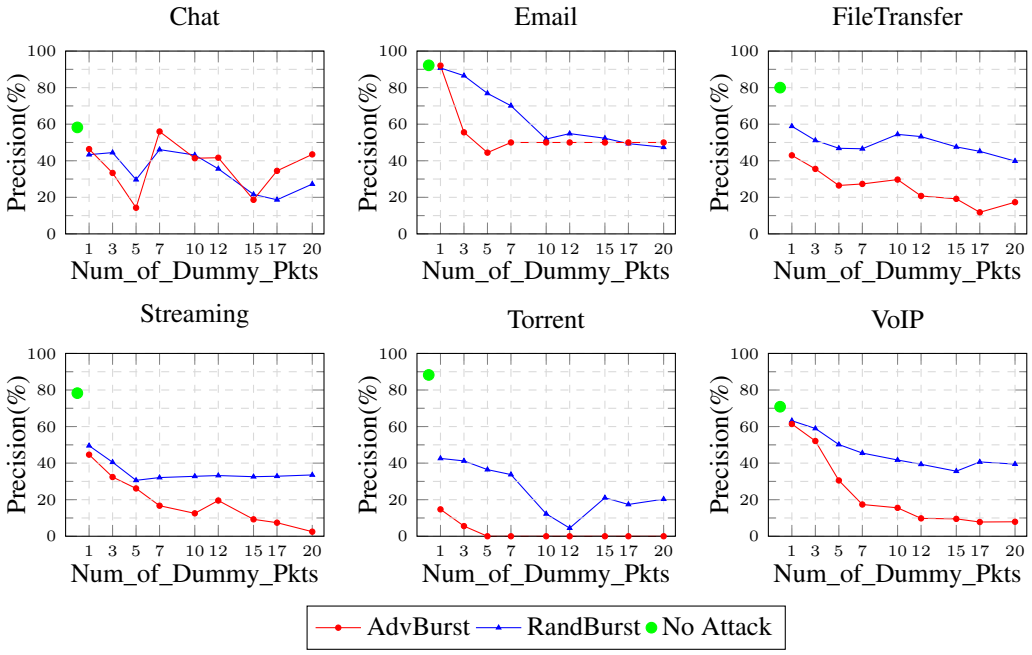

Figure 15: The precision of FTSC-PS under different attacks over various numbers of dummy packets. The legends show the kinds of attacks that have been applied to FTSC-PS.

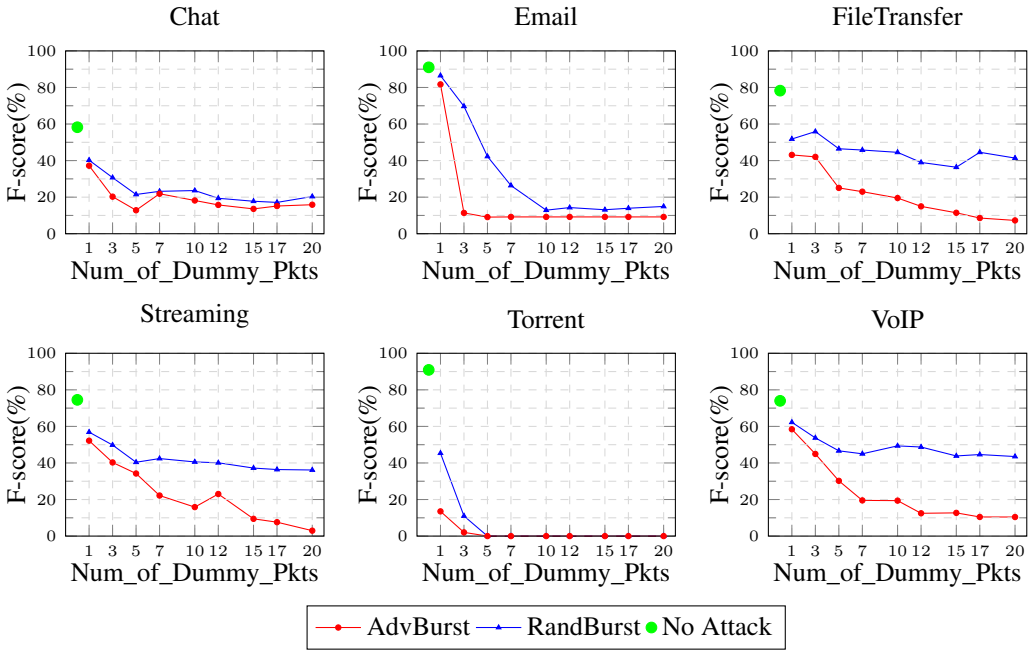

Figure 16: The F-score of FTSC-PS under different attacks over various numbers of dummy packets. The legends show the kinds of attacks that have been applied to FTSC-PS.

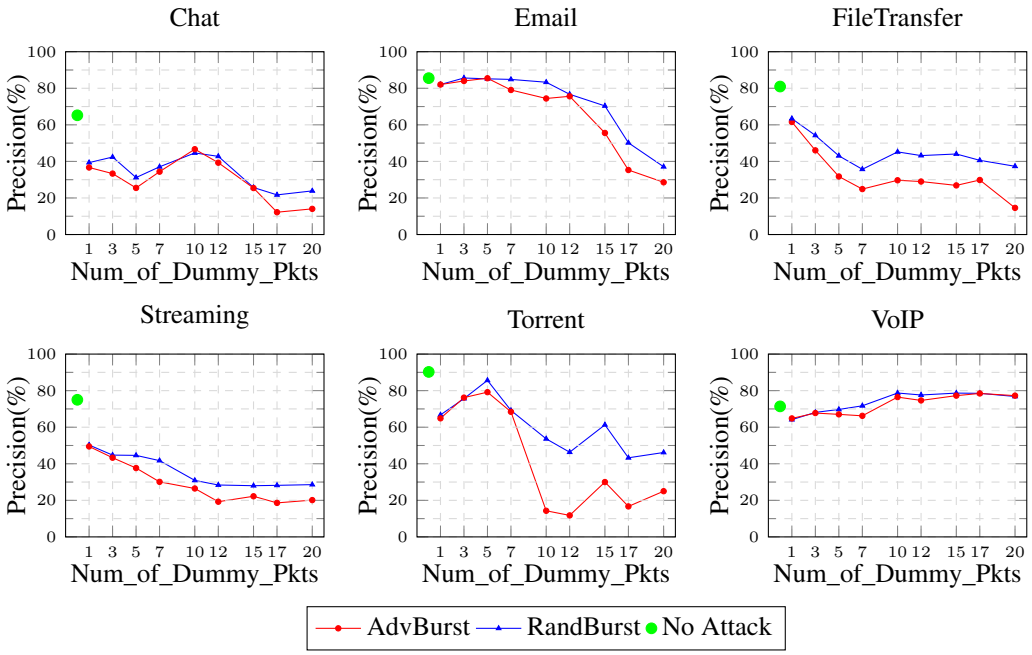

Figure 17: The precision of FTSC-IAT under different attacks over various numbers of dummy packets. The legends show the kinds of attacks that have been applied to FTSC-IAT.

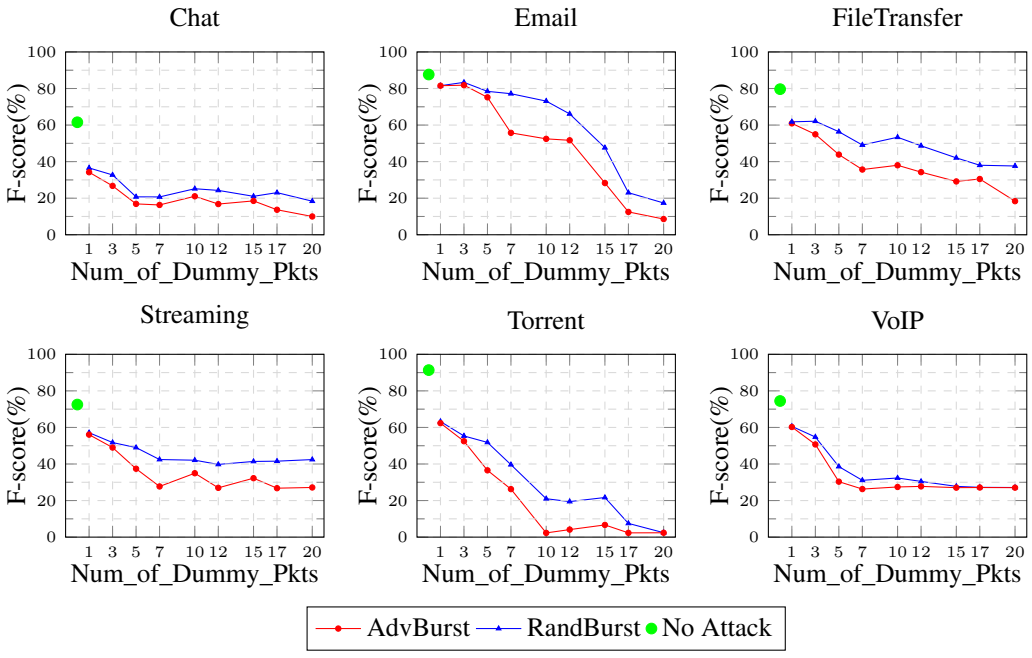

Figure 18: The F-score of FTSC-IAT under different attacks over various numbers of dummy packets. The legends show the kinds of attacks that have been applied to FTSC-IAT.