Evaluating Shallow and Deep Networks for Secure Shell (SSH)Traffic Analysis

Vinayakumar R[1], K.P Soman[1] and Prabaharan Poornachandran[2]

[1]Centre for Computational Engineering and Networking (CEN), Amrita School of Engineering, Coimbatore, Amrita Vishwa Vidyapeetham,
Amrita University, India.

[2]Center for Cyber Security Systems and Networks, Amrita School of Engineering, Amritapuri, Amrita Vishwa Vidyapeetham,
Amrita University, India.

Outline

- Introduction
- Methodology
- Description of the data set and Results
- Summary
- Future Work
- References

Introduction

- Traffic classification serves as a primary mechanism for numerous network management activities counting from knowing simple network statistics to quality of service provisioning.
- Most commonly used methods are port based, payload based and flow features statistics.

Methodology

- Flow feature statistics such as protocol, minimum packet length (f&b), minimum inter-arival time (f&b), duration mean packet length (f&b), mean inter-arival time (f&b), total packets (f&b), maximum packet length (f&b), maximum inter-arival time (f&b), total bytes (f&b), standad deviation of packet lengths (f&b), standad deviation of inter-arival times (f&b) are passed to machine learning and deep learning algorithms such as recurrent neural networks (RNN) and long short-term memory (LSTM).

Contd.

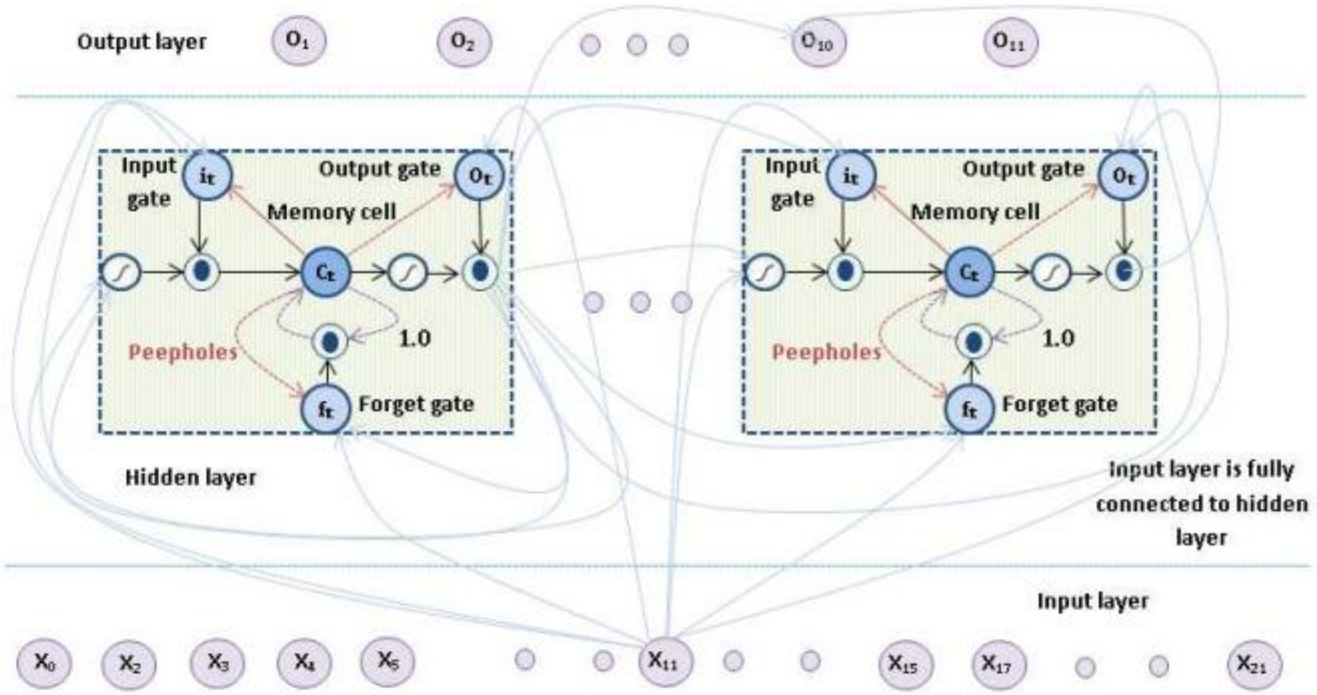
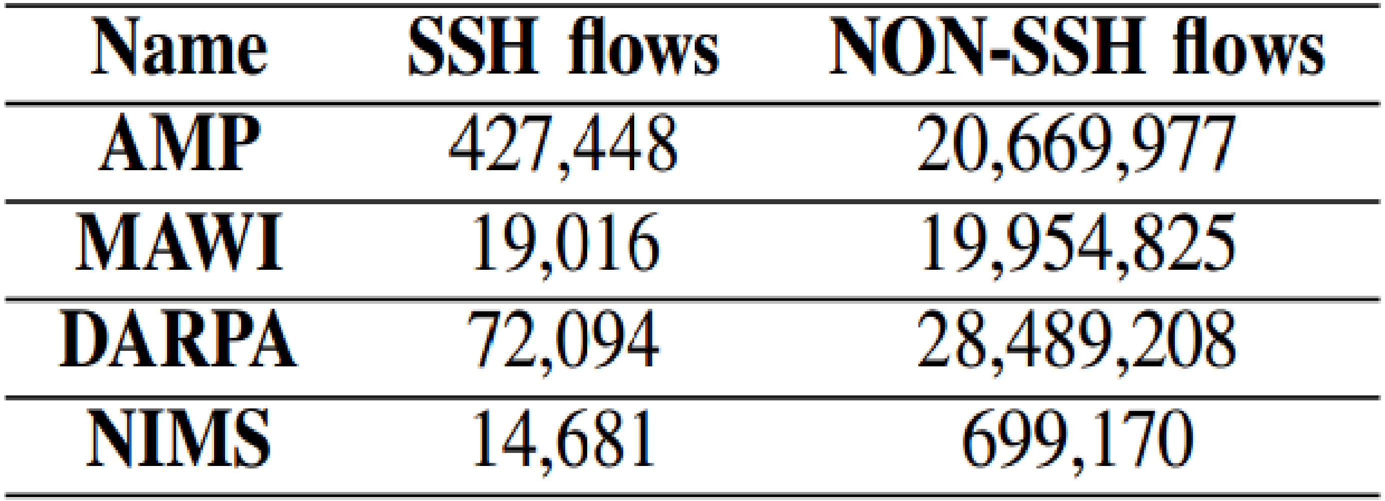


Figure 1. Proposed LSTM Architecture

Description of the data set and Results

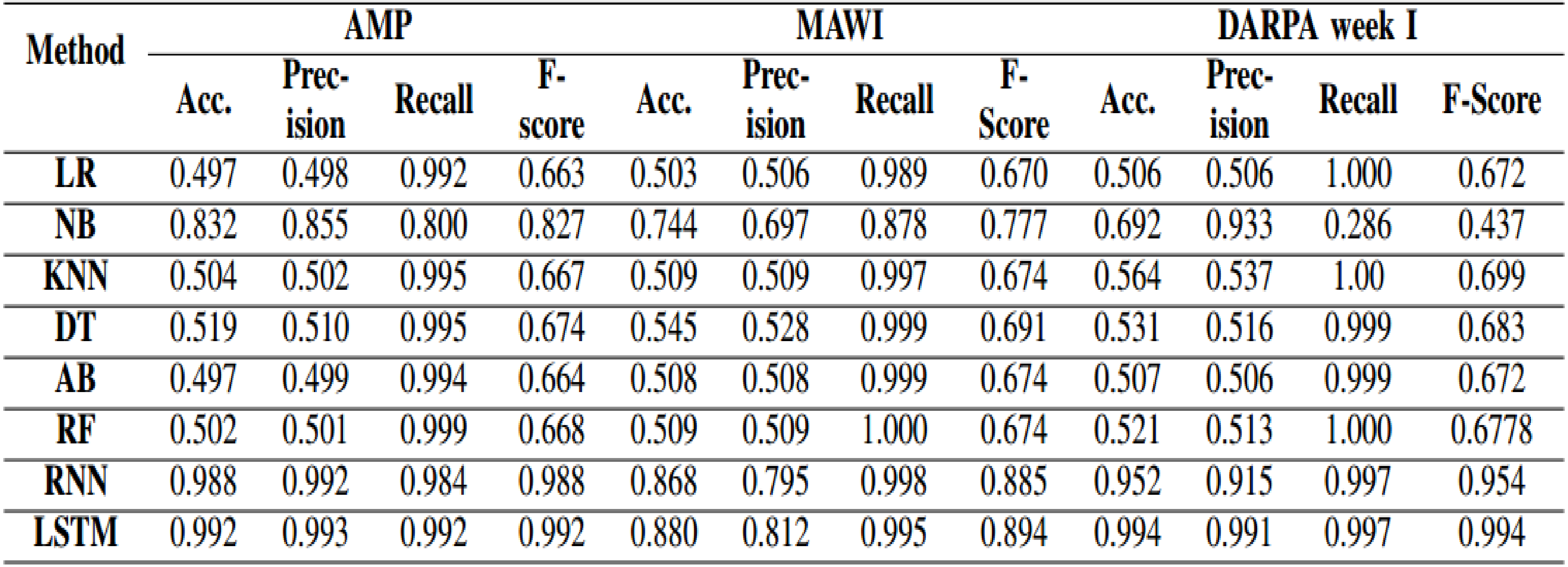
Public traces are from the NLANRs (National Laboratory for Applied Network Research) Active Measurement Project (AMP) [1] and Measurement and Analysis on the WIDE Internet (MAWI) [2] and DARPA Week 1 and Week 3 [5]. Private trace is from Network Information Management and Security Group (NIMS) [3], [4].

Table 1. Description of Data set



Name	SSH flows	NON-SSH flows
AMP	427,448	20,669,977
MAWI	19,016	19,954,825
DARPA	72,094	28,489,208
NIMS	14,681	699,170

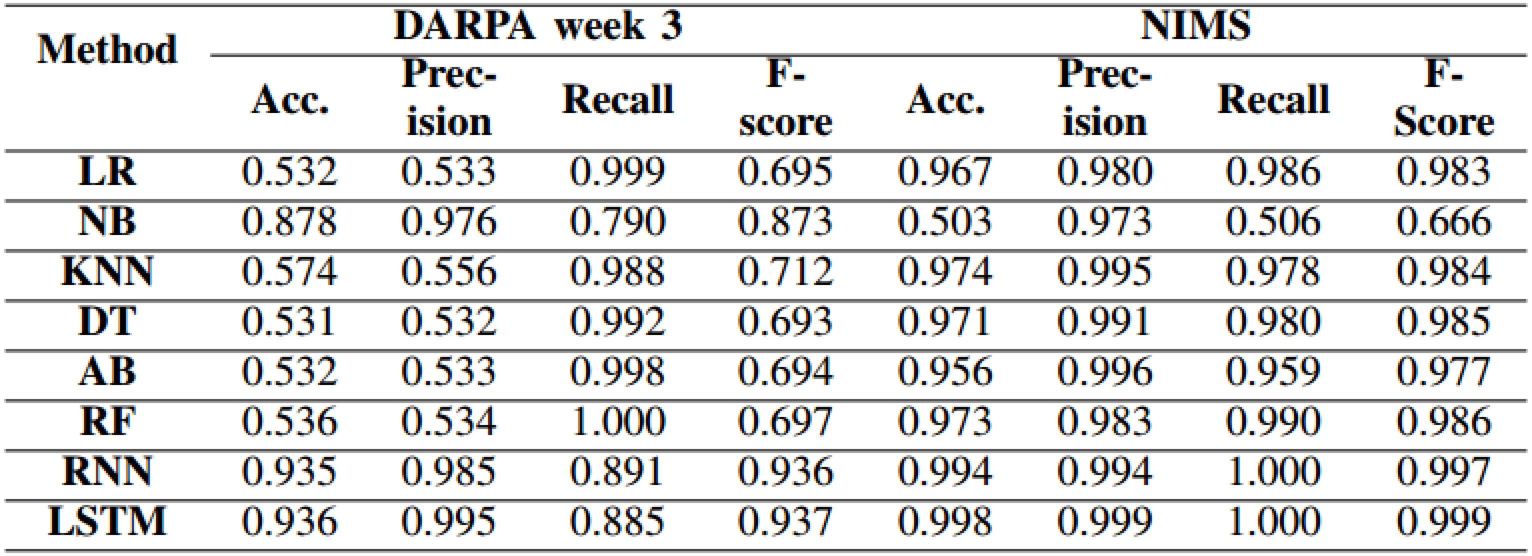
Results



Method	AMP				MAWI				DARPA week I			
	Acc.	Precision	Recall	F-score	Acc.	Precision	Recall	F-Score	Acc.	Precision	Recall	F-Score
LR	0.497	0.498	0.992	0.663	0.503	0.506	0.989	0.670	0.506	0.506	1.000	0.672
NB	0.832	0.855	0.800	0.827	0.744	0.697	0.878	0.777	0.692	0.933	0.286	0.437
KNN	0.504	0.502	0.995	0.667	0.509	0.509	0.997	0.674	0.564	0.537	1.00	0.699
DT	0.519	0.510	0.995	0.674	0.545	0.528	0.999	0.691	0.531	0.516	0.999	0.683
AB	0.497	0.499	0.994	0.664	0.508	0.508	0.999	0.674	0.507	0.506	0.999	0.672
RF	0.502	0.501	0.999	0.668	0.509	0.509	1.000	0.674	0.521	0.513	1.000	0.6778
RNN	0.988	0.992	0.984	0.988	0.868	0.795	0.998	0.885	0.952	0.915	0.997	0.954
LSTM	0.992	0.993	0.992	0.992	0.880	0.812	0.995	0.894	0.994	0.991	0.997	0.994

Table 2. Summary of test results with training data set NIMS and testing data set AMP, MAWI, DARPA week 1, DARPA week 3

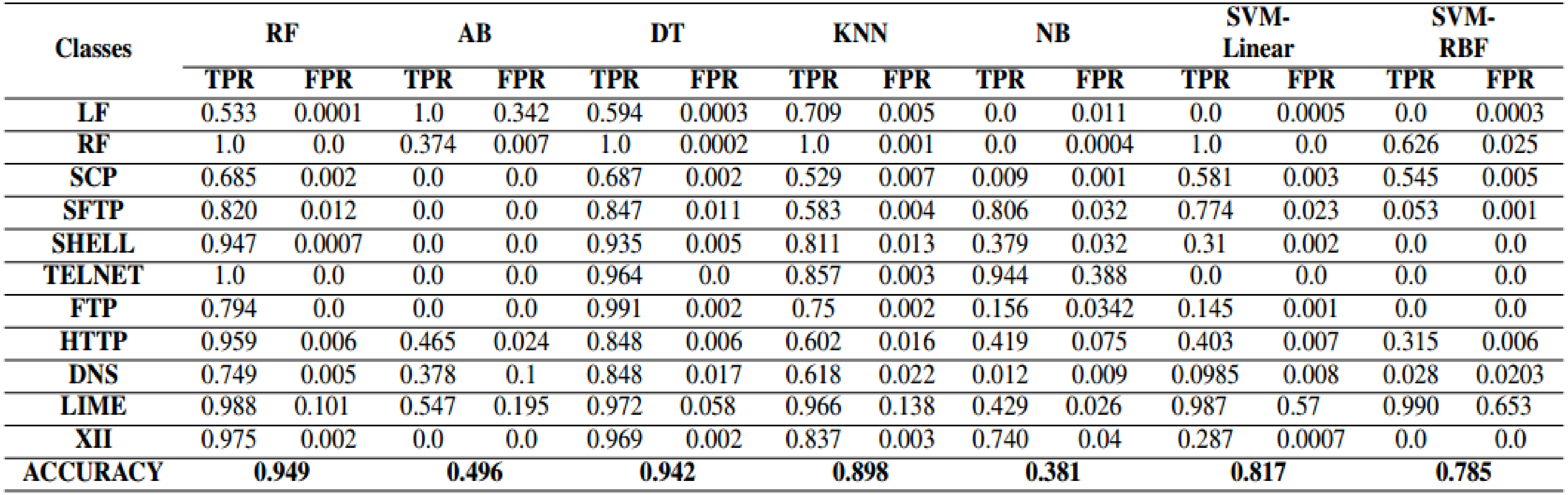
Contd.



Method	DARPA week 3				NIMS			
	Acc.	Precision	Recall	F-score	Acc.	Precision	Recall	F-Score
LR	0.532	0.533	0.999	0.695	0.967	0.980	0.986	0.983
NB	0.878	0.976	0.790	0.873	0.503	0.973	0.506	0.666
KNN	0.574	0.556	0.988	0.712	0.974	0.995	0.978	0.984
DT	0.531	0.532	0.992	0.693	0.971	0.991	0.980	0.985
AB	0.532	0.533	0.998	0.694	0.956	0.996	0.959	0.977
RF	0.536	0.534	1.000	0.697	0.973	0.983	0.990	0.986
RNN	0.935	0.985	0.891	0.936	0.994	0.994	1.000	0.997
LSTM	0.936	0.995	0.885	0.937	0.998	0.999	1.000	0.999

Table 3. Summary of test results with training data set AMP, MAWI, DARPA week 1, DARPA week 3 and testing data set NIMS

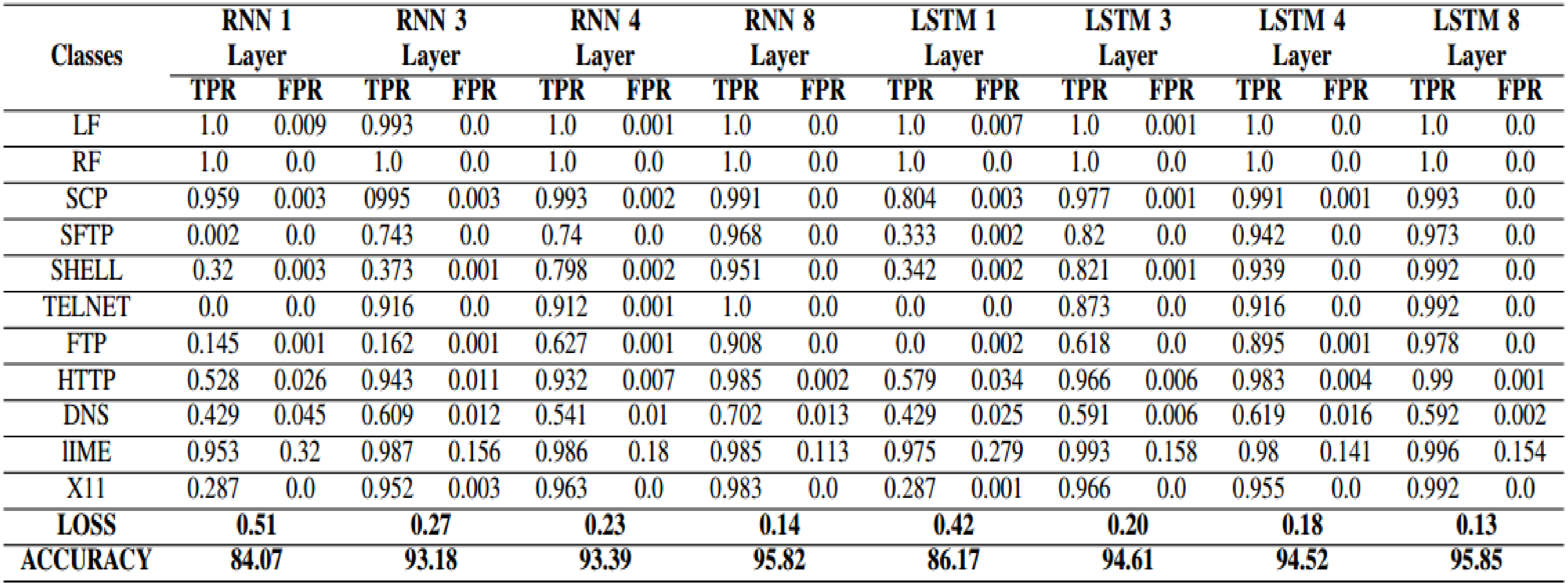
Contd.



Classes	RF		AB		DT		KNN		NB		SVM-Linear		SVM-RBF	
	TPR	FPR	TPR	FPR	TPR	FPR	TPR	FPR	TPR	FPR	TPR	FPR	TPR	FPR
LF	0.533	0.0001	1.0	0.342	0.594	0.0003	0.709	0.005	0.0	0.011	0.0	0.0005	0.0	0.0003
RF	1.0	0.0	0.374	0.007	1.0	0.0002	1.0	0.001	0.0	0.0004	1.0	0.0	0.626	0.025
SCP	0.685	0.002	0.0	0.0	0.687	0.002	0.529	0.007	0.009	0.001	0.581	0.003	0.545	0.005
SFTP	0.820	0.012	0.0	0.0	0.847	0.011	0.583	0.004	0.806	0.032	0.774	0.023	0.053	0.001
SHELL	0.947	0.0007	0.0	0.0	0.935	0.005	0.811	0.013	0.379	0.032	0.31	0.002	0.0	0.0
TELNET	1.0	0.0	0.0	0.0	0.964	0.0	0.857	0.003	0.944	0.388	0.0	0.0	0.0	0.0
FTP	0.794	0.0	0.0	0.0	0.991	0.002	0.75	0.002	0.156	0.0342	0.145	0.001	0.0	0.0
HTTP	0.959	0.006	0.465	0.024	0.848	0.006	0.602	0.016	0.419	0.075	0.403	0.007	0.315	0.006
DNS	0.749	0.005	0.378	0.1	0.848	0.017	0.618	0.022	0.012	0.009	0.0985	0.008	0.028	0.0203
LIME	0.988	0.101	0.547	0.195	0.972	0.058	0.966	0.138	0.429	0.026	0.987	0.57	0.990	0.653
XII	0.975	0.002	0.0	0.0	0.969	0.002	0.837	0.003	0.740	0.04	0.287	0.0007	0.0	0.0
ACCURACY	**0.949**		**0.496**		**0.942**		**0.898**		**0.381**		**0.817**		**0.785**	

Table 4. Summary of test results in classifying background applications running over SSH and NON-SSH using machine learning

Contd.



Classes	RNN 1 Layer		RNN 3 Layer		RNN 4 Layer		RNN 8 Layer		LSTM 1 Layer		LSTM 3 Layer		LSTM 4 Layer		LSTM 8 Layer	
	TPR	FPR	TPR	FPR	TPR	FPR	TPR	FPR	TPR	FPR	TPR	FPR	TPR	FPR	TPR	FPR
LF	1.0	0.009	0.993	0.0	1.0	0.001	1.0	0.0	1.0	0.007	1.0	0.001	1.0	0.0	1.0	0.0
RF	1.0	0.0	1.0	0.0	1.0	0.0	1.0	0.0	1.0	0.0	1.0	0.0	1.0	0.0	1.0	0.0
SCP	0.959	0.003	0995	0.003	0.993	0.002	0.991	0.0	0.804	0.003	0.977	0.001	0.991	0.001	0.993	0.0
SFTP	0.002	0.0	0.743	0.0	0.74	0.0	0.968	0.0	0.333	0.002	0.82	0.0	0.942	0.0	0.973	0.0
SHELL	0.32	0.003	0.373	0.001	0.798	0.002	0.951	0.0	0.342	0.002	0.821	0.001	0.939	0.0	0.992	0.0
TELNET	0.0	0.0	0.916	0.0	0.912	0.001	1.0	0.0	0.0	0.0	0.873	0.0	0.916	0.0	0.992	0.0
FTP	0.145	0.001	0.162	0.001	0.627	0.001	0.908	0.0	0.0	0.002	0.618	0.0	0.895	0.001	0.978	0.0
HTTP	0.528	0.026	0.943	0.011	0.932	0.007	0.985	0.002	0.579	0.034	0.966	0.006	0.983	0.004	0.99	0.001
DNS	0.429	0.045	0.609	0.012	0.541	0.01	0.702	0.013	0.429	0.025	0.591	0.006	0.619	0.016	0.592	0.002
IIME	0.953	0.32	0.987	0.156	0.986	0.18	0.985	0.113	0.975	0.279	0.993	0.158	0.98	0.141	0.996	0.154
X11	0.287	0.0	0.952	0.003	0.963	0.0	0.983	0.0	0.287	0.001	0.966	0.0	0.955	0.0	0.992	0.0
LOSS	**0.51**		**0.27**		**0.23**		**0.14**		**0.42**		**0.20**		**0.18**		**0.13**	
ACCURACY	**84.07**		**93.18**		**93.39**		**95.82**		**86.17**		**94.61**		**94.52**		**95.85**	

Table 5. Summary of test results in classifying services and background traffics running over of SSH and NON-SSH using deep learning

Summary

- The performance of machine learning and deep learning approaches are evaluated on classifying SSH.
- To know how machine learning and deep learning works on completely unseen data, we have trained both machine learning and deep learning approaches on public traces such as AMP, MAWI, DARPA Week 1 and DARPA week 3 and the performance of them is evaluated on private trace such as NIMS and vice versa.
- Deep learning algorithms performed well in comparison to the machine learning algorithms in all the experimental settings.

Future Work

- The internet and its applications mainly peer-2-peer (P2P), voice over internet protocol (VOIP), multi-media are following constant transformation. Thus, the patterns of traffic are very dynamic. Thus the proposed technique can be applied on the recently released data set.

References

[1] “Nlanr,” available at http://pma.nlanr.net/special.

[2] “Mawi,” available at http://tracer.csl.sony.co.jp/mawi/.

[3] R. Alshammari and A. N. Zincir-Heywood, “A flow based approach for ssh traffic detection,” in Systems, Man and Cybernetics, 2007. ISIC. IEEE International Conference on. IEEE, 2007, pp. 296–301.

[4] “Can encrypted traffic be identified without port numbers, ip addresses and payload inspection?” Computer networks, vol. 55, no. 6, pp. 1326–1350, 2011

[5] L. Didaci, G. Giacinto, and F. Roli, “Ensemble learning for intrusion detection in computer networks,” in Workshop Machine Learning Methods Applications, Siena, Italy, 2002