Evaluating Shallow and Deep Networks for Ransomware Detection and Classification

Vinayakumar R[1], K.P Soman[1] and K.K.Senthil Velan[2], Shaunak Ganorkar[2]

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

Amrita University, India.

[2]Lakhshya Cyber Security Labs Pvt Ltd,

Coimbatore, India.

Outline

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

Introduction

- Ransomware - An old malware causing havoc in recent times.
- A malware which holds your system and files hostage and demands ransom in exchange of release the system/files.
- Signature based detection is the most commonly used mechanism.
- Signature based detection fails at detecting the new ransomware or variants of existing ransomware.

Methodology

- Cerber, Cryptolocker, CryptoWall, Maktub, Sage, Ransomware, Torrentlocker - in all 7 Ransomware Families over 1300 unique variants of Ransomwares.
- Cuckoo Sandbox is widely popular sandbox used for malware analysis.
- The reports were processed - Frequency of API calls made by the samples were taken in account.
- Around 130-150 API Calls were taken into account with corresponding Frequency counts.
- These reports further preprocessed so that our SVM and MLP will accept them as inputs.

Contd.

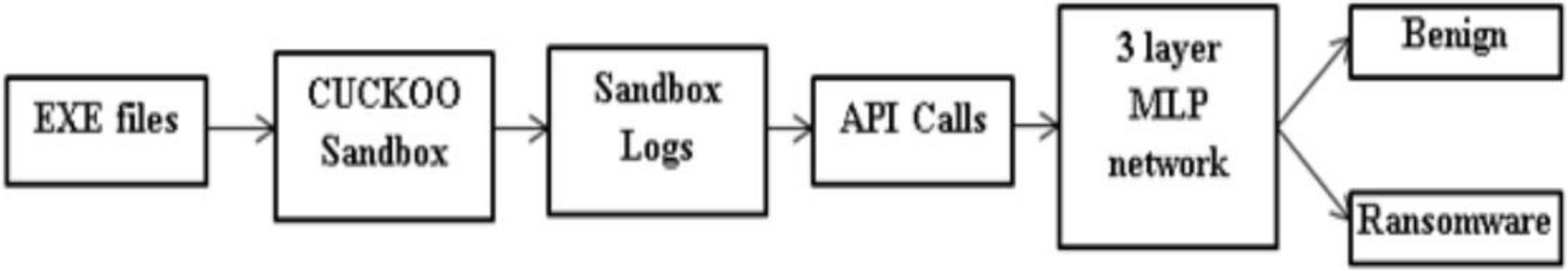
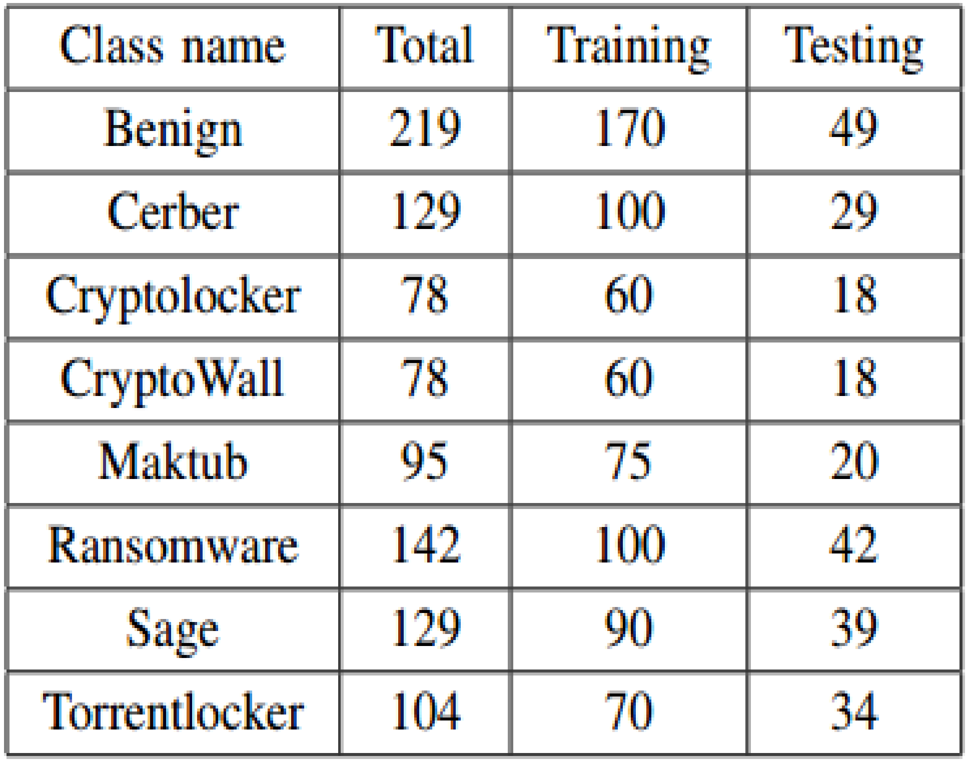


Figure 1. MLP architecture for ransomware detection and classification.

Description of the data set and Results

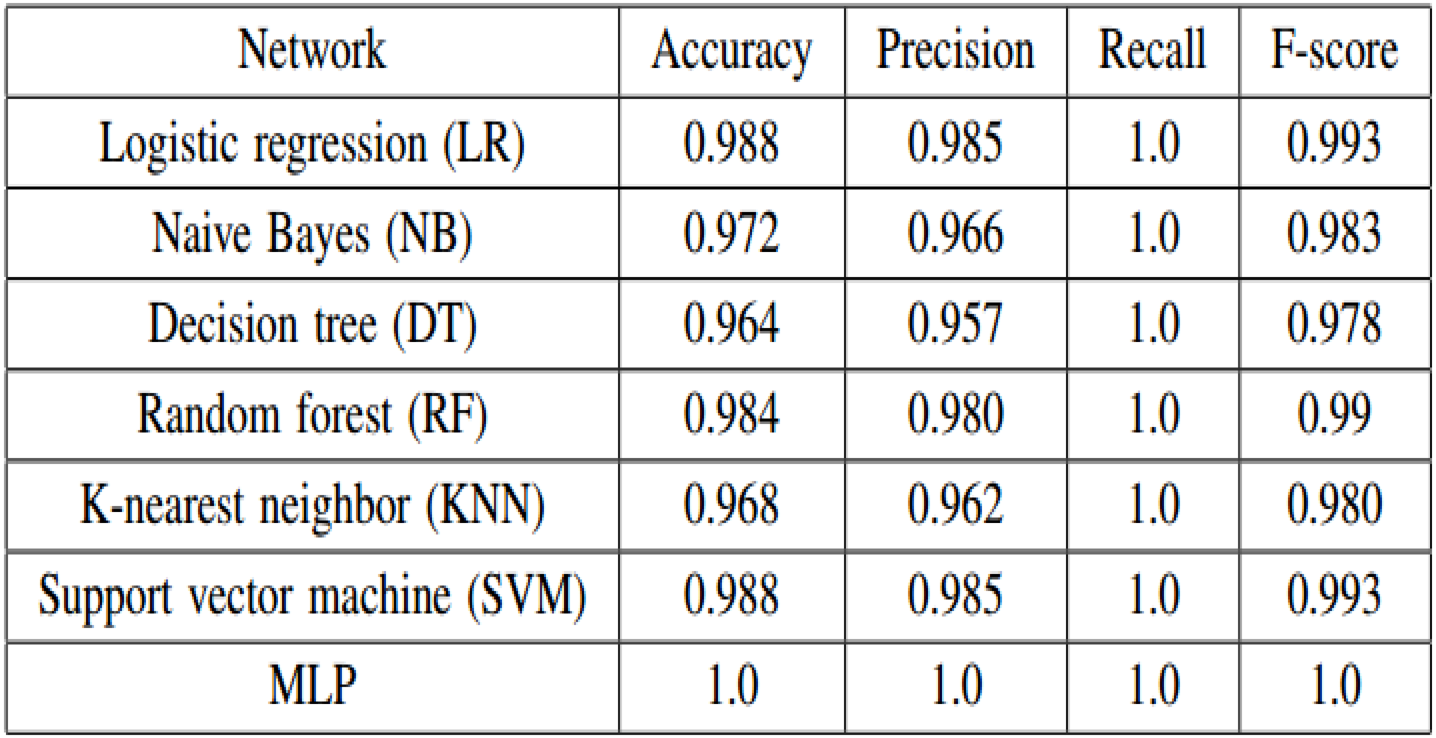
We have collected 7 different ransomware families through various sources such as Open Malware [1], Contagio Malware Dump [2], Malwr [3], theZoo aka Malware DB [4], VirusTotal [5] and VirusShare [6].

Table 1 Description of Data set



Class name	Total	Training	Testing
Benign	219	170	49
Cerber	129	100	29
Cryptolocker	78	60	18
CryptoWall	78	60	18
Maktub	95	75	20
Ransomware	142	100	42
Sage	129	90	39
Torrentlocker	104	70	34

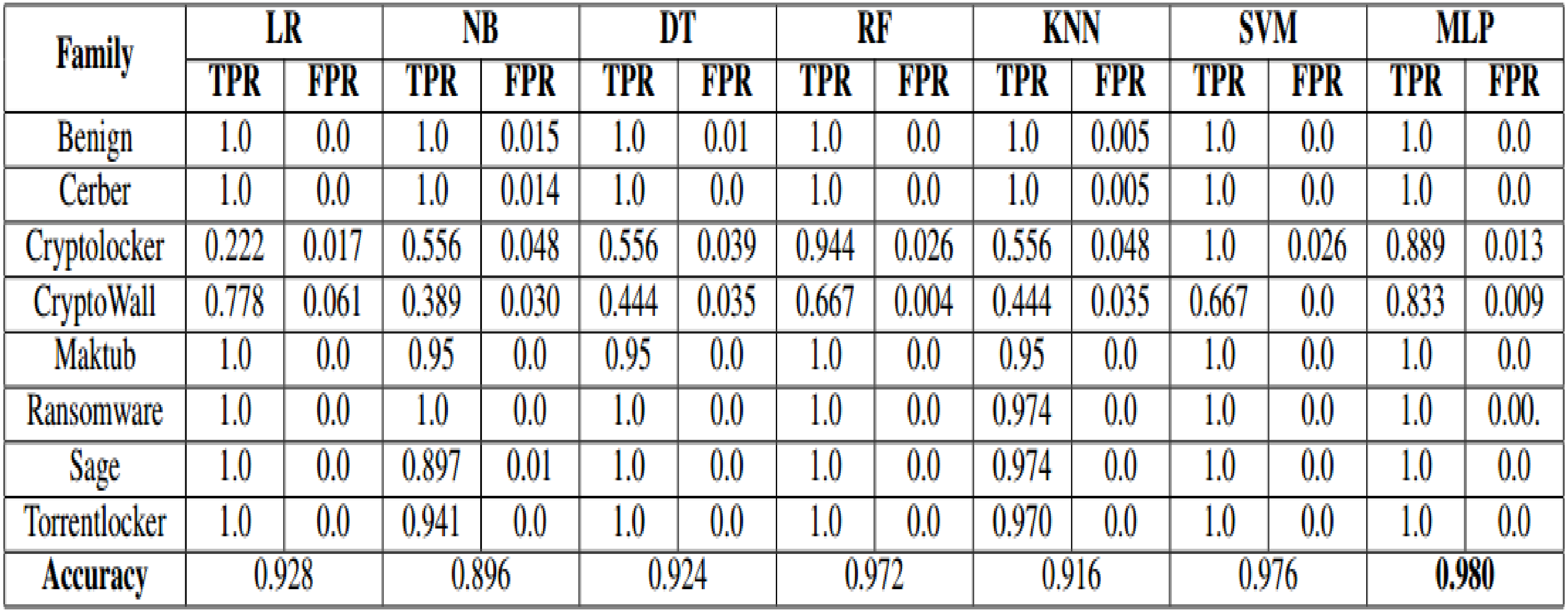
Contd.



Network	Accuracy	Precision	Recall	F-score
Logistic regression (LR)	0.988	0.985	1.0	0.993
Naive Bayes (NB)	0.972	0.966	1.0	0.983
Decision tree (DT)	0.964	0.957	1.0	0.978
Random forest (RF)	0.984	0.980	1.0	0.99
K-nearest neighbor (KNN)	0.968	0.962	1.0	0.980
Support vector machine (SVM)	0.988	0.985	1.0	0.993
MLP	1.0	1.0	1.0	1.0

Table 2. Summary of test results for binary classification

Contd.



Family	LR		NB		DT		RF		KNN		SVM		MLP	
	TPR	FPR	TPR	FPR	TPR	FPR	TPR	FPR	TPR	FPR	TPR	FPR	TPR	FPR
Benign	1.0	0.0	1.0	0.015	1.0	0.01	1.0	0.0	1.0	0.005	1.0	0.0	1.0	0.0
Cerber	1.0	0.0	1.0	0.014	1.0	0.0	1.0	0.0	1.0	0.005	1.0	0.0	1.0	0.0
Cryptolocker	0.222	0.017	0.556	0.048	0.556	0.039	0.944	0.026	0.556	0.048	1.0	0.026	0.889	0.013
CryptoWall	0.778	0.061	0.389	0.030	0.444	0.035	0.667	0.004	0.444	0.035	0.667	0.0	0.833	0.009
Maktub	1.0	0.0	0.95	0.0	0.95	0.0	1.0	0.0	0.95	0.0	1.0	0.0	1.0	0.0
Ransomware	1.0	0.0	1.0	0.0	1.0	0.0	1.0	0.0	0.974	0.0	1.0	0.0	1.0	0.00.
Sage	1.0	0.0	0.897	0.01	1.0	0.0	1.0	0.0	0.974	0.0	1.0	0.0	1.0	0.0
Torrentlocker	1.0	0.0	0.941	0.0	1.0	0.0	1.0	0.0	0.970	0.0	1.0	0.0	1.0	0.0
Accuracy	0.928		0.896		0.924		0.972		0.916		0.976		**0.980**	

Table 3. Summary of test results for multiclass classification

Summary

- We designed a Proof-of-Concept model for Detection of Ransomware Samples using Multi-layer percptron (MLP). The MLP architecture used system API calls to detect ransomware samples.
- For comparison, the other classical machine learning classifiers are used. MLP network performed well in comparison to the other classical machine learning algorithms.
- The binary classification gives 99% accuracy. This shows that the model is able to identify a sample as a malware or a benign sample.

Future Work

- Number of ransomware families can even more, provided that the number of variants per family is large.
- Various other features such as Registry Operations, Mutex accessed, Strings,etc can also be considered for classification of Ransomware samples. This will be helpful to counter the new variants of ransomware.

References

[1] http://www.offensivecomputing.net/

[2] http://contagiodump.blogspot.in/

[3] https://malwr.com/

[4] https://github.com/ytisf/theZoo/

[5] https://virustotal.com/

[6] https://virusshare.com/