Ensuring Reproducibility in AGD Detection: A Model Comparison Framework

Tomás Pelayo-Benedet[1], Ricardo J. Rodríguez[1], Carlos H. Gañán[2]

[1]Dpto. de Informática e Ingeniería de Sistemas, Universidad de Zaragoza, Spain
[2]Delft University of Technology, The Netherlands

Domain Generation Algorithms (DGAs)

- First time detected in the `Conficker` malware family [2]
- A DGA generates domain names following behaviour similar to a pseudo-random number generator. These domain names are known as *Algorithmically Generated Domains* (AGDs)
- Few examples of AGDs [3]: `atuqhuswcvjehti.com`, `test.takemuchexpression.club`, `ecfd206098b6b12d069f58e4da6d66c5f2.cc`, `gzauh44cvh14f52i35m29m29crn10g63dqd30e51c39.biz`

Detection Techniques

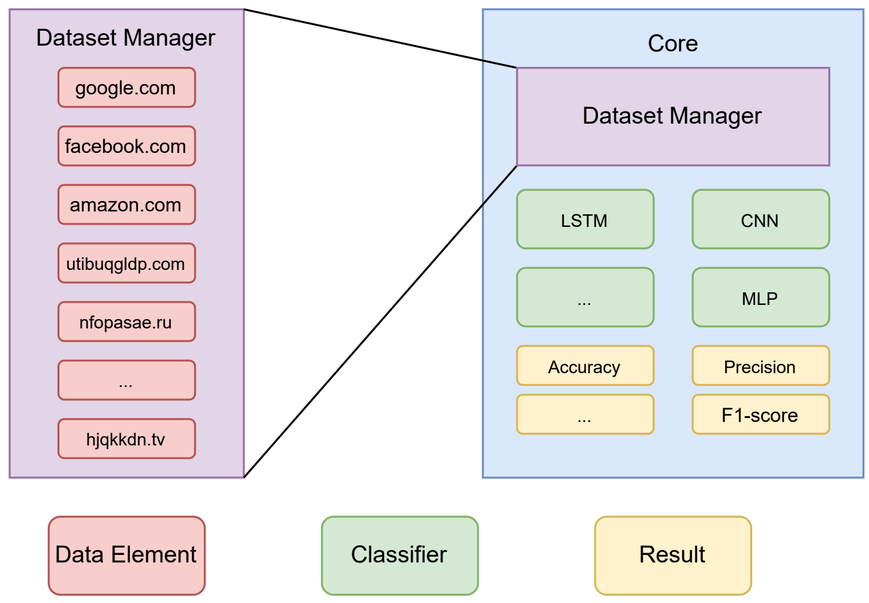
- Different solutions in the literature, the most recent ones adopting machine learning and deep learning approaches
- Most studies are based on the analysis of the domain name only
- There is no common methodology for comparing models with each other, even though the datasets follow the same structure

Preliminary Results

Model (Year)	Acc	Prec	Rec	F1	FPR	TPR	MCC	κ
LSTM [5] (2016)	95.42	97.39	95.69	96.53	5.12	95.69	89.82	0.8045
LSTM [7] (2017)	95.44	97.25	95.87	96.55	5.40	95.87	89.84	0.8059
CNN [7] (2017)	94.96	97.39	94.98	96.17	5.07	94.98	88.86	0.7849
LSTM [6] (2018)	95.02	96.82	95.67	96.24	6.27	95.67	88.88	0.7896
CNN [6] (2018)	92.94	96.29	92.99	94.61	7.16	92.99	84.49	0.7056
CMU [8] (2018)	94.87	**97.46**	94.77	96.10	**4.92**	94.77	88.69	0.7810
MIT [8] (2018)	**95.48**	96.96	**96.23**	**96.59**	6.03	**96.23**	**89.87**	**0.8083**
Parallel CNN [8] (2018)	93.48	96.64	93.48	95.03	6.49	93.48	85.68	0.7265
Baseline [8] (2018)	86.51	93.36	85.87	89.46	12.19	85.87	71.31	0.4745
MLP [8] (2018)	92.59	96.41	92.32	94.32	6.86	92.32	83.84	0.6907
CNN [1] (2019)	95.28	97.08	95.81	96.44	5.76	95.81	89.48	0.7998
Max Pooling [1] (2019)	90.48	95.62	89.84	92.64	8.21	89.84	79.53	0.6107
LSTM [1] (2019)	92.40	96.98	91.44	94.13	5.68	91.44	83.67	0.6804
LSTM + CNN [1] (2019)	83.88	94.12	80.87	86.99	10.09	80.87	67.44	0.3796
Bidireccional [1] (2019)	93.40	95.92	94.10	95.00	8	94.10	85.33	0.7261
DBD [4] (2019)	94.19	96.92	94.28	95.58	5.98	94.28	87.18	0.7545

Acc: Accuracy; **Prec**: Precision; **Rec**: Recall; **F1**: F1-score; **FPR**: False Positive Rate; **TPR**: True Positive Rate; **MCC**: Matthews's Correlation Coefficient; κ: Cohen's Kappa Score

Framework



- *Core* module includes all execution logic
- *Dataset Manager* manages dataset processing
- New models must adhere to the *Classifier* schema
- *Data Element* and *Result* represent training data and evaluation metrics, respectively. They are user-defined

Conclusions

- Our framework allows anyone to train and compare models in a simple and fast way
- Simpler models (but Baseline) tend to achieve better results when considering a large number of different malware families
  - They need to generalize more
  - More robust to detect DGAs from different families

References

[1] Berman, D.S.: DGA CapsNet: 1D Application of Capsule Networks to DGA Detection. Information **10**(5) (2019)

[2] Porras, P.A., Saïdi, H., Yegneswaran, V.: A Foray into Conficker's Logic and Rendezvous Points. LEET **9**, 7 (2009)

[3] Tuan, T.A., Anh, N.V., Luong, T.T., Long, H.V.: UTL_DGA22 A Dataset for DGA Botnet Detection and Classification. Computer Networks **221**, 109508 (2023)

[4] Vinayakumar, R., Soman, K.P., Poornachandran, P., Alazab, M., Jolfaei, A.: DBD: Deep Learning DGA-Based Botnet Detection, pp. 127–149. Springer (2019)

[5] Woodbridge, J., Anderson, H.S., Ahuja, A., Grant, D.: Predicting Domain Generation Algorithms with Long Short-Term Memory Networks. CoRR **abs/1611.00791** (2016)

[6] Yang, L., Liu, G., Zhai, J., Dai, Y., Yan, Z., Zou, Y., Huang, W.: A Novel Detection Method for Word-Based DGA. In: Sun, X., Pan, Z., Bertino, E. (eds.) Cloud Computing and Security. pp. 472–483. Springer International Publishing (2018)

[7] Yu, B., Gray, D.L., Pan, J., Cock, M.D., Nascimento, A.C.A.: Inline DGA Detection with Deep Networks. In: 2017 IEEE International Conference on Data Mining Workshops (ICDMW). pp. 683–692. IEEE (2017)

[8] Yu, B., Pan, J., Hu, J., Nascimento, A., De Cock, M.: Character Level based Detection of DGA Domain Names. In: 2018 International Joint Conference on Neural Networks (IJCNN). pp. 1–8. IEEE (2018)

Acknowledgements

This research was supported in part by TED2021-131115A-I00 (MIMFA), funded by MCIN/AEI/10.13039/501100011033, by the Recovery, Transformation and Resilience Plan funds, financed by the European Union (Next Generation), by the Spanish National Cybersecurity Institute (INCIBE) under *Proyectos Estratégicos de Ciberseguridad – CIBERSEGURIDAD EINA UNIZAR*, by the University, Industry and Innovation Department of the Aragonese Government under *Programa de Proyectos Estratégicos de Grupos de Investigación* (DisCo research group, ref. T21-23R), and by the RAPID project (Grant No. CS.007) financed by the Dutch Research Council (NWO).

Try It!

Contact data: reverseame@unizar.es