POSTER: Enabling Fair ML Evaluations for Security {References for Poster Timeline}

FEARGUS PENDLEBURY, Royal Holloway University of London & King's College London
FABIO PIERAZZI, Royal Holloway University of London & King's College London
ROBERTO JORDANEY, Royal Holloway University of London & King's College London
JOHANNES KINDER, Royal Holloway University of London
LORENZO CAVALLARO, King's College London

1 LIST OF REFERENCES

Table 1 reports the list of papers that may violate constraints C1, C2 and C3 that we describe in [37]. This is an extended and updated version of the list contained in [39], and applied to the constraints of our `Tesseract` framework [37]. The list considers papers published in the last 10 years but is not intended to be exhaustive. Please do not hesitate to contact us if you believe if some ✓ or ✗ need changes.

Table 1. References for papers in the timeline of poster [36]. We report papers between 2009 and 2018.

Year	Venue	C1	C2	C3	Domain	Notes
2018	USENIX Sec [33]	✗	✗	–	Tunneling	We were not able to assess C3 from the paper.
2018	USENIX Sec [38]	✗	✗	✓	Android	
2018	NDSS [35]	✗	✗	✓	Vulnerabilities	
2018	S&P [34]	✗	✗	✓	Scams	
2017	NDSS [31]	✓	✗	✗	Android	
2017	S&P [28]	✗	✗	✓	IP Domains	
2017	ESORICS [30]	✗	✗	✓	Android	
2017	CODASPY [32]	✗	✗	✓	Android	
2017	TDSC [29]	✗	✗	–	Android	C3 does not apply as they consider only malware clustering.
2016	CCS [27]	✗	✗	✓	Android	
2016	NDSS [24]	✗	✗	–	PDFs	We were not able to assess C3 from the paper.
2016	NDSS [26]	✗	✗	✗	Bots	
2016	S&P-MoST [25]	✗	✗	–	Android	C3 does not apply as they consider only malware clustering.
2015	CCS [23]	✗	✗	✗	Twitter	
2015	NDSS [22]	✗	✗	✓	Fake Accounts	
2014	NDSS [20]	✗	✗	✓	Android	
2014	WATeR [21]	✗	✗	✓	x86	
2013	CCS [19]	✗	✗	✓	URLs	
2013	CCS-AISec [17]	✗	✗	✓	Android	
2013	ICASSP [16]	✗	✗	✓	x86	
2013	NDSS [18]	✓	–	✗	PDF	We were not able to assess C2 from the paper.
2012	JMLR [14]	✗	✗	✗	x86	

Table 1. References for papers in the timeline of poster [36]. We report papers between 2009 and 2018.

Year	Venue	C1	C2	C3	Domain	Notes
2012	RAID [15]	✗	✗	✓	ActionScript	–
2012	ISSTA [12]	✗	✗	–	x86	We were not able to assess C3 from the paper.
2012	ACSAC [13]	✓	–	✓	PDF	We were not able to assess C2 from the paper.
2011	ACSAC [11]	✗	✗	✓	PDF	
2011	USENIX Sec [9]	✗	✗	✓	JavaScript	
2011	RAID [10]	✗	✗	✓	JavaScript	
2011	WWW [8]	✗	✗	✓	URLs	
2010	ESSoS [7]	✗	✗	✓	x86	
2010	ACSAC [6]	✗	✗	✓	JavaScript	
2010	NSDI [5]	✓	–	✓	Traffic	We were not able to assess C2 from the paper.
2010	WWW [4]	✗	✗	✓	JavaScript	
2009	MALWARE [3]	✗	✗	✓	x86	
2009	RAID [2]	✗	✗	✓	x86	
2009	ICC [1]	✗	✗	✗	Android	

REFERENCES

[1] A-D Schmidt, Rainer Bye, H-G Schmidt, Jan Clausen, Osman Kiraz, Kamer A Yuksel, Seyit Ahmet Camtepe, and Sahin Albayrak. Static analysis of executables for collaborative malware detection on android. In *IEEE ICC*, 2009.

[2] M Zubair Shafiq, S Momina Tabish, Fauzan Mirza, and Muddassar Farooq. PE-miner: Mining structural information to detect malicious executables in realtime. In *RAID*, 2009.

[3] Ronghua Tian, Lynn Batten, Rafiqul Islam, and Steve Versteeg. An automated classification system based on the strings of trojan and virus families. In *IEEE MALWARE*, 2009.

[4] Marco Cova, Christopher Kruegel, and Giovanni Vigna. Detection and analysis of drive-by-download attacks and malicious JavaScript code. In *WWW*, 2010.

[5] Roberto Perdisci, Wenke Lee, and Nick Feamster. Behavioral Clustering of HTTP-Based Malware and Signature Generation Using Malicious Network Traces. In *NSDI*, 2010.

[6] Konrad Rieck, Tammo Krueger, and Andreas Dewald. Cujo: Efficient detection and prevention of drive-by-download attacks. In *ACSAC*, 2010.

[7] Igor Santos, Felix Brezo, Javier Nieves, Yoseba K Penya, Borja Sanz, Carlos Laorden, and Pablo G Bringas. Idea: Opcode-sequence-based malware detection. In *ESSoS*, 2010.

[8] Davide Canali, Marco Cova, Giovanni Vigna, and Christopher Kruegel. Prophiler: A Fast Filter for the Large-scale Detection of Malicious Web pages. In *WWW*, 2011.

[9] Charlie Curtsinger, Benjamin Livshits, Benjamin G Zorn, and Christian Seifert. ZOZZLE: Fast and Precise In-Browser JavaScript Malware Detection. In *USENIX Security*, 2011.

[10] Mario Heiderich, Tilman Frosch, and Thorsten Holz. IceShield: Detection and mitigation of malicious websites with a frozen DOM. In *RAID*, 2011.

[11] Pavel Laskov and Nedim Šrndić. Static detection of malicious JavaScript-bearing PDF documents. In *ACSAC*, 2011.

[12] Davide Canali, Andrea Lanzi, Davide Balzarotti, Christopher Kruegel, Mihai Christodorescu, and Engin Kirda. A quantitative study of accuracy in system call-based malware detection. In *ISSTA*, 2012.

[13] Charles Smutz and Angelos Stavrou. Malicious pdf detection using metadata and structural features. In *ACSAC*, 2012.

[14] Gil Tahan, Lior Rokach, and Yuval Shahar. Mal-id: Automatic malware detection using common segment analysis and meta-features. *JMLR*, 2012.

[15] Timon Van Overveldt, Christopher Kruegel, and Giovanni Vigna. FlashDetect: ActionScript 3 Malware Detection. In *RAID*, 2012.

[16] George E Dahl, Jack W Stokes, Li Deng, and Dong Yu. Large-scale malware classification using random projections and neural networks. In *ICASSP*. IEEE, 2013.

[17] Hugo Gascon, Fabian Yamaguchi, Daniel Arp, and Konrad Rieck. Structural detection of android malware using embedded call graphs. In *AISec workshop*, 2013.

[18] Nedim Šrndic and Pavel Laskov. Detection of malicious pdf files based on hierarchical document structure. In *NDSS*, 2013.

[19] Gianluca Stringhini, Christopher Kruegel, and Giovanni Vigna. Shady Paths: Leveraging surfing crowds to detect malicious web pages. In *CCS*, 2013.

[20] Daniel Arp, Michael Spreitzenbarth, Malte Hubner, Hugo Gascon, and Konrad Rieck. DREBIN: Effective and Explainable Detection of Android Malware in Your Pocket. In *NDSS*, 2014.

[21] Zane Markel and Michael Bilzor. Building a machine learning classifier for malware detection. In *IEEE WATeR*, 2014.

[22] Yazan Boshmaf, Dionysios Logothetis, Georgos Siganos, Jorge Lería, Jose Lorenzo, Matei Ripeanu, and Konstantin Beznosov. Integro: Leveraging Victim Prediction for Robust Fake Account Detection in OSNs. In *NDSS*, 2015.

[23] Jonghyuk Song, Sangho Lee, and Jong Kim. CrowdTarget: Target-based detection of crowdturfing in online social networks. In *CCS*, 2015.

[24] Curtis Carmony, Xunchao Hu, Heng Yin, Abhishek Vasisht Bhaskar, and Mu Zhang. Extract Me If You Can: Abusing PDF Parsers in Malware Detectors. In *NDSS*, 2016.

[25] Santanu Kumar Dash, Guillermo Suarez-Tangil, Salahuddin Khan, Kimberly Tam, Mansour Ahmadi, Johannes Kinder, and Lorenzo Cavallaro. Droidscribe: Classifying Android Malware Based on Runtime Behavior. In *MoST-SPW*. IEEE, 2016.

[26] Eunjo Lee, Jiyoung Woo, Hyoungshick Kim, Aziz Mohaisen, and Huy Kang Kim. You are a Game Bot!: Uncovering Game Bots in MMORPGs via Self-similarity in the Wild. In *NDSS*, 2016.

[27] Ziyun Zhu and Tudor Dumitras. FeatureSmith: Automatically engineering features for malware detection by mining the security literature. In *CCS*. ACM, 2016.

[28] Sumayah Alrwais, Xiaojing Liao, Xianghang Mi, Peng Wang, XiaoFeng Wang, Feng Qian, Raheem Beyah, and Damon McCoy. Under the shadow of sunshine: Understanding and detecting bulletproof hosting on legitimate service provider networks. In *IEEE Symp. S&P*, 2017.

[29] Tanmoy Chakraborty, Fabio Pierazzi, and VS Subrahmanian. EC2: Ensemble clustering and classification for predicting android malware families. *IEEE Trans. Dependable and Secure Computing (TDSC)*, 2017.

[30] Kathrin Grosse, Nicolas Papernot, Praveen Manoharan, Michael Backes, and Patrick McDaniel. Adversarial examples for malware detection. In *ESORICS*. Springer, 2017.

[31] Enrico Mariconti, Lucky Onwuzurike, Panagiotis Andriotis, Emiliano De Cristofaro, Gordon Ross, and Gianluca Stringhini. MaMaDroid: Detecting Android Malware by Building Markov Chains of Behavioral Models. In *NDSS*, 2017.

[32] Guillermo Suarez-Tangil, Santanu Kumar Dash, Mansour Ahmadi, Johannes Kinder, Giorgio Giacinto, and Lorenzo Cavallaro. DroidSieve: Fast and Accurate Classification of Obfuscated Android Malware. In *ACM CODASPY*, 2017.

[33] Diogo Barradas, Nuno Santos, and Luís Rodrigues. Effective Detection of Multimedia Protocol Tunneling using Machine Learning. In *USENIX Security*, 2018.

[34] Amin Kharraz, William Robertson, and Engin Kirda. Surveylance: Automatically Detecting Online Survey Scams. In *IEEE Symp. S&P*, 2018.

[35] Zhen Li, Deqing Zou, Shouhuai Xu, Xinyu Ou, Hai Jin, Sujuan Wang, Zhijun Deng, and Yuyi Zhong. VulDeePecker: A Deep Learning-Based System for Vulnerability Detection. In *NDSS*, 2018.

[36] Feargus Pendlebury, Fabio Pierazzi, Roberto Jordaney, Johannes Kinder, and Lorenzo Cavallaro. POSTER: Enabling Fair ML Evaluations for Security. In *CCS*, 2018.

[37] Feargus Pendlebury, Fabio Pierazzi, Roberto Jordaney, Johannes Kinder, and Lorenzo Cavallaro. TESSERACT: Eliminating Experimental Bias in Malware Classification across Space and Time. *arXiv*, 2018.

[38] Octavian Suciu, Radu Mărginean, Yiğitcan Kaya, Hal Daumé III, and Tudor Dumitraş. When Does Machine Learning FAIL? Generalized Transferability for Evasion and Poisoning Attacks. *USENIX Security*, 2018.

[39] Bradley Austin Miller. *Scalable Platform for Malicious Content Detection Integrating Machine Learning and Manual Review*. University of California, Berkeley, 2015.