Fabrício José de Oliveira Ceschin

Federal University of Paraná, Paraná, Brazil
Department of Informatics, Computer Science
Curitiba, PR 82590-300

Email: fabricioceschin@gmail.com
Website: fabriciojoc.github.io
LinkedIn: linkedin.com/in/fabricioceschin
GitHub: github.com/fabriciojoc

Education

Feb 2018 – Present **Federal University of Paraná**
PhD, Machine Learning applied to Security
Curitiba, Paraná, Brazil
Honors: Fully-funded through Coordination for the Improvement of Higher Education Personnel (CAPES)

Feb 2016 – Feb 2018 **Federal University of Paraná**
MS, Machine Learning applied to Security
Curitiba, Paraná, Brazil
Honors: Fully-funded through Coordination for the Improvement of Higher Education Personnel (CAPES)

Feb 2012 – Dec 2015 **Federal University of Paraná**
BSc, Computer Science
Curitiba, Paraná, Brazil

Thesis

Ceschin, Fabrício, Gomes, Murilo Heitor, Oliveira, Luiz S., Grégio, André. Machine Learning (In) Security: A Stream of Problems. Pre-Dissertation Project. Abril 2021. Advisors: André Grégio, Heitor Murilo Gomes, and Luiz S. Oliveira.

Fabrício Ceschin, David Menotti, André Grégio. Need for Speed: Analysis of Brazilian Malware Classifiers' Expiration Date. February 2018. Master's Degree Thesis. Advisors: André Grégio and David Menotti.

Fabrício Ceschin, Bruno Müller. Web services and atomic transactions in distributed systems. December 2015. Undergraduate Thesis. Advisor: Bruno Müller.

Research & Professional Experience

Jul 2022 – Present **Federal University of Paraná and Samsung Research Brasil**
Researcher
Collaborative research with Samsung Research Brasil in mobile security.

Jun 2022 – Present **C3SL – Scientific Computing Center and Free Software, Federal University of Paraná, and National Service of Industrial Apprenticeship (SENAI)**
Researcher
Collaborative research with National Service of Industrial Apprenticeship (SENAI) in machine learning and natural language processing applied to jobs recommendation.

Feb – Mar 2020 **The University of Waikato, School of Computing and Mathematical Sciences**
Visitor PhD Student
Hamilton, Waikato, New Zealand
Collaborative research in machine learning applied to cybersecurity, data streams, and development of algorithms for scikit-multiflow library.

May 2019 and Aug 2018 **University of Florida, Department of Electrical and Computer Engineering**
Visitor PhD Student
Gainesville, Florida, United States
Collaborative research in machine learning applied to cybersecurity.

Mar 2014 – Feb 2016 **C3SL – Scientific Computing Center and Free Software**
Full Stack Web Developer
Curitiba, Paraná, Brazil
Full stack web development for Brazilian ministry of health systems.

May 2013 – Jan 2014 **ECOMP – Junior Computing Company**
Project Advisor
Curitiba, Paraná, Brazil
Manage and develop websites for academic entities.

Teaching & Presentations

2020 Course Assistant, Data Science for Cybersecurity, Federal University of Paraná.

2020 Presenter, CiDWeek - CiDAMO's I Data Science Week, Federal University of Paraná.

2020 Presenter, "No Need to Teach New Tricks to Old Malware: Winning an Evasion Challenge with XOR-based Adversarial Samples", Reversing and Offensive-Oriented Trends Symposium (ROOTS).

2019 Presenter, "Shallow Security: On the Creation of Adversarial Variants to Evade Machine Learning-Based Malware Detectors", Reversing and Offensive-Oriented Trends Symposium (ROOTS).

2019 Course Assistant, Data Science Course, Informatics Academic Week, Federal University of Paraná.

2019 Presenter, Machine Learning applied to Cybersecurity Course, Brazilian Security Symposium (SBSEG).

2017 Monitor, Undergraduate Algorithm Class, Federal University of Paraná.

Academic Service

2017, 2019, 2020	Reviewer for Brazilian Security Symposium (SBSeg) Conference.
2018	Reviewer for Annual Computer Security Applications Conference (ACSAC).
2019	Reviewer for Conference on Detection of Intrusions and Malware & Vulnerability Assessment (DIMVA).
2019	Reviewer for IEEE SecDev Conference.
2019	Reviewer for IEEE Security & Privacy Magazine.
2020	Reviewer for Australasian Conference on Information Security and Privacy (ACISP).
2020	Reviewer for The European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML).
2021	Reviewer for Expert Systems With Applications (ESWA) Journal.
2021	Reviewer for Frontiers in Big Data Journal.
2021	Reviewer for The International Conference on Availability, Reliability and Security (ARES).
2021	Reviewer for USENIX Security 2021 Artifact Evaluation.
2022	Reviewer for USENIX Security & Privacy 2022.

Awards, Grants & Competitions

Sep 2021	Machine Learning Security Evasion Competition (MLSEC) 2021, Attacker's Challenge – 1st Place
Sep 2021	Machine Learning Security Evasion Competition (MLSEC) 2021, Defender's Challenge – 1st Place
Sep 2020	Machine Learning Security Evasion Competition (MLSEC) 2020, Attacker's Challenge – 1st Place
Sep 2020	Machine Learning Security Evasion Competition (MLSEC) 2020, Defender's Challenge – 2nd Place
Sep 2019	Machine Learning Security Evasion Competition (MLSEC) 2019 – 2nd Place
May 2019	Secure and Private AI Scholarship from Facebook AI
Jan 2019	USENIX Enigma 2019 - Diversity Grant
Aug 2017	Google Research Awards for Latin America

Skills

Programming Languages	C, Python, Ruby, JavaScript, Java, PHP
Technical Skills	Machine Learning, Applied Machine Learning, Python Libraries for Data Science (such as Scikit-Learn, TensorFlow, Keras, Pytorch, Scikit-Multiflow,

	Pandas, NumPy, SciPy, Matplotlib, and Seaborn), Open-Source projects (already contributed to Scikit-Multiflow and TensorFlow), Data Streams, Concept Drift, Natural Language Processing, Adversarial Machine Learning, Malware Detection, Malware Classification, Computer Security, Web Development
Languages	English (Proficient), Portuguese (Native)

Projects

Fast & Furious: Malware Detection Data Stream Datasets. These datasets (DREBIN and AndroZoo) are contributions to the paper "Fast & Furious: On the Modelling of Malware Detection as an Evolving Data Stream". Note that these datasets are different from their original versions. The original DREBIN dataset does not contain the samples' timestamps, which I collected using VirusTotal API. My version of the AndroZoo dataset is a subset of reports from their dataset previously available in their APK Analysis API, which was discontinued. With samples' timestamps, we can simulate real world data streams.

Scikit-Multiflow. Scikit-Multiflow is an open-source machine learning package for streaming data. It extends the scientific tools available in the Python ecosystem. Scikit-Multiflow is intended for streaming data applications where data is continuously generated and must be processed and analyzed on the go. Data samples are not stored, so learning methods are exposed to new data only once.

Corvus_. Corvus_ is a dynamic analysis system for malware targeting Windows, Linux, and Android, basically the result of multiple developments created by my research group (SECRET). Some machine learning models for malware detection developed by me are available as detection engines for Windows Portable Executable (PE) files. Corvus is currently available in a beta version, so any feedback and/or bug report is appreciated.

Machine Learning applied to Cyber Security Course. The massive amount of data produced by security solutions have been creating a strong dependency on automated methods for knowledge discovery. Attacks against computer systems make use of several transmission channels and formats (e.g., network traffic, binary files, text, chained system calls etc.), which difficult their observation among unsuspicious data. Machine learning techniques are a great aid for separating data into classes, but they need to be correctly deployed. In the Machine Learning applied to Cyber Security Course, I show how to adequately apply machine learning algorithms to the security data science process. To do so, it discusses key concepts about the subject and present practical examples with free, open-source tools.

Brazilian Malware Dataset. The Brazilian Malware Dataset contains thousands (~50K) of malware and goodware collected in the Brazilian cyberspace over years and it is frequently updated with new samples. This dataset was used in the paper 'The Need for Speed: An Analysis of Brazilian Malware Classifiers'.

Publications

Ceschin, Fabrício, Botacin, Marcus, Gomes, Heitor Murilo, Pinagé, Felipe, Oliveira, Luiz S., Grégio, André. Fast & Furious: Modelling Malware Detection as Evolving Data Streams. Expert Systems with Applications. https://doi.org/10.1016/j.eswa.2022.118590. August 2022.

Giovanini, Luiz, Ceschin, Fabrício, Silva, Mirela, Chen, Aokun, Kulkarni, Ramchandra, Banda, Sanjay, Lysaght, Madison, Qiao, Heng, Sapountzis, Nikolaos, Sun, Ruimin, Matthews, Brandon, Wu, Dapeng Oliver, Grégio, André, Oliveira, Daniela. Online Binary Models are Promising for Distinguishing Temporally Consistent Computer Usage Profiles. IEEE Transactions on Biometrics, Behavior, and Identity Science. 10.1109/TBIOM.2022.3179206. June 2022.

Botacin, Marcus, Domingues, Felipe Duarte, Ceschin, Fabrício, Machnicki, Raphael, Alves, Marco Antonio Zanata, de Geus, Paulo Lício, Grégio, André. AntiViruses under the Microscope: A Hands-On Perspective. Computers & Security, pp. 102500, 2021, ISSN: 0167-4048. 10.1016/j.cose.2021.102500. January 2022.

Castanhel, Gabriel R., Heinrich, Tiago, Ceschin, Fabrício, Maziero, Carlos. Taking a Peek: An Evaluation of Anomaly Detection Using System calls for Containers. 2021 IEEE Symposium on Computers and Communications (ISCC). 10.1109/ISCC53001.2021.9631251. December 2021.

Silva, Mirela, Ceschin, Fabrício, Shrestha, Prakash, Brant, Christopher, Gilda, Shlok, Fernandes, Juliana, Silva, Catia S., Grégio, André, Oliveira, Daniela, Giovanini, Luiz. People Still Care About Facts: Twitter Users Engage More with Factual Discourse than Misinformation--A Comparison Between COVID and General Narratives on Twitter. arXiv:2012.02164. September 2021.

Botacin, Marcus, Moia, Vitor Hugo Galhardo, Ceschin, Fabricio, Henriques, Marco Amaral A, Grégio, André. Understanding uses and misuses of similarity hashing functions for malware detection and family clustering in actual scenarios. Forensic Science International: Digital Investigation, 38 , pp. 301220, 2021, ISSN: 2666-2817. 10.1016/j.fsidi.2021.301220. September 2021.

Botacin, Marcus; Ceschin, Fabricio; Sun, Ruimin; Oliveira, Daniela; Grégio, André. Challenges and pitfalls in malware research. Computers & Security, pp. 102287, 2021, ISSN: 0167-4048. 10.1016/j.cose.2021.102287. July 2021.

Ceschin, Fabrício, Botacin, Marcus, Lüders, Gabriel, Gomes, Heitor Murilo, Oliveira, Luiz S., Grégio, André. No need to teach new tricks to old malware: Winning an evasion challenge withxor-based adversarial samples. Proceedings of the 3rd Reversing and Offensive-Oriented Trends Symposium, Association for Computing Machinery, Vienna, Austria, 2020, ISBN: 9781450377751. 10.1145/3433667.3433669. November 2020.

Castanhel, Gabriel Ruschel, Heinrich, Tiago, Ceschin, Fabrício, Maziero, Carlos. The Impact of Trace Size in Anomaly Detection System for Containers Through Machine Learning. ERRC - WRSeg 2020. 10.5753/errc.2020.15203. November 2020.

Ceschin, Fabrício, Gomes, Murilo Heitor, Botacin, Marcus, Bifet, Albert, Pfahringer, Bernhard, Oliveira, Luiz S., Grégio, André. Machine Learning (In) Security: A Stream of Problems. arXiv:2010.16045. October 2020.

Lüders, Gabriel, Botacin, Marcus, Ceschin, Fabrício, Grégio, André. Breaking Good: Injeção de Payloads Legítimos em Binários Maliciosos para Teste de Robustez de Antivírus contra Evasão. IV Salão de Ferramentas - SBSEG 2020. 10.5753/sbseg_estendido.2020.19273. October 2020.

Castanhel, Gabriel Ruschel, Heinrich, Tiago, Ceschin, Fabrício, Maziero, Carlos. Detecção de Anomalias: Um Estudo Voltado na Identificação de Ataques no Ambiente de Contêiner. XIV Workshop de Trabalhos de Iniciação Científica e de Graduação (WTICG) - SBSEG 2020. 10.5753/sbseg_estendido.2020.19283. October 2020.

Botacin, Marcus, Ceschin, Fabricio, de Geus, Paulo, Grégio, André. We Need to Talk About AntiViruses: Challenges & Pitfalls of AV Evaluations. Computers & Security, pp. 101859, 2020, ISSN: 0167-4048. 10.1016/j.cose.2020.101859. August 2020.

Ceschin, Fabrício, Botacin, Marcus, Gomes, Heitor Murilo, Oliveira, Luiz S, Grégio, André. Shallow Security: On the Creation of Adversarial Variants to Evade Machine Learning-Based Malware Detectors. Proceedings of the 3rd Reversing and Offensive-Oriented Trends Symposium, Association for Computing Machinery, Vienna, Austria, 2019, ISBN: 9781450377751. 10.1145/3375894.3375898. November 2019.

Ceschin, Fabrício, S. Oliveira, Luiz, Grégio, André. Aprendizado de Máquina para Segurança: Algoritmos e Aplicações. Mini Cursos - XIX Simposio Brasileiro de Segurança da Informacão ao e de Sistemas Computacionais - SBSeg 2019. 10.5753/sbc.8589.4. September 2019.

Beppler, Tamy, Botacin, Marcus, Ceschin, Fabrício, Oliveira, Luiz S, Grégio, André. L(a)ying in (Test)Bed: How Biased Datasets Produce Impractical Results for Actual Malware Families' Classification. Information Security, pp. 381–401, Springer International Publishing, Cham, 2019, ISBN: 978-3-030-30215-3. 10.1007/978-3-030-30215-3_19. September 2019.

Botacin, Marcus, Galante, Lucas, Ceschin, Fabricio, Santos, Luigi Carro Paulo Cesar, de Geus, Paulo Licio, Gregio, Andre, Zanata, Marco. The AV says: Your hardware definitions were updated! 14th International Symposium on Reconfigurable Communication-centric Systems-on-Chip (ReCoSoC 2019), IEEE, 2019, ISBN: 978-1-7281-4770-3. 10.1109/ReCoSoC48741.2019.9034972. July 2019.

Ceschin, Fabrício, Pinage, Felipe, Castilho, Marcos, Menotti, David, Oliveira, Luis S, Gregio, André. The need for speed: An analysis of brazilian malware classifers. IEEE Security Privacy, 16 (6), pp. 31-41, 2018, ISSN: 1540-7993. 10.1109/MSEC.2018.2875369. November 2018.

Ceschin, Fabrício, Menotti, David, Castilho, Marcos, Grégio, André. Avaliação da Eficácia de Classificadores de Malware ao Longo do Tempo. Workshop de Forense Computacional - SBSEG 2017. Anais do XVII SBSeg, 2017. p. 1-10. November 2018.