Fabrício Ceschin

+1 470-927-2955 | fabricioceschin@gmail.com | linkedin.com/in/fabricioceschin | github.com/fabriciojoc

Education

Federal University of Paraná — Curitiba, PR, Brazil

Ph.D. in Computer Science, Machine Learning applied to Security — *Feb. 2018 – Dec. 2022*

- Presented talk "Spotting the Differences: Quirks of Machine Learning (in) Security" at USENIX Enigma 2023
- Three-time winner of Microsoft's Machine Learning Security Evasion Competition (MLSEC) in 2020 and 2021
- Secure and Private AI Scholarship from Facebook AI in 2019
- Wrote and published 20 articles based on thesis research in peer-reviewed journals and conferences
- Best thesis award by the Federal University of Paraná's Computer Science department.

Federal University of Paraná — Curitiba, PR, Brazil

M.S. in Computer Science, Machine Learning applied to Security — *Feb. 2016 – Feb. 2018*

- Awarded by Google Research Awards for Latin America in 2017
- Wrote and published 2 articles based on thesis research in peer-reviewed journals and conferences

Federal University of Paraná — Curitiba, PR, Brazil

BSc in Computer Science — *Feb. 2012 – Dec. 2015*

- Ranked 1st in a class of 80 students, GPA 7.7/10

Experience

Research Scientist — Jul. 2023 – Present

Georgia Institute of Technology — *Atlanta, GA, USA*

- Developed a Large Language Model for forensics analysis using system and application logs
- Explored different strategies to handle logs with Large Language Models using PyTorch and Hugging Face

Researcher — Jan. 2023 – Jul. 2023

Federal University of Paraná & Samsung Research Brazil — *Curitiba, PR, Brazil*

- Developed a realistic evaluation framework for cybersecurity considering its particularities, leading the project using agile development methodology (SCRUM), with Scikit-Learn modules, API documentation, and unit tests

Visitor Researcher — Feb. 2020 – Mar. 2020

The University of Waikato, School of Computing and Mathematical Sciences — *Hamilton, Waikato, New Zealand*

- Developed a delayed evaluation module for Scikit-Multiflow, an open-source machine learning package for streaming data in Python
- Conducted collaborative research on machine learning and data streams applied to cybersecurity

Visitor Researcher — May 2019 and Aug. 2018

University of Florida, Department of Electrical and Computer Engineering — *Gainesville, Florida, United States*

- Conducted collaborative research on machine learning applied to cybersecurity, focusing on continuous authentication challenges

Full Stack Web Developer — Mar. 2014 – Feb. 2016

C3SL – Scientific Computing Center and Free Software — *Curitiba, PR, Brazil*

- Developed a full-stack web application using Ruby, Javascript, and PostgreSQL to the Brazilian ministry of health

Project Advisor — May 2013 – Jan. 2014

ECOMP – Federal University of Paraná Junior Computing Company — *Curitiba, PR, Brazil*

- Developed a REST API using FastAPI and PostgreSQL to store data from learning management systems
- Developed a full-stack web application using Flask, React, PostgreSQL and Docker to analyze GitHub data
- Explored ways to visualize GitHub collaboration in a classroom setting

Projects

Corvus_ | *Python, Javascript, Django, RabbitMQ, PostgreSQL, HTML/CSS, Git* Feb. 2017 – Mar. 2023

- Developed a web-based dynamic analysis system for malware, including some tools developed in the research group
- Developed a REST API using swagger with Django rest Framework, integrating with other tools and automating submissions and analysis
- Deployment of all the machine learning models developed in the Machine Learning Security Evasion Competition (MLSEC) to detect Windows Portable Executable (PE) malware

Scikit-Multiflow | *Python, Git* Feb. 2020 – May 2020

- Developed a delayed evaluation module that consider samples‘ timestamps in the stream
- Collaborated with the library authors to make the module available in the Python package repository, developing unit tests and a synthetic dataset for them

Malware Detection Data Stream Datasets | *Python* Jul. 2019 – Jan. 2020

- Alternative versions of the origional DREBIN and AndroZoo datasets containing 10 years of benign and malicious Android applications temporal information
- Created a script that collected more than 400K samples including their first appearance date in the wild, a crucial information for a realistic evaluation

Machine Learning applied to Cyber Security Course | *Python, Jupyter Notebook, Git* Jul. 2019 – Ago. 2019

- Developed the course to introduce machine learning applied to cyber security considering the pipeline and all the steps required to build a solution
- Base material for the course “CI305/CI1030 – Data Science for Security” at Federal University of Paraná, Brazil

Brazilian Malware Dataset | *Python* May. 2018 – Nov. 2020

- Collected 50K malign and benign Windows applications collected in the Brazillian cyberspace over years to evaluate concept drift in malware detectors

Selected Publications

Machine Learning (In) Security: A Stream of Problems
Digital Threats: Research and Practice 2023

- Fabrício Ceschin, Marcus Botacin, Albert Bifet, Bernhard Pfahringer, Luiz S. Oliveira, Heitor Murilo Gomes, André Grégio

Fast & Furious: Modelling Malware Detection as Evolving Data Streams
Expert Systems with Applications 2022

- Fabrício Ceschin, Marcus Botacin, Heitor Murilo Gomes, Felipe Pinagé, Luiz S. Oliveira, André Grégio

Online Bin. Models are Promising for Distinguishing Temp. Consistent Computer Usage Profiles
IEEE Transactions on Biometrics, Behavior, and Identity Science 2022

- Luiz Giovanini and Fabrício Ceschin, Mirela Silva, Aokun Chen, Ramchandra Kulkarni, Sanjay Banda, Madison Lysaght, Heng Qiao, Nikolaos Sapountzis, Ruimin Sun, Brandon Matthews, Dapeng Oliver Wu, André Grégio, Daniela Oliveira

Challenges and pitfalls in malware research
Computers & Security 2021

- Marcus Botacin, Fabrício Ceschin, Ruimin Sun, Daniela Oliveira, André Gregio

Winning an evasion challenge with xor-based adversarial samples
Reversing and Offensive-Oriented Trends Symposium 2020

- Fabrício Ceschin, Marcus Botacin, Gabriel Lüders, Heitor Murilo Gomes, Luiz S. Oliveira, André Gregio

On the Creation of Adversarial Variants to Evade Machine Learning-Based Malware Detectors
Reversing and Offensive-Oriented Trends Symposium 2019

- Fabrício Ceschin, Marcus Botacin, Heitor Murilo Gomes, Luiz S. Oliveira, André Gregio

The need for speed: An analysis of brazilian malware classifers
IEEE Security & Privacy 2018

- Fabrício Ceschin, Felipe Pinage, Marcos Castilho, David Menotti, Luis S. Oliveira, André Gregio

Technical Skills

Languages: Python, Java, C/C++, C#, SQL, JavaScript, Ruby, Skala, PHP, HTML/CSS
Frameworks: Django, Flask, Vue.js, WordPress
Developer Tools: Git, Docker, Swagger, Amazon AWS Cloud, Microsoft Azure Cloud, VS Code, Visual Studio, PyCharm, IntelliJ, Eclipse, Jupyter Notebook, Android Studio
Libraries: Scikit-Learn, Scikit-Multiflow, TensorFlow, Keras, PyTorch, PyArrow, Pandas, NumPy, SciPy, Matplotlib