Luis Vargas

http://luisvargas.me

Email : lvarg025@gmail.com
Mobile : 305-934-4931

Education

- **University of Florida** — Gainesville, FL
  Doctor of Philosophy: Computer Engineering – Cybersecurity; Advisor: Patrick Traynor — *Summer 2021*
- **University of Florida** — Gainesville, FL
  Master of Science, Computer Science — *Fall 2020*
- **Florida International University** — Miami, FL
  Bachelor of Science, cum laude, Computer Engineering — *Fall 2015*

Relevant Graduate Courses

- Advanced Data Science, Computer Networks Security, Introduction to Cryptography, Computer Information Security, Mobile Security, Engineering Entrepreneurship

Skills

- **Programming Languages**: Proficient: (Python); Familiar: (MATLAB, SQL, Java, Elixir)
- **Technologies**: Software: (LaTeX, Git, MongoDB, Spark, Audacity, NLP packages (Gensim, SpaCy, Huggingface, SparkNLP), TensorFlow (Keras), Dash, Ansible, NetworkX, Zeek); OS: (OS X, Ubuntu, Windows)
- **Languages**: English (fluent), Spanish (fluent)

Experience

- **Data Scientist** — Washington D.C. - Remote
  Alethea Group — *October 2021 – Present*
  - Developed machine learning/deep learning tools (NLP techniques) to automatically label, tag, and visualize large data corpora to improve and facilitate intelligence extraction
  - Designed an analytic dashboard to help track trends, network coordination, and statistical changes of disinformation narratives.
  - Communicated (presentation/reports) results of analytic and data science tools to executives and analysts of non-technical backgrounds
- **Ph.D. Research Assistant** — Gainesville, FL
  University of Florida — *June 2016 – August 2021*
  - My Ph.D. research interest lies in the intersection between data science, cybersecurity, and disinformation. My focus is on better understanding the deployment challenges of tools used to detect the spread of fake news and the detections of information operations (disinformation campaigns).
  - Researched the SMS usage of many online services based on 900,000 messages sent to public gateways. We uncovered phone verification fraud, poor entropy generation of one-time codes and sensitive information being sent over an insecure medium
  - Developed a data pipeline (all aspects of E.T.L. and data warehousing) and analyzed traffic (Zeek DNS/TLS/Conn logs) from a real hospital network to detect and classify potentially compromised devices
  - Proposed and implemented a systems that uses time, instead of a secret, as the encryption mechanism (i.e., fast encryption/slow decryption) to protect long term data at rest
- **Data Science - Threat Intelligence (Information Operations) Intern** — Reston, VA
  FireEye — *Summer 2019 and Summer 2020*
  - Developed python-based analysis tools to extract and visualize key insights from a large corpora of unstructured data to improve intelligence reporting
  - Developed and analyzed deepfake image detection models using Keras to enhance intelligence reporting
  - Created web crawlers with a database integration to record historical events and streamline the threat intelligence analysis process

- Communicated (presentations/reports) results from the analysis tools to audiences with multiple non-technical backgrounds

- **Flight Controls R&D Intern** — Mountain View, CA
  NASA Ames Research Center — *Fall 2014 and Summer 2015*
  - Researched possible function allocations in aircraft separation to support future growth for air traffic controllers
  - Created MATLAB automated scripts to compare and streamline the metric analysis process of multiple simulations with various parameters
  - Developed scripts (JavaScripts) to dynamically view simulated air traffic to identify bottleneck points and macro patterns in an airspace
  - Gave multiple technical presentation to a room full of researchers to convey ideas and progress

- **Manufacturing Engineering Systems Intern** — Dearborn, MI
  Ford Motor Company — *Summer 2014*
  - Processed and refined Big Data from the Connected Vehicle Analytic Challenge using High Performance Computing and SQL scripting
  - Trained Natural Language Processing models (entity recognition) to support more accurate business output from Securities and Exchange Commission files
  - Installed a distributive computing network that allowed faster processing of manufacturing Big Data that improved processing speed by 60%

Projects

- **Measuring the Impact of Language in the Fake News Detection Problem**: Created a data analysis pipeline to measure the difference in the diffusion of Spanish and English-driven spread of online news articles. The pipeline included data collection using Twitter's APIv2, cleaning and parsing query results, extracting diffusion networks from tweet activity, featurizing the diffusion networks, and training machine learning models.
- **Detection of Information Operation (disinformation campaign) Activity in Twitter**: Developed an early detection mechanism to detect when Twitter accounts coordinate in a similar disinformation campaigns to identify possible malicious accounts. The data processing included collection of benign and malicious "coordinated" Twitter communities, extraction of coordination networks from the accounts within the communities, featurizing the coordination networks, and training machine learning models.
- **Hospital Network Traffic Analysis**: Created the data analysis pipeline for network traffic collected (Zeek logs) from a hospital. This included all parts of the ETL process: making crawlers to various OSINT website, cleaning/filtering data, storing data in a data warehouse model. The big data was then analyzed using Spark and Python to provide a security analysis of the devices inside the hospital.

Selected Publications

[1] Logan Blue, Kevin Warren, Hadi Abdullah, Cassidy Gibson, **Luis Vargas**, Jessica O'Dell, Kevin Butler, and Patrick Traynor. Who Are You (I Really Wanna Know)? Detecting Audio DeepFakes Through Vocal Tract Reconstruction. In *31st USENIX Security Symposium (USENIX Security 22)*, 2022.

[2] **Luis Vargas**, Patrick Emami, and Patrick Traynor. On the Detection of Disinformation Campaign Activity with Network Analysis. In *Proceedings of the 2020 ACM SIGSAC Cloud Computing Security Workshop*, CCSW '20, 2020.

[3] **Luis Vargas**, Logan Blue, Vanessa Frost, Chritopher Patton, Nolen Scaife, Kevin R.B. Butler, and Patrick Traynor. Digital Healthcare-Associated Infection: A Case Study on the Security of a Major Multi-Campus Hospital System. In *Proceedings of the 2019 Network and Distributed System Security Symposium*, NDSS '19, 2019.

[4] Bradley Reaves, **Luis Vargas**, Nolen Scaife, Dave Tian, Logan Blue, Patrick Traynor, and Kevin R.B. Butler. Characterizing the Security of the SMS Ecosystem with Public Gateways. *ACM Transactions on Privacy and Security*, 2018.

[5] **Luis Vargas**, Gyan Hazarika, Rachel Culpepper, Kevin R.B. Butler, Thomas Shrimpton, Doug Szajda, and Patrick Traynor. Mitigating Risk while Complying with Data Retention Laws. In *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*, CCS '18, 2018.