Kulsoom Abdullah, PhD

Email : kulsoom@gatech.edu

Profile websites: LinkedIn, Github, Google Scholar, Meetup, Personal

Skills summary

- **Computer Languages & Tools**: Python, Spark (PySpark and Scala), Hive SQL, Java, JOGL (Java OpenGL extensions), C, Hping/Tcl, Sockets, LaTeX **Proficiency in**: R, AWS, D3.js
- **Foreign Languages**: Pashto (basic proficiency), Urdu & Arabic (limited proficiency)

Education

- **Georgia Institute of Technology** Atlanta, GA
  PhD in Electrical & Computer Engineering; GPA: 3.61/4.0 *Aug. 2000 – May 2006*
  - **Dissertation**: Scaling and Visualizing Network Data to Facilitate in Intrusion Detection Tasks
    - ∗ Developed two systems that visualize packet header & IDS data. Visualization improved the efficiency at which attacks & anomalies are identified by the network security analyst.
- **Georgia Institute of Technology** Atlanta, GA
  Masters of Science in Electrical Engineering ; GPA: 3.57/4.0 *Aug. 1998 – May 2000*
- **University of Central Florida** Orlando, FL
  Bachelors of Science in Computer Engineering ; GPA: 3.8/4.0 *Aug. 1994 – May 1998*

Selected Experience

- **ADP** Atlanta, GA
  Data Scientist *March 2016 - Present*
  - **Ventures/Monetization**: A lead data scientist within ADP Ventures team, which developed and delivers new patented data products delivered via UI and API feeds serving real estate and capital markets. Product based on geo and sector insights by aggregating and anonymizing ADP's payroll data of 30 million US employees.
    - ∗ Led new efforts to machine datasets to observe and present employee workforce migration patterns down to the block level (patented product).
    - ∗ Collaborated with peers, Sr. Leadership and product owners to iterate products based on additional feedback from clients and sales team, resulting in the execution of ADP's data monetization strategy.
    - ∗ Managed data quality deliverables to identify missing, incomplete data and determined outliers to help remove noise and add additional signal furthering product reach.
    - ∗ Demonstrated the potential for improved predictive models using ADP data versus Census data.
  - **Open Data Science**: Supported ADP's Open Data science initiatives by researching and prototyping projects, questioning existing processes, updating and documenting findings. Educated peers, interns, business owners and third party vendors. Former member of ADP's Data Science Innovation lab in partnership with Georgia Tech to select and mentor graduate students on various data science initiatives.
  - **Tools used**: Hive, Spark (Python, some Scala), Python and libraries such as Scikit-Learn, NLP to classify companies (TF/IDF, Word2Vec), employee census geocoding, Unix shell scripting, Bitbucket/Github, Jupyter, Confluence/wiki documentation, VIM, Tableau (EDA)
- **Damballa, Inc.** Atlanta, GA
  Research Scientist *Jun 2014 - Aug 2015*
  - **Infected Client Tracking Experiment**:
    Analyzed malware botnet command & control (c2) domains and network traffic using machine learning, domain attribution and classification to identify those that are malicious.
    - ∗ Discovered 50% of IP address churn (IP address changing for a client) during the data exploration stage.
    - ∗ Examined client query behavior by independently designing and coding experiments to test hypothesis. Used the results to come up with statistical features to classify clients that are following malicious domains specific to malware campaigns deployed by botnet operators.
  - **Tools used**: Mining DNS data, Map/Reduce with Python streaming api, Hadoop, R (EDA and hierarchical clustering), LDA to cluster domains into topics, Unix Shell scripting, GitHub, SQLITE, Jupyter

- **Georgia Tech-CAP Group** Atlanta, GA
  Visiting Scholar *July 2009-Sep 2015*
  - **Visualizing Domain Reputation & Attribution**: Leveraged DNS agility, used by malware command-and-control (C2) system, for reputation & attribution. Clustered domains on network features and relationships, then visualized the most important relationships. Historical IP and domain data that was used for this project came from a security company's passive DNS database.
  - **Tools used**: Python, D3.js, Sci-kit learn,Github
  - **Parallel 3D Coordinate System**: Researched & developed a network security stereoscopic 3D visualization and machine learning for inexperienced users (command/navigation recommendation based on expert users), and user evaluation. Reference the publications.
  - **Tools used**: Hping/Tcl, Wireshark, Snort IDS
- **Georgia State University** Atlanta, GA
  Consultant *Dec 2012-Apr 2013*
  - **CSEM Research**: Research in the Internet's role in Commercial Sexual Exploitation of Minors (CSEM), mostly self-supported. Collaborated with Dr. Mary Finn at Georgia State University
  - **Training**: Educated & trained investigators & research assistants about the current communication technologies & how they interface, assisted with technology content-related questions in qualitative interview, assist with the collection & analysis of web data.
  - **Data Collection**: Scraped web forum discussion data HTML pages of over one year from 3 boards. Tokenized each post & parameters (parent & child posts, userid, subject, date, time, message) from each HTML page, storing into MongoDB using Python. Used cluster analysis for classification results, along with professor feedback, to help reduce the dictionary. E.g. Atlanta forum was 39K posts by 113K unique words.
  - **Tools used**: Sci-kit learn, TF-IDF, MongoDB
- **Scientific Research Corporation** Atlanta, GA
  Engineer III *Jan 2008-Jun 2009*
  - **Software Defined Radios**: Assisted in developing an adaptive power control algorithm in Qualnet, validating a thorough network testing plan & tested SRC's revised Mobile Route (MANET software).
  - **JTRS Training Network Analysis**: Assessed the unicast and geocast network loading for FCS training & JTRS radios & waveforms to carry this load.
  - **SPAWAR SATCOM Analysis**: Generated VTRPE input files for varying frequencies, wind speed, sea conditions to simulate a general channel model to test Reliable Link Protocol (RLP) performance over the sea.
- **Georgia Tech** Atlanta, GA
  Graduate Research Assistant *Aug 2001-May 2006*
  - **Research topics**: wireless communications, network security, & visualization.
    - Information visualization, GUI design, & user study research.
  - **Network Security Data Visualization**: Dealt with general visualization & HCI issues of network security data - scaling 65535 TCP & UDP ports & 4 billion possible IPv4 addresses, time scales for various activity detection.
  - **Tech and Data used**: Georgia Tech campus's IDS output, Java and OpenGL

Talks/Teaching/Volunteer/Personal

- **Personal projects**: GitHub
  - Regression
    - Predicting bike rental count: EDA, feature significance, and comparison of regression models
    - Predicting superconductor temperature: EDA, linear regression deep dive and comparison of regression models
  - Classification
    - Recipe cuisine classification (from scratch): serverless to deploy - data collection (web scraping), using AWS Lambda functions, processing and cleaning the data. Comparison of shallow ML and Deep Learning model results, deployment as a cloud service, scheduled re-run of the system to get any new recipes, and retrain the model. TDE blog post, GitHub
    - Strength movement quality classification: compared models for both binary and multiclass cases, random and participant based test/train splitting, window and lag time features
- **Talks**:

  - Docker for Data Science, Atlanta Code Camp, Atlanta, October 2018
  - Thematic panel session presenter: "Understanding the Role of the Internet in the Sale of Sex with Minors," M. Finn, L. Stalans, & K. Abdullah. 2013 American Society of Criminology Meeting – Expanding the Core: Neglected Crimes, Groups, Causes and Policy Approaches, November 20-23, 2013, Atlanta, GA, USA.

- **Volunteer**: Mentor at Anidata (non-profit) in Atlanta, GA from Jan 2016 - July 2018. Guided future data scientists on data projects that improve and benefit the community. Assisted in Python workshop, curriculum and teaching.
- **Hobbies**: Olympic Weightlifting, Photography, Chorus, general diet/health/fitness
- **Bio**: Kulsoom Abdullah is a Pakistani-American competitive Olympic Weightlifter and received her Crossfit Level I certification. Her website LiftingCovered.com and Facebook page document her experiences weightlifting in an effort to compete at U.S. national competitions. She advocated competing in clothing that adheres to religious codes, opening the door for women from cultures around the world to compete and move beyond preconceived notions of gender, race, and religion. Her efforts led to an invitation to deliver remarks following then Secretary of State Hillary Clinton at the U.S. State Department's Eid ul Fitr reception 2011. She represented Pakistan as the first female at the international level to compete at the 2011 World Weightlifting Championships. She resides in Atlanta, GA and is currently taking a break from major competitions while continuing training. She continues to support causes which focus on empowerment of the marginalized and speaks at events, and to the news media about her story and insights.

Selected Publications and Reports

1. Troy Nunnally, Kulsoom Abdullah, A. Selcuk Uluagac, John A. Copeland, and Raheem Beyah. NAVSEC: A Recommender System for 3D Network Security Visualizations. In *Proceedings of the Tenth Workshop on Visualization for Cyber Security*, VizSec '13, pages 41–48, New York, NY, USA, 2013. ACM

2. T. Nunnally, P. Chi, K. Abdullah, A. S. Uluagac, J. A. Copeland, and R. Beyah. P3D: A parallel 3D coordinate visualization for advanced network scans. In *2013 IEEE International Conference on Communications (ICC)*, pages 2052–2057, June 2013

3. K. Abdullah, G. Conti, and R. Beyah. A Visualization Framework for Self-Monitoring of Web-Based Information Disclosure. In *2008 IEEE International Conference on Communications*, pages 1700–1707, May 2008

4. Cited in: Greg Conti. *Security Data Visualization: Graphical Techniques for Network Analysis*. No Starch Press, 2007

5. Kulsoom Abdullah, Gregory Conti, and Edward Sobiesk. Self-monitoring of Web-based Information Disclosure. In *Proceedings of the 2007 ACM Workshop on Privacy in Electronic Society*, WPES '07, pages 56–59, New York, NY, USA, 2007. ACM

6. Kulsoom Abdullah and John A. Copeland. Tool Update: High Alarm Count Issues in IDS Rainstorm. In *Proceedings of the 3rd International Workshop on Visualization for Computer Security*, VizSEC '06, pages 61–62, New York, NY, USA, 2006. ACM

7. G. Conti, K. Abdullah, J. Grizzard, J. Stasko, J. A. Copeland, M. Ahamad, H. L. Owen, and C. Lee. Countering security information overload through alert and packet visualization. *IEEE Computer Graphics and Applications*, 26(2):60–70, March 2006

8. K. Abdullah, C. P. Lee, G. Conti, J. A. Copeland, and J. Stasko. IDS RainStorm: visualizing IDS alarms. In *IEEE Workshop on Visualization for Computer Security, 2005. (VizSEC 05).*, pages 1–10, Oct 2005

9. Kulsoom Abdullah, C. Lee, G. Conti, and J. A. Copeland. Visualizing network data for intrusion detection. In *Proceedings from the Sixth Annual IEEE SMC Information Assurance Workshop*, pages 100–108, June 2005

10. Gregory Conti and Kulsoom Abdullah. Passive Visual Fingerprinting of Network Attack Tools. In *Proceedings of the 2004 ACM Workshop on Visualization and Data Mining for Computer Security*, VizSEC/DMSEC '04, pages 45–54, New York, NY, USA, 2004. ACM

11. Kulsoom Abdullah, Cherita Corbett, and John Copeland. Lucent wavelan throughput testing. Technical report, March 2002