Sheharyar Khalid

sheharyar@virginia.edu | linkedin.com/sheharyar-khalid1/ | sheharyarkhalid.com

Education

University of Virginia (UVA) — Charlottesville, VA, USA
Ph.D. in Computer Science — *Aug. 2023 – Present*

Lahore University of Management Sciences (LUMS) — Lahore, Pakistan
B.S. in Computer Science — *Aug. 2019 – June 2023*

Research Projects

Bloom Filter Based Streaming Graph Matching

University of Virginia (UVA) — *Charlottesville, VA, USA*

- Led a project with Prof. Kevin Skadron and Prof. Wajih Ul Hassan focused on accelerating threat graph matching workloads on streaming log data, significantly enhancing the efficiency and effectiveness of APT-detection systems capable of scaling to line rates of 100Gbps
- Developed a novel bloom-filter-based tracking algorithm that instantaneously builds and matches attack graphs from log data, enabling real-time sub-second graph matching for dynamic threat detection
- Engineered a stateless tracking engine that utilizes less than 10% of computational resources and drastically reduces graph inference times to less than 1ms and attacker dwell time in systems by 99%
- Architected a highly scalable and efficient system for managing high-volume data streams, supporting real-time dynamic graph updates and achieving an 8x performance improvement over existing systems
- In submission to the USENIX Security Symposium 2025

Hardware-Accelerated Streaming Log Data Matching

University of Virginia (UVA) — *Charlottesville, VA, USA*

- Worked with Prof. Kevin Skadron and Prof. Wajih Ul Hassan to employ FPGA to accelerate Security information and event management (SIEM) systems workloads by pushing computation to an FPGA chip and reducing the database overhead while improving the detection latency
- Conducted comprehensive analysis of SIEM systems to identify and assess theoretical bottlenecks. Designed and implemented a new system architecture, achieving a 400% increase in efficiency by optimizing data processing workflows
- Built a robust automation framework designed to optimize security operations by translating 2,800 existing threat rules into 16,000 actionable literals. This innovation drastically improved the precision and data processing, thereby enhancing system-wide threat detection and response
- Engineered a robust, scalable end-to-end system capable of processing millions of logs daily, leveraging advanced data analytics and literal engines to enhance online threat detection and achieve a 200% improvement in threat identification speed.
- **This work received Best Poster Award at UVA CS Research Symposium** *2023*
- In-submission to USENIX Security Symposium *2025*

Digital Image Recognition on Real World Data

University of Virginia (UVA) — *Charlottesville, VA, USA*

- Developed a classification model on real-world handwriting data using deep learning models, specifically Siamese Networks and Convolutional Neural Networks (CNNs) improving detection accuracy by 9%
- Strategically analyzed and curated a diverse dataset of 100 real-world handwritten samples from diverse participants, building a data pool
- Addressed complex challenges in automated handwriting recognition, adapting and refining models to handle real-world variability in handwriting styles, resulting in 90% accuracy

Privacy Preserving and Transparent Health Application

University of Virginia (UVA) — *Charlottesville, VA, USA*

- Developed a privacy-preserving mobile health application that empowers users with enhanced control over their data sharing practices, significantly boosting user satisfaction and trust
- Conducted detailed user interviews to identify privacy and data sharing concerns within health applications, leading to the design of an innovative app that enhances user autonomy by enabling informed decisions on data sharing and privacy

- Engineered a health application that significantly boosted user satisfaction by 50% and enhanced privacy awareness by 90%, demonstrating tangible benefits in user engagement and data security education.

Provenance Graph Builder

Lahore University of Management Sciences (LUMS) *Lahore, PK*

- Developed a real-time provenance graph from streaming log data using Wazuh, enabling dynamic threat hunting and immediate security incident response
- Engineered a distributed data collection and visualization framework utilizing the lightweight Wazuh agent, deployed across host machines to enhance real-time monitoring and analysis on a central server
- Enhanced graph-building speeds by 30% in a distributed environment, optimizing system performance and efficiency

User Perceptions of Privacy Policies

Lahore University of Management Sciences (LUMS) *Lahore, PK*

- Developed an analysis framework for evaluating the privacy policies of health applications, ensuring compliance with the Health Insurance Portability and Accountability Act (HIPAA) and aligning user expectations with the actual functionalities of these applications
- Designed and implemented a comprehensive framework to analyze 260 privacy policies, creating a structured user study design that accurately captures user perceptions and reactions in real-world scenarios
- Conducted a qualitative study to assess user perceptions of privacy policies, leading to the development of a framework that enhances policy transparency and user-friendliness

Urdu Speech Recognition

Lahore University of Management Sciences (LUMS) *Lahore, PK*

- Developed an Urdu Speech Recognizer leveraging the advanced Wav2Vec2 model, utilizing real-world datasets to create a practical, real-life speech processing application tailored for native Urdu speakers
- Collected 3,500 real-world speech recordings through a coordinated effort involving a team of five participants, effectively building a robust data pool for speech recognition development
- Trained and rigorously tested an automatic speech recognition model, achieving a notable 80% accuracy and maintaining a 20% word error rate, demonstrating significant effectiveness in real-world applications

Other Projects

Early Stage Diabetes Detection | *Jupyter, Scikit Learn, Pandas*

- Developed a logistic regression model for early-stage diabetes detection, achieving a 30% improvement over the baseline, highlighting the model's enhanced predictive capabilities
- Uncovered critical insights through the model, demonstrating how age combined with gender affects diabetes risk, and achieved a high prediction accuracy of 90%

Crime Prediction | *Jupyter, Scikit Learn, Pandas*

- Developed a KModes clustering model to categorize Boston area groups by the most prevalent crime, enhancing strategic law enforcement responses and community safety initiatives
- Conducted detailed data analysis and feature extraction, successfully identifying five distinct crime clusters across various Boston districts, facilitating targeted crime prevention strategies

Discussion and Social Platform | *Python, Node, React, MongoDB, Express, Figma*

- Developed a comprehensive MERN stack web application tailored for LUMS University students, featuring integrated functionalities such as a marketplace, events calendar, forums, and user authentication, significantly enhancing campus engagement and communication
- Implemented a secure sign-up and login system utilizing token-based authentication to ensure high levels of data security and user privacy
- Architected and deployed the application's backend using Node.js/Express, MongoDB, and React, employing tools like Postman, Git, and Trello for streamlining the development process

Cargo Delivery Application | *Python, Node, React, MongoDB, Express, Figma*

- Engineered a robust MERN stack web application for cargo pickup and shipment, incorporating multiple user roles including admin, user, and delivery personnel, enhancing operational efficiency
- Implemented secure JWT authentication to manage multiple user sessions simultaneously
- Developed a full-stack solution using Node.js/Express, MongoDB, and React. Leveraged tools like Postman, Git, and Trello for effective project management, optimizing both development workflow and team productivity.

Work Experience

External Reviewer

University of Virginia (UVA) *Charlottesville, VA, USA*

- IEEE Symposium on Security and Privacy 2024
- Network and Distributed System Security (NDSS) Symposium 2025

Head Teaching Assistant - Network Security *CS 473*

Lahore University of Management Sciences (LUMS) *Lahore, PK*

- Managed and facilitated all aspects of course delivery, including labs, homework, assignments, and quizzes, ensuring a high standard of educational support for students
- Developed and implemented a communication platform to streamline interactions between instructors and students, significantly enhancing the clarity and efficiency of course communications
- Proactively explored and implemented workflow optimization strategies for other teaching assistants, improving efficiency and reducing administrative overhead

Teaching Assistant - Computational Problem Solving *CS 100*

Lahore University of Management Sciences (LUMS) *Lahore, PK*

- Redesigned the course curriculum, introducing innovative assignments and labs that challenged traditional formats and enhanced learning outcomes
- Conducted thorough research and implemented strategic improvements to enhance the student experience in online educational settings, focusing on engagement and interactivity

Technical Skills

Languages: Python, C, C++, SQL, HTML, Haskell, Java, Javascript, MATLAB
Frameworks: Node, React, Numpy, Scikit Learn
Development Stacks: ELK, MERN, Pytorch
Tools & Misc: Git, Docker

Awards & Honors

UVA Computer Science PhD Fellowship: Awarded for the first year of my PhD program at UVA
Best Poster Award - UVA Research Symposium 2023: Best first year poster award
Leadership Representative for Computer Science & Engineering Graduate Student Group (CSGSG)
Dean's Honor List: Awarded for excellent academic performance in Fall 2021-2022