Sonjoy Kumar Paul

skpaul@tamu.edu | LinkedIn | sonjoykp.github.io | GitHub | +1 (979) 402-9790

Career Summary

- Conducting advanced research in **Neuro Security and Brainwave Privacy**, focusing on privacy attacks and defenses on consumer BCI devices at SPIES Lab, advised by Dr. Nitesh Saxena.
- 5 years of professional experience as a C++ software engineer on a **2M+ LOC portfolio and credit analysis system** (Investortools), used by Fortune 500 banks.
- Solid background in data structures, algorithms, systems design, software lifecycle, and agile development.

Research Interests

Neuro-Security; Brainwave Privacy; Computer & Network Security; Machine Learning for Security

Education

Texas A&M University, College Station, TX — Jan 2024 – Present
Ph.D., Computer Science (SPIES Research Lab, Advisor: Dr. Nitesh Saxena)

University of Michigan-Dearborn, MI — Jan 2023 – Dec 2023
Ph.D., Computer and Information Science (Transfer Out)

Bangladesh University of Engineering and Technology (BUET), Dhaka — Feb 2013 – Sep 2017
B.Sc., Computer Science and Engineering

Undergraduate Thesis

Solving Sudoku as a Constraint Satisfaction Problem and Analyzing Effects of Constraint Ordering Explored NP-complete Sudoku as CSP; studied constraint-ordering strategies and their effect on solution efficiency (time & backtrack count). *Supervisor: Abu Wasif, Assistant Professor, BUET.*

Work Experience

Graduate Assistant-Research, Texas A&M University — Jan 2024 – Present

- Designing experiments for **EEG-based speech/digit inference (BrainPhone)** and **security warning comprehension detection** via multimodal signals (eye/mouse tracking).
- Identified **three novel QR-/phone-based attacks** on chat device linking; proposed defenses.

Graduate Student Research Assistant, Univ. of Michigan-Dearborn — Jan 2023 – Dec 2023

- Analyzed GitHub commit histories of serverless apps, detecting inefficiency patterns.
- Improved GPU utilization for compute-heavy serverless workloads (DL & data processing).

Senior Software Engineer, CodeCrafters Intl. (Investortools) — Jul 2022 – Dec 2022

- Enhanced a multi-format **template engine** (Word/Excel/Email) with robust expression parsing.
- Designed and led development of an internal **Auction Management System** for training.

Software Engineer, CodeCrafters Intl. — Jul 2019 – Jun 2022

- Improved **Expression Parser** performance (core engine of the platform).
- Built a **User Folder Migration** system enabling 200+ clients to migrate reports/graphs seamlessly.
- Contributed features across reporting, CLI automation, and proprietary OODBMS.

Software Developer, CodeCrafters Intl. — Jan 2018 – Jun 2019

- Developed secure **report folder permission system** with granular access control.
- Delivered GUI modules using MFC with multi-process and background functionality.
- Diagnosed/resolved critical bugs, improving system stability.

Publications

Peer-Reviewed

- M. A. Munny, M. Alam, **S. K. Paul**, D. Timko, M. L. Rahman, N. Saxena. *Infrastructure Patterns in Toll Scam Domains.* APWG eCrime 2025 (to appear).

Under Review

- A. Mandal, C. Arisoy, **S. K. Paul**, and N. Saxena. *BrainPhone: A Human-Centered Investigation of Speech and Identity Snooping via BCI Headsets*
- M. Alam, A. Hussain, **S. K. Paul**, A. W. Hays, M. I. Huq, N. Saxena. *SoK: AI-FLARE: AI Fuzzing via LLM Reasoning.*
- M. Alam, M. L. Rahman, **S. K. Paul**, A. W. Hays, A. Hussain, M. I. Huq, N. Saxena. *SoK: PHILTER: AI-based Phishing Detection Gaps via LLM Reasoning.*

Selected Projects

TANGO: Low-Resource NLP Augmentation

- Conducted **7.8M-sentence augmentation** across 65 languages; validated with human studies.
- Proposed **prompt-based LLM evaluation** (semantic preservation, error severity, diversity).

Machine Learning-Based Malware Detection and Attack Challenge

- Developed malware detection models under strict constraints (≤1 GB RAM, ≤5s/query).
- Created evasive malware binaries in both black-box and white-box attack settings.

Improving CLIP Training (Vision–Language)

- Benchmarked optimizers (Adam, AdamW, SGD) with advanced contrastive losses (CLIP, CyCLIP, SogCLR, VICReg, OnlineCLR).
- Achieved best performance with **AdamW + CyCLIP** on CC3M & MSCOCO.

Deep Learning Coursework (CSCE 636)

- Ridge vs. LASSO on E2006-tfidf (CV, error curves, sparsity analysis).
- Implemented SGD with LR schedulers (step, cosine, polynomial) for logistic regression.
- Multi-class Logistic Regression on MNIST: minibatch, weight decay, overfitting analysis.

NLP Coursework (CSCE 638)

- Implemented BPE tokenization and sinusoidal positional encodings.
- Built CNN classifier with GloVe embeddings; fine-tuned BERT-base.
- Developed Seq2Seq LSTM for Quora paraphrasing with GloVe init + sampling generation.

Software Security Labs

- Exploited buffer overflow + ROP, format-string, fuzz testing, symbolic execution.
- Implemented defenses: bound checking, control-flow integrity, software fault isolation.

Systems & Applications

- Hadoop + Spark cluster deployment for distributed apps.
- Microservices refactor of a vending app using Docker.

Other Academic Projects

- AI-powered Reversi with alpha-beta pruning (Java Swing).
- SAP-style 4-bit CPU (28 instructions) in Proteus.
- Hall Management System (Java Swing + Laravel/MySQL).

- Multiplayer 29 Card Game (Java sockets, multithreading).

Technical Skills

Languages: C, C++, Java, Python, JavaScript
ML/DL: PyTorch, scikit-learn, TensorFlow, LibAUC
Data/DB: PostgreSQL, MySQL, Oracle, Cassandra, proprietary OODB
Cloud/Big Data: AWS, Hadoop, Spark
Frameworks/Tools: Spring Boot, Docker, Kubernetes, Serverless, Git, Linux

Licenses & Certifications

Neural Networks and Deep Learning, Coursera (Credential ID: L6CHNWH3FQAC), Aug 2020

Awards

2nd Runner-up, Software Project Show, Int'l Conf. on Networking Systems and Security (NSysS), 2016. Recognized for Hall Management System project (efficiency and usability).