




Shoumik Saha

 shoumiksaha.github.io

 smksaha@umd.edu

 in/shoumik-saha

 /ShoumikSaha

RESEARCH INTEREST

- **Machine Learning** - Security & Reliability of GenAI, LLM Alignment & Hallucination
- **Computer Security** - Adversarial Attacks & Defenses, Malware

EXPERIENCE

Applied Scientist Intern <i>Amazon AWS</i> Manager - Zijian Wang	Full Time <i>Jun 2024 – Aug 2024</i>
Graduate Research Assistant <i>Department of Computer Science</i> University of Maryland Supervisor - Dr. Soheil Feizi	Full Time <i>Aug 2023 – Present</i>
Graduate Research Assistant <i>Maryland Cybersecurity Center</i> Supervisor - Dr. Tudor Dumitras	Full Time <i>Aug 2022 – Jul 2022</i>
Lecturer <i>Department of Computer Science and Engineering</i> United International University	Full Time <i>Jul 2021 – Jul 2022</i>
Research Assistant <i>Data Science and Engineering Research Lab</i> Bangladesh University of Engineering and Technology (BUET)	Part Time <i>Mar 2021 – Jul 2022</i>

EDUCATION

University of Maryland <i>Ph.D. in Computer Science and Engineering (Ongoing)</i> ○ CGPA – 3.814/4.0 ○ Advisor – Dr. Soheil Feizi	<i>Aug 2022 – Present</i>
Bangladesh University of Engineering and Technology (BUET) <i>BSc. in Computer Science and Engineering</i> ○ Advanced GPA – 3.79/4.0 ○ Thesis supervisor – Dr. Atif Hasan Rahman	<i>Mar 2016 – Feb 2021</i>

PUBLICATIONS

- **LLM-Check: Investigating Detection of Hallucinations in Large Language Models** [Gaurang Sriramanan, Siddhant Bharti, Vinu Sankar Sadasivan, **Shoumik Saha**, Priyatham Kattakinda, Soheil Feizi]
NeurIPS 2024 (Conference on Neural Information Processing Systems)
We introduced efficient techniques that analyze internal states, attention maps, and output probabilities to detect hallucinations from a single response, significantly improving detection performance while being less computationally expensive than previous methods.
- **Fast Adversarial Attacks on Language Models In One GPU Minute** [Vinu Sankar Sadasivan, **Shoumik Saha**, Gaurang Sriramanan, Priyatham Kattakinda, Atoosa Chegini, Soheil Feizi]
ICML 2024 (International Conference on Machine Learning) (arxiv) (The Register)
We proposed a novel approach in adversarial attack on LLMs, namely BEAST, that can jailbreak,

cause hallucination, and membership inference attacks. Our approach can find jailbreaking prompts within one minute under a resource-constrained setting.

- **DRSM: De-Randomized Smoothing on Malware Classifier Providing Certified Robustness** [Shoumik Saha, Wenxiao Wang, Yigitcan Kaya, Soheil Feizi, Tudor Dumitras]
ICLR 2024 (*International Conference on Learning Representations*) ([arxiv](#)) ([OpenReview](#))

We are the first to propose certified robustness in the domain of static malware detection from executables. We demonstrated both theoretical and empirical robustness of our proposed DRSM framework. Besides, we published a new benign dataset, named PACE.

- **MAlign: Explainable Static Raw-byte Based Malware Family Classification using Sequence Alignment** [Shoumik Saha, Sadia Afroz, Atif Rahman]
Computers & Security Journal ([Link](#))

We proposed a novel approach, namely MAlign, incorporating concepts from Bioinformatics into Malware Security. We developed a static raw-byte-based malware family classifier with better accuracy and robustness. MAlign also provides explainability by relocating the code blocks responsible for malicious attacks.

- **Demystifying Behavior-Based Malware Detection at Endpoints** [Yigitcan Kaya, Yizheng Chen, Shoumik Saha, Fabio Pierazzi, Lorenzo Cavallaro, David Wagner, Tudor Dumitras]
Under Review ([Arxiv](#))

We presented a quantitative study of how sandbox traces differ from real-world ones, and how it impacts machine learning models. We identified this distribution shift and proposed a solution for ML models that boosted the TPR from 14% to 20% @ 1% FPR.

- **Contrastive Self-Supervised Learning Based Approach for Patient Similarity: A Case Study on Atrial Fibrillation Detection from PPG Signal** [Shoumik Saha, Subangkar Shanto, Atif Rahman, Mohammad Masud, Mohammed Eunus Ali]
Under Review ([Arxiv](#))

We proposed Contrastive Learning based Deep Learning solution to find the patient similarity using physiological signal. We demonstrated the efficacy of our method by applying it to detect Atrial Fibrillation using PPG signals collected from smartwatches.

SCHOLARSHIPS & ACHIEVEMENTS

- Awarded 'Dean's Fellowship' from the University of Maryland
- Awarded 'Innovation Fund' for research from the ICT division of Bangladesh government
- Achieved 'Merit Stipends' from BUET in five out of seven terms
- Got 'Dean's Award' in Junior year for extra-ordinary result
- Achieved 'Talent-Pool Scholarship' in High School

TEACHING

Teaching Assistant

University of Maryland

Fall 2023

Data 200: Knowledge in Society: Science, Data and Ethics

Lecturer

United International University

Fall 2021 - Summer 2022

Discrete Mathematics, Data Structure and Algorithm, Operating Systems

ACTIVITIES

- Reviewer, ICML 2024 & ICLR 2025
- External Reviewer, USENIX 2023
- Participant, ICPC Dhaka Regional
- General Secretary, BUET Photographic Society