Gargi Mitra web: https://gargi-mitra.github.io/website/ gargimiitm@gmail.com

Research Statement

I investigate the security and privacy of *modern networked systems*, focusing on those that integrate their legacy system components with emerging technologies. Such integrations are increasingly common– embedding Machine Learning (ML)-based software into legacy medical devices, integrating cloud-based services with industrial automation systems and IoT applications, or incorporating predictive streaming techniques into services like Netflix. My research reveals that *such integrations create new vulnerabilities that traditional security and privacy measures were not designed to address.* These vulnerabilities arise from *fundamental mismatches between the legacy design assumptions, and the security, privacy, performance, and interoperability requirements of modern technologies*, and they often invalidate legacy threat models and trust boundaries. In safety-critical environments, such vulnerabilities can directly endanger human lives or disrupt essential services; in domains like IoT and web applications, they can lead to serious privacy breaches and regulatory violations.

Yet, these modern technologies are meant to benefit society in substantial ways. AI/ML-driven medical diagnosis and treatment enable quick, low-cost access to life-saving treatments, remote monitoring supports continuous patient care, automation reduces the physical burden on workers, and cloud-based analytics allow organizations to scale while lowering operational costs. *My long-term vision is to ensure that this technological progress remains secure, safe, and trustworthy, enabling society to reap its benefits without any hidden risks.* The overarching goal of my research, therefore, is to determine whether systems retain their security, safety, and privacy guarantees when modern technologies are integrated, and if not, to design practical, minimally disruptive strategies to restore them. This agenda is guided by three foundational questions:

1. How can we define and thoroughly evaluate the "security" and "privacy" of a system?
2. How does the integration of new technologies expand the attack surface of a system and introduce new threat models that existing security and privacy risk assessment techniques do not capture?
3. When are legacy security and privacy measures sufficient, and when must new strategies be devised to meet the security and privacy requirements of modern technologies?

I follow a systems-oriented approach to address these questions. This involves analyzing end-to-end interactions between a system's legacy and modern components to uncover emergent vulnerabilities, through modeling the data and control flows of the modern system, updating its threat models, designing and empirically evaluating the impact of novel attack techniques, and experimentally demonstrating their feasibility. Finally, I produce actionable insights that security practitioners can adopt to mitigate the newly identified vulnerabilities with minimal disruptions. Next, I describe how applying this approach to answer the three questions has shaped my contributions across multiple system domains.

1 Doctoral Research: Encrypted Web Traffic Analysis

Driven by growing concerns about network surveillance, in my early research works, I examined encrypted traffic analysis (ETA)-based attacks, a class of attacks that infer a user's online activities only from their encrypted traffic metadata (packet size, direction, and inter-packet timings) without breaking encryption or compromising the communication endpoints. This enables long-term, stealthy surveillance by on-path adversaries positioned anywhere along the network path (e.g., a compromised ISP or backbone router) between a web server and its users.

The central theme that emerged from my PhD research is that *information leakage from encrypted traffic metadata cannot be attributed solely to weaknesses in encryption techniques. The design of web applications and network protocols plays a critical role as well. Specifically,*

optimizations and newly introduced features in the applications and protocols might introduce privacy risks that remain unnoticed without rigorous threat re-modeling and risk assessment.

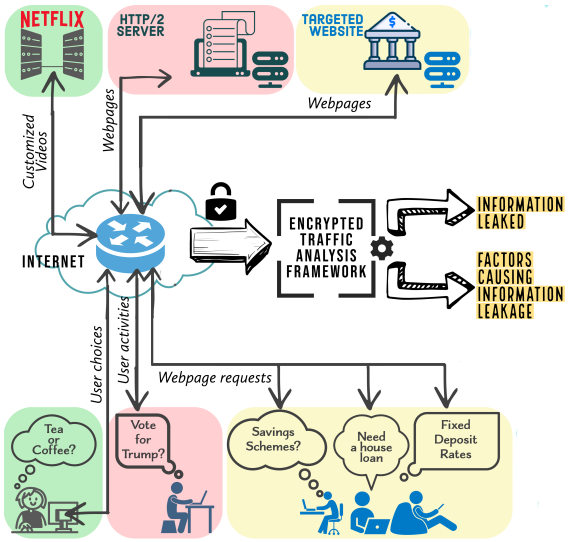

Fig. 1. Contributions of my PhD thesis

My work in this space highlights a known but underexplored facet of privacy: For an individual, leaked browsing data may often seem benign. But when *webpage-level* browsing patterns are aggregated *across a population*, an adversary can infer demographic trends–product preferences, socio-political leanings, content consumption habits–which they can then exploit for targeted campaigns or selective news exposure, as seen in the Cambridge Analytica scandal.

The earliest ETA techniques attempted webpage-level mass-scale surveillance of HTTPS traffic, but they were built when the web ecosystem was much simpler–sites had far fewer webpages, no embedded cookies, and the techniques assumed highly constrained user behavior (e.g., restricted navigation paths, or a single OS/browser/caching configuration). The techniques used ML models trained and tested on traffic collected under these constrained conditions. I found that models trained under such constrained assumptions fail when tested on modern websites with thousands of pages, embedded cookies, and more relaxed user-behavior assumptions that capture mass-scale diversity (varied navigation paths, heterogeneous client platforms, and differences in caching and cookie configurations). However, expanding training data to cover this diversity required collecting an impractically large number of traffic samples and risked triggering server-side DoS defenses. To overcome these limitations, we developed **Snoopy** [1] (IEEE TDSC'22), an ETA framework that uses statistical techniques to approximate how each webpage's encrypted-traffic fingerprint varies across cache and cookie settings, browser/OS combinations, and navigation paths. We showed that learning these variations requires far fewer traffic samples than adapting prior techniques to account for user diversity, making Snoopy practical and highly accurate.

In **Depending on HTTP/2 for Privacy?** [2] (DSN'20), we debunked the belief that HTTP/2's parallel request processing inherently protects encrypted traffic from ETA attacks. We showed that this misconception stemmed from an outdated threat model. Prior ETA work assumed purely passive adversaries, overlooking those who can introduce small perturbations in network traffic without any additional privilege, and without getting detected. We demonstrated that such minimally active adversaries can disrupt HTTP/2's parallelism and make the traffic vulnerable to known ETA attacks.

In **White Mirror** [3] (ACM SIGCOMM'19 Posters and Demos), we developed ETA techniques to demonstrate that Netflix's predictive streaming technique for interactive videos leaks users' on-screen choices through encrypted traffic metadata. This raised significant privacy concerns, especially as Netflix planned to commercialize its interactive video creation software across other industries, potentially amplifying the privacy risks across a wider range of applications. We disclosed our findings to Netflix; they initially acknowledged the issue but considered network-layer vulnerabilities outside their scope. Our work gained wider attention after coverage in Wired, which in turn, sparked discussions on Reddit. About a year later, we found that Netflix had adopted one of our recommended countermeasures (changing the user-choice encoding scheme). We validated all our ETA-based research findings with extensive experiments on real-world websites and network traffic collected from consenting participants.

2 Postdoctoral Research: Security of Safety-Critical Systems

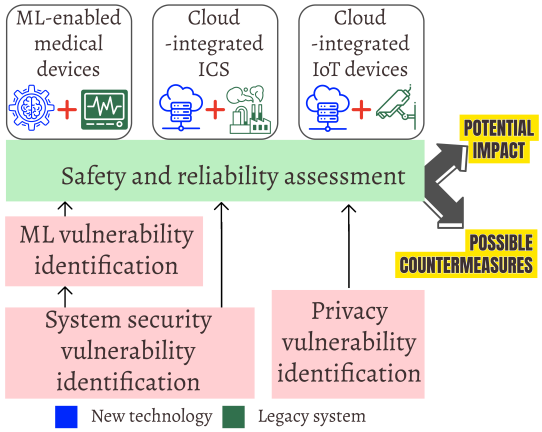

Fig. 2. Postdoctoral research contributions

During my postdoctoral research at UBC, I built on the insights gained during my PhD while broadening my research scope in two directions. First, I *extended ETA techniques on domains beyond web applications*, for example, in modern (cloud-integrated) industrial control systems (ICS). Second, I examined how the *integration of modern technologies with legacy systems introduces new security and privacy risks* beyond ETA, that require revisiting risk assessment techniques and designing new security measures. I focused on two representative domains undergoing rapid technological modernization: medical devices that incorporate machine learning (ML) for personalizing treatment or enhancing diagnosis accuracy, and IoT applications that continuously process data whose privacy sensitivity changes in real time.

In our recent work [4], we show that modern ICSes, now increasingly offloading their operational workflows to the cloud, are vulnerable to ETA-based attacks despite using VPNs and standard security practices to protect site-cloud communications. We demonstrate this through **ICS-Sniper**, a stealthy and targeted blackhole attack that applies a novel ETA technique on the VPN traffic metadata to identify packets likely carrying safety-critical information, and then selectively drops them to disrupt ICS operations. The key insight is that *modern ICSes are vulnerable to a fundamentally new threat model*: the adversary has less system visibility than those targeting web applications, and has lesser privileges than the intrusive adversaries assumed in existing ICS attacks. Yet, the adversary can endanger physical safety and operational reliability, as well as lead to huge financial losses. Existing defenses against ETA-based attacks in web applications are unsuitable for ICS, as they impose timing overheads that violate the low-latency requirements of ICS communications. This work exposes a previously overlooked, high-impact threat to cloud-integrated ICSes.

Similarly, we identified new high-impact attack vectors in modern ML-enabled medical devices. These devices rely on large volumes of physiological data for generating personalized diagnosis and treatment plans, both during model training and at inference time. This data is collected from peripheral devices (e.g., wearable devices and monitoring equipment). This creates a highly interconnected, multi-vendor ecosystem with an expanded attack surface. Existing risk-assessment techniques fail to account for vulnerabilities arising from peripheral components that feed data into ML pipelines at different stages of training and inference. To address this gap, we conducted a **systematic assessment** [5] (IEEE/ACM CHASE'24, Springer Nature HealthSec'24 Journal) of commercial ML-enabled devices and showed that adversaries can exploit known security vulnerabilities in peripheral components to launch stealthy inference-time false data injection attacks on the ML model, which can potentially endanger patient safety. Experts with the necessary interdisciplinary knowledge to identify such vulnerabilities are rare, and assembling such teams is costly for manufacturers. To address this practical problem, we developed **MedAIScout** [6] (Red Teaming Workshop, NEURIPS'24) and **SAM** [7] (Workshop on Cybersecurity in Healthcare, ACM CCS'24), two LLM-aided tools that partially automate this vulnerability-discovery process, reducing the security analyst's time and effort. In our most recent work [8], we showed that conventional anomaly detectors (ADs) perform poorly in securing medical ML models against evasion attacks. These ADs are typically trained indiscriminately on large volumes of benign data to cover diverse scenarios. However, in the medical domain, patient physiological features vary widely, even among those with the same condition,

causing ADs to misclassify adversarial inputs as benign patient data, which leads to low recall.

Today, general-purpose IoT devices are increasingly deployed in public environments, such as surveillance cameras in offices or patient-monitoring cameras in hospitals and care facilities. These devices may capture data from individuals who are entirely unknown to the developer at design time, and then transmit this data to backend servers at unknown locations. This combination creates a substantial risk of violating privacy regulations, particularly in regions with stringent frameworks such as the GDPR. To help developers detect data flows that may inadvertently breach privacy guidelines due to the unpredictable nature of both the captured data and the backend infrastructure, we developed **Turnstile** (Eurosys'26), a hybrid information flow control framework. Turnstile identifies privacy-sensitive code paths through static taint analysis and augments the application with a dynamic information flow tracking mechanism. This approach protects privacy-sensitive data flows while saving the developer's time and effort.

Overall, my Postdoctoral research marked a deliberate shift toward safety-critical systems, where the consequences of security and privacy breaches might directly compromise user safety, degrade system reliability, and trigger harmful physical outcomes. A central insight emerging from this research is that existing security measures often do not cover the expanded attack surface when modern technologies are retrofitted into legacy systems. This underscores the need for security research that involves cross-layer, domain-aware understanding of modern safety-critical systems. Effective defense in these environments requires not just patching vulnerabilities or hardening networks, but a fundamental rethinking of threat models, validating system behavior, and designing domain-specific, safety-aligned security mechanisms.

3 Future Research Agenda

My research goal is to ensure that technological advancement remains safe, secure, and trustworthy, enabling society to embrace its benefits without unintended consequences. Building on the insights from my prior work, I plan to pursue the following research directions that would collectively advance my contributions towards this overarching goal.

1. **Cost–benefit analysis of security research in the industry:** During my work, I had the opportunity to interact closely with security practitioners from industry (e.g., GE Healthcare, Medcrypt, Forescout, CS2AI) and regulatory agencies such as the FDA. These interactions revealed that many high-impact security and privacy vulnerabilities persist because security investments are frequently shaped by short-term revenue goals and competitive pressures. Integrating emerging technologies with legacy system components further expands the attack surface, and mitigating them would be expensive, owing to the substantial time and resources they require, which can potentially delay product release or reduce market competitiveness. Consequently, organizations struggle to determine which security issues to prioritize and how much financial investment is justified. To address this challenge, I aim to develop a decision-support framework that would help industry security teams prioritize vulnerabilities and determine the most effective way to mitigate them among all available options. This framework will incorporate economic impact, operational burden, robustness of the mitigation techniques, and safety implications, enabling decision-makers to systematically evaluate trade-offs across various mitigation options and allocate resources in a way that yields the greatest combined benefit for both end users and manufacturers. I plan to leverage my existing network of industrial security practitioners to gather real-world insights into operational and economic constraints and to foster long-term collaborations around this research effort.
2. **Security for AI-agent-enabled systems:** As AI agents become more deeply embedded in everyday applications, adversaries have started to exploit vulnerabilities in cross-agent and agent-system interactions. For example, researchers recently demonstrated that Google Gemini can be hijacked via a malicious calendar invite, tricking it into executing smart-home commands by embedding hidden prompts in an innocuous email invitation [9]. The

recent exploitation of Claude [10] for a large-scale cyber-espionage campaign is another such example. The key insight from both of these incidents is the same: the AI agents did not truly understand the broader context of the data or control flows among connected components. Instead, they acted on fragmented instructions that appeared harmless in isolation, but when combined, resulted in unsafe or malicious behavior. These incidents underscore the need for a systematic investigation into the attack surfaces that emerge when AI agents interact with legacy software and operational infrastructure. To this end, I plan to develop methods for modeling system-wide cross-agent and agent-system interactions, identifying potential contextual gaps, determining how adversaries can exploit these gaps to trigger unsafe actions, characterizing these exploitable vulnerabilities, and designing effective countermeasures.

3. **AI for continuous security risk assessment in safety-critical systems:** My research shows that attack surface of safety-critical systems continually evolves as new technologies are integrated. Yet, current safety and security standards and industrial practices that treat security as an afterthought to cut expenses rarely account for this ongoing adaptation. This creates security gaps that adversaries often exploit to cause damage. To address this gap, I aim to develop methodologies for continuous threat assessment in evolving safety-critical domains. For this, I want to build an AI-agent-enabled automated risk assessment toolkit that can be personalized across various domains. This toolkit will be able to (i) translate domain-specific safety constraints into security policies; (ii) dynamically model system attack surfaces every time a new technology is integrated with it; and (iii) interface with vulnerability databases to perform near–real-time assessments of newly reported vulnerabilities and their impact on the system. Together, these contributions will enable continuous, evidence-driven security evaluation in modern safety-critical systems. The use of agentic AI will keep implementation costs low while incorporating domain-specific security expertise.

References

[1] G. Mitra, P. K. Vairam, S. Saha, N. Chandrachoodan, and V. Kamakoti, "Snoopy: a webpage fingerprinting framework with finite query model for mass-surveillance," *IEEE Transactions on Dependable and Secure Computing*, vol. 20, no. 5, pp. 3734–3752, 2022.

[2] G. Mitra, P. K. Vairam, P. Slpsk, N. Chandrachoodan, et al., "Depending on HTTP/2 for privacy? Good luck!" In *2020 50th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*, IEEE, 2020, pp. 278–285.

[3] G. Mitra, P. K. Vairam, P. Slpsk, N. Chandrachoodan, and K. V, "White mirror: Leaking sensitive information from interactive Netflix movies using encrypted traffic analysis," in *Proceedings of the ACM SIGCOMM 2019 Conference Posters and Demos*, 2019, pp. 122–124.

[4] G. Mitra, P. Dash, Y. E. Yao, A. Mehta, and K. Pattabiraman, "ICS-Sniper: A Targeted Blackhole Attack on Encrypted ICS Traffic," *arXiv preprint arXiv:2312.06140*, 2023.

[5] M. Elnawawy, M. Hallajiyan, G. Mitra, S. Iqbal, and K. Pattabiraman, "Systematically Assessing the Security Risks of AI/ML-enabled Connected Healthcare Systems," in *2024 IEEE/ACM Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)*, 2024, pp. 97–108. DOI: 10.1109/CHASE60773.2024.00019.

[6] A. P. Dharmalingam and G. Mitra, "MedAIScout: Automated Retrieval of Known Machine Learning Vulnerabilities in Medical Applications," in *Red Teaming GenAI: What Can We Learn from Adversaries?*.

[7] M. Hallajiyan, A. P. Dharmalingam, G. Mitra, H. Alemzadeh, S. Iqbal, and K. Pattabiraman, "SAM: Foreseeing Inference-Time False Data Injection Attacks on ML-enabled Medical Devices," in *Proceedings of the 2024 Workshop on Cybersecurity in Healthcare*, 2023, pp. 77–84.

[8] M. Elnawawy, G. Mitra, S. Iqbal, and K. Pattabiraman, "Learning from the Good Ones: Risk Profiling-Based Defenses Against Evasion Attacks on DNNs," *arXiv preprint arXiv:2505.06477*, 2025.

[9] B. Nassi, S. Cohen, and O. Yair, "Invitation Is All You Need! Promptware Attacks Against LLM-Powered Assistants in Production Are Practical and Dangerous," *arXiv preprint arXiv:2508.12175*, 2025.

[10] Anthropic, *Disrupting the first reported AI-orchestrated cyber espionage campaign*, URL: https://www.anthropic.com/news/disrupting-AI-espionage, Last accessed: Nov 27, 2025.