Securing AI for Science

Innovating Reliable and Secure Supercomputing Infrastructure

Realizing resilient and secure supercomputing systems to advance open-science research, in the presence of accidental failures and cyber-attacks is my research goal. The primary motivations are: i) high rates of AI training/inference disruptions in hyperscalers [1, 2], with OpenAI reports GPUs "*melting*" during extreme ChatGPT loads [3], ii) cross-stack hardware failures, software bugs, and interconnect congestions [4] disrupting exascale scientific simulations, iii) and increasing billion-scale cyberthreats targeted at open-science infrastructure, e.g., NOIRLab security incident when operating the Vera Rubin Observatory [5].

Leading cybersecurity research at the National Center for Supercomputing Applications (NCSA), a hub for the nation's scientific computing at the University of Illinois Urbana-Champaign, I witness firsthand the challenges of ensuring the uninterrupted operation of $100M-scale supercomputers (BlueWaters) and modern research computing (DeltaAI) in accumulated uncertainty from underlying hardware to AI applications.

Funding. I have raised more than $1.5*M* in research funding from the National Science Foundation (NSF) as PI, industry partners (IBM Research), and published a dozen papers in top conferences such as Supercomputing, USENIX Security, IFIP International Conference on Dependable Systems and Networks, and IEEE Quantum Computing and Engineering [6–32].

My research portfolio includes four NSF grants [15, 19], e.g., Cybersecurity Innovation for Cyberinfrastructure (CICI: Security Data Lake), Research on Research Security (RoRS: Uncover HPC allocation abuses), and Formal Method in the Fields (FMiTF: Formalizing Federated Authentication); HPC-Quantum Computing observability (`PQSee.com`) is being commercialized through initial support of Campus Cyberinfrastructure (CC*: Post Quantum Cryptography Network Instrument). As a fellow of the NSF Cybersecurity Center of Excellence, he provides resiliency expertise to industry partners (IBM Research), mid-scale national testbeds (FABRIC), and DOE national labs.

Intellectual Merit. The key driver of my research is cross-stack measurements of high-speed interconnects, AI application logs, and kernel probes from sustained petascale supercomputers (Blue Waters), HPC/GPU clusters (DeltaAI, Polaris), and the exascale Aurora machine at the Argonne National Lab. These decade-long, curated datasets (1PB) enable cyberattack detection, error propagation modeling across GPU architectures (Hopper/Blackwell/Rubin) and vendors, and an understanding of quantum-resistant cryptography adoption, as published in top conferences such as Supercomputing, USENIX Security, and IEEE Quantum Computing and Engineering.

Broader Impact. My research has contributed to the development of a reliable and secure AI infrastructure, ensuring the integrity of extreme-scale scientific workflows across the NSF's advanced research computing, DOE supercomputers, and industry-funded data lakes. My research has been recognized with awards for my teaching and mentorship, including an IEEE Dependable Systems and Networks Best Paper Award (2014), Outstanding Mentor for CyberCorps: Scholarship for Service and Fiddler Innovation Fellowship recipients (2025), and an Art of HPC exhibit at Supercomputing 2024.

Future of Resilient and Secure AI/HPC/Supercomputing infrastructure. I am investigating the reliability and security challenges of next-generation accelerators, such as Quantum Processing Units (QPU), being integrated with HPC.

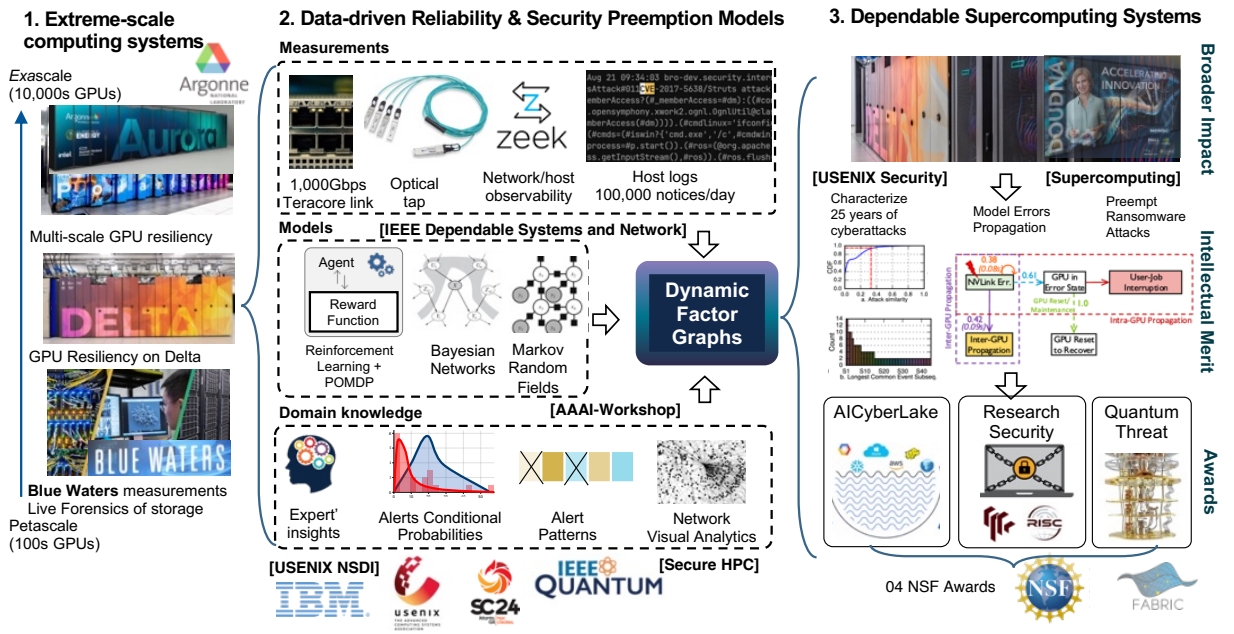

Figure 1: Real-world data of latest GPUs (Vera Rubin planned) from supercomputers (DeltaAI, Aurora) are collected from national labs (ANL, NERSC) to drive security modeling, reliability taxonomy, and standardization with NIST.

Foundation: Reliability and Security Data Curation from Real-World Systems

The key driver for my research is to curate decade-long forensic data on nascent Secure Shell (SSH) backdoors [25, 26] (2000s), FBI Major Case Stakkato [33] security incidents (2010s), threats leveraging Machine Learning/AI to masquerade attacks as regular network traffic [20] (2020s), and to anticipate Quantum-driven adversaries that might break traditional encryptions [16] (2030 onward). In the process of collecting security data, I also capture reliability data, e.g., one Petabyte of BlueWaters performance measurements is available on Globus for open-science research. My efforts have been recognized with several NSF grants that I am leading as PI, an award for Best Paper at IEEE DSN, and publications in top system conferences (Supercomputing, USENIX Security/NSDI, IEEE Quantum Computing Engineering). Artistic explorations also provide a visual analytic view into the curated data, with my Art of HPC exhibited at the Denver Museum of Art / showcased at the Georgia International Convention Center.

Method: Probabilistic Modeling

The intellectual merit is to provide measurements of real-world, system-wide outages that validate traditional control-theoretic modeling. The crux of my method is to augment the utility of probabilistic graphical models (Bayesian Networks, Markov Random Fields, and Factor Graphs, which subsume both) with raw measurements while safeguarding critical attack-preemption/failure-detection decisions with domain insights/knowledge. The high fidelity of attacks' evolution and errors propagation across the hardware stack to application layers via high-speed interconnects is therefore captured, for the first time, in state-of-the-art NVIDIA Hopper GPU accelerators in the DeltaAI system. My method is applicable not only in the resiliency domain, but also in medical diagnostics, evidenced by publications and funding with leading U.S. clinics (Mayo) and international hospitals in Vietnam.

Impact: Realiable and Secure Supercomputers

The broader impact of my work has been disseminated mainly through the NSF TrustedCI, Cybersecurity Center of Excellence. As a TrustedCI fellow, I provide input on frameworks that advise on cybersecurity implementation for NSF Major Facilities, Mid-scale Research Infrastructure (FABRIC), and SLAC, which traditionally lack a full security operations team relying on OmniSOC. Completed engagements include

assessing quantum risk to science at the National Center for Atmospheric Research (NCAR), giving security log analysis tutorials at Lawrence Berkeley National Laboratory/CMU, and driving adoption of GPU resiliency tools for Aurora, one of the three exascale supercomputers, at Argonne Leadership Computing Facility through the Joint Laboratory for Extreme Scale Computing (JLESC).

Future: Reliable HPC-QC integration and Safeguarding Open-Science Infrastructure

The convergence of novel accelerators such as Quantum Processing Units (QPUs), HPC/exascale computing (Rubin/MI450/GPUMax), and the ongoing shifts in the distribution of AI training and inference workloads in the next decade will pose a unique set of challenges.

(C1) *How to measure the reliability of HPC-enabled Quantum and Spaceflight computing systems?*

(C2) *How to safeguard the integrity of HPC allocations and data security of open science?*

(C3) *How to keep fundamental security research open and collaborative?*

Hereafter, I will outline technical and management approaches to address the challenges outlined above.

To address challenge C1, Fig. 1 shows an example pipeline and approach to scale up our existing resiliency model from petascale (BlueWaters, Delta) to exascale (Aurora, 2026-onward) by using both a federation of AI agents to assist with unstructured log parsing and workflow coordination. In addition, I will use counterfactual reasoning to simulate "what-if" scenarios of catastrophic failures and irreversible attack impacts so that both monitoring tools and operators can anticipate and be prepared for them.

A crucial direction I will address is the correctness of scientific computing in the presence of new accelerators (Quantum Processing Units) and hostile environments (RISC-V architecture in the presence of space radiation). Counterfactual reasoning, coupled with formal reasoning, will be the enabling technique for NASA/JPL's Spaceflight computing systems being developed with DARPA and will roll out before 2040s. I will further adapt those techniques to the workloads of the next-generation HPC-Quantum Computing integration that NCSA has begun developing with PsiQuantum and IQM.

The principle of the above resiliency model provides the basis for addressing challenge C2, because monitoring the health of supercomputers will reveal the efficacy of HPC allocations, identify potential misuse, and increase the utilization of supercomputers for productive scientific work. I will translate the intellectual merits above into research security policies, educational materials exemplifying unproductive use cases of HPC allocations, and work with interagency partners through NSF's Research on Research Security Program (RoRS) to disseminate the impact (#2537355).

Finally, to assist with high-fidelity data storage, I already secured funding from the NSF Cybersecurity Innovation for Cyberinfrastructure program to build a data lake (#2530738) that is applicable to both resiliency and security. I am involved in defining open security standards for research data sharing and collaborating with the National Institute of Standards and Technology on a special publication on HPC Security (800-223). We will draw concrete case studies from emerging attacks such as quantum-resistant cryptography, AI backdoors, etc., to demonstrate the artifact on the Open Science Data Federation (OSDF).

References Cited

[1] X. Jiao, A. Pandey, K. Pattabiraman, and F. Lin, "Large-scale AI infra reliability: Challenges, strategies, and Llama 3 training experience," in *2025 55th Annual IEEE/IFIP International Conference on Dependable Systems and Networks-Supplemental Volume (DSN-S)*, pp. 140–146, IEEE, 2025. https://doi.org/10.1109/DSN-S65789.2025.00054.

[2] D. Pariag and A. Phillabaum, "Building resilient monitoring at Meta," 2023. https://atscaleconference.com/building-resilient-monitoring-at-meta/; accessed 2025-09-01.

[3] C. Welch, "OpenAI says 'our GPUs are melting' as it limits ChatGPT image generation requests," *The Verge*, March 2025. https://www.theverge.com/news/637542/chatgpt-says-our-gpus-are-melting-as-it-puts-limit-on-image-generation-requests.

[4] S. Jha, A. Patke, J. Brandt, A. Gentile, B. Lim, M. Showerman, G. Bauer, L. Kaplan, Z. Kalbarczyk, W. Kramer, *et al.*, "Measuring congestion in {High-Performance} datacenter interconnects," in *17th USENIX Symposium on Networked Systems Design and Implementation (NSDI 20)*, pp. 37–57, 2020.

[5] info@noirlab.edu, "Cybersecurity incident at nsf noirlab | noirlab." https://noirlab.edu/public/announcements/ann23022/. [Online; accessed 2025-09-26].

[6] B. Baheri, E. Giusto, S. Xu, K. N. Smith, E. Younis, and P. Cao, "Grand challenges of secure & trustworthy quantum computing," 2025.

[7] E. Giusto, S. Núñez-Corrales, K. N. Smith, P. Cao, E. Younis, P. Rech, F. Vella, B. Baheri, A. Cilardo, B. Montrucchio, *et al.*, "Dependable classical-quantum computing systems engineering," *Frontiers in Computer Science*, vol. 7, p. 1520903, 2025.

[8] S. Cui, A. Patke, H. Nguyen, A. Ranjan, Z. Chen, P. Cao, B. Bode, G. Bauer, C. Di Martino, S. Jha, C. Narayanaswami, D. Sow, Z. T. Kalbarczyk, and R. K. Iyer, "Characterizing GPU resilience and impact on AI/HPC systems," *arXiv preprint arXiv:2503.11901*, 2025. https://arxiv.org/abs/2503.11901.

[9] S. Cui, A. Patke, H. Nguyen, A. Ranjan, Z. Chen, P. Cao, B. Bode, G. Bauer, C. Di Martino, S. Jha, *et al.*, "Story of two gpus: Characterizing the resilience of hopper h100 and ampere a100 gpus," *Supercomputing, ACM/IEEE International Conference for High Performance Computing, Networking, Storage, and Analysis*, 2025.

[10] R. Gupta, S. Liu, R. Zhang, X. Hu, X. Wang, H. Benkraouda, P. Kommaraju, P. Cao, N. Feamster, and K. Nahrstedt, "Generative active adaptation for drifting and imbalanced network intrusion detection," *arXiv preprint arXiv:2503.03022*, 2025.

[11] E. G. Campolongo, Y.-T. Chou, E. Govorkova, W. Bhimji, W.-L. Chao, C. Harris, S.-C. Hsu, H. Lapp, M. S. Neubauer, J. Namayanja, *et al.*, "Building machine learning challenges for anomaly detection in science," *arXiv preprint arXiv:2503.02112*, 2025.

[12] L. Phan, A. Gatti, Z. Han, N. Li, J. Hu, H. Zhang, C. B. C. Zhang, M. Shaaban, J. Ling, S. Shi, *et al.*, "Humanity's last exam," *arXiv preprint arXiv:2501.14249*, 2025.

[13] P. Cao, "Jupyter notebook attacks taxonomy: Ransomware, data exfiltration, and security misconfiguration," in *SC24-W: Workshops of the International Conference for High Performance Computing, Networking, Storage and Analysis*, pp. 750–754, IEEE, 2024.

[14] P. Cao, Z. Kalbarczyk, and R. K. Iyer, "Security testbed for preempting attacks against supercomputing infrastructure," in *SC24-W: Workshops of the International Conference for High Performance Computing, Networking, Storage and Analysis*, pp. 1781–1788, IEEE, 2024.

[15] P. M. Cao, "Cc* integration-small: Quantum-resistant cryptography in supercomputing scientific applications," *NSF Award Number 2430244. Directorate for Computer and Information Science and Engineering*, vol. 24, no. 2430244, p. 30244, 2024.

[16] J. Sowa, B. Hoang, A. Yeluru, S. Qie, A. Nikolich, R. Iyer, and P. Cao, "Post-quantum cryptography (PQC) network instrument: Measuring PQC adoption rates and identifying migration pathways," in *2024 IEEE International Conference on Quantum Computing and Engineering (QCE)*, vol. 1, pp. 1835–1846, IEEE, 2024. https://doi.org/10.1109/QCE60285.2024.00213.

[17] L. Yang, Z. Chen, C. Wang, Z. Zhang, S. Booma, P. Cao, C. Adam, A. Withers, Z. Kalbarczyk, R. K. Iyer, and G. Wang, "True attacks, attack attempts, or benign triggers? An empirical measurement of network alerts in a security operations center," in *33rd USENIX Security Symposium (USENIX Security 24)*, pp. 1525–1542, 2024. https://www.usenix.org/conference/usenixsecurity24/presentation/yang-limin.

[18] V. Tay, X. Li, D. Mashima, B. Ng, P. Cao, Z. Kalbarczyk, and R. K. Iyer, "Taxonomy of fingerprinting techniques for evaluation of smart grid honeypot realism," in *2023 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*, pp. 1–7, IEEE, 2023.

[19] P. M. Cao, "Fmitf: Track ii: Bringing verification-aware languages and federated authentication to enable secure computing for scientific communities," *NSF Award Number 2319190. Directorate for Computer and Information Science and Engineering*, vol. 23, no. 2319190, p. 19190, 2023.

[20] K. Chung, P. Cao, Z. T. Kalbarczyk, and R. K. Iyer, "stealthml: Data-driven malware for stealthy data exfiltration," in *2023 IEEE International Conference on Cyber Security and Resilience (CSR)*, pp. 16–21, IEEE, 2023.

[21] P. Lougovski, O. D. Parekh, J. Broz, M. Byrd, J. C. Chapman, Y. Chembo, W. A. de Jong, E. Figueroa, T. S. Humble, J. Larson, *et al.*, "Report for the ascr workshop on basic research needs in quantum computing and networking," tech. rep., Sandia National Lab.(SNL-NM), Albuquerque, NM (United States), 2024.

[22] Y. Wu, P. Cao, A. Withers, Z. T. Kalbarczyk, and R. K. Iyer, "Poster: Mining threat intelligence from billion-scale ssh brute-force attacks," *Proc. Netw. Distrib. Syst. Security*, pp. 1–3, 2020.

[23] Y. Cao, P. Cao, H. Chen, *et al.*, "Predicting ICU admissions for hospitalized COVID-19 patients with a factor graph-based model," in *2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 2577–2582, IEEE, 2021.

[24] J. Basney, P. Cao, and T. Fleury, "Investigating root causes of authentication failures using a SAML and OIDC observatory," in *2020 IEEE 6th International Conference on Dependability in Sensor, Cloud and Big Data Systems and Application (DependSys)*, pp. 119–126, IEEE, 2020. https://doi.org/10.1109/DependSys51298.2020.00026.

[25] P. M. Cao, Y. Wu, S. S. Banerjee, J. Azoff, A. Withers, Z. T. Kalbarczyk, and R. K. Iyer, "CAUDIT: Continuous auditing of SSH servers to mitigate brute-force attacks," in *16th USENIX Symposium*

on Networked Systems Design and Implementation (NSDI 19), pp. 667–682, 2019. https://www.usenix.org/conference/nsdi19/presentation/cao.

[26] P. Cao, "On preempting advanced persistent threats using probabilistic graphical models," *arXiv preprint arXiv:1903.08826*, 2019.

[27] S. Chen, M. McCutchen, P. Cao, S. Qadeer, and R. K. Iyer, "Svauth–a single-sign-on integration solution with runtime verification," in *Runtime Verification: 17th International Conference, RV 2017, Seattle, WA, USA, September 13-16, 2017, Proceedings 17*, pp. 349–358, Springer, 2017.

[28] P. Cao, E. C. Badger, Z. T. Kalbarczyk, R. K. Iyer, A. Withers, and A. J. Slagell, "Towards an unified security testbed and security analytics framework," in *Proceedings of the 2015 Symposium and Bootcamp on the Science of Security*, pp. 1–2, 2015.

[29] P. Cao, E. Badger, Z. Kalbarczyk, R. Iyer, and A. Slagell, "Preemptive intrusion detection: Theoretical framework and real-world measurements," in *Proceedings of the 2015 Symposium and Bootcamp on the Science of Security*, pp. 1–12, 2015.

[30] C. Pham, Z. Estrada, P. Cao, Z. Kalbarczyk, and R. K. Iyer, "Reliability and security monitoring of virtual machines using hardware architectural invariants," in *2014 44th Annual IEEE/IFIP International Conference on Dependable Systems and Networks*, pp. 13–24, IEEE, 2014.

[31] P. Cao, H. Li, K. Nahrstedt, Z. Kalbarczyk, R. Iyer, and A. J. Slagell, "Personalized password guessing: a new security threat," in *Proceedings of the 2014 Symposium and Bootcamp on the Science of Security*, pp. 1–2, 2014.

[32] C. Pham, P. Cao, Z. Kalbarczyk, and R. K. Iyer, "Toward a high availability cloud: Techniques and challenges," in *IEEE/IFIP International Conference on Dependable Systems and Networks Workshops (DSN 2012)*, pp. 1–6, IEEE, 2012.

[33] K. Ricker, J. Barlow, and C. Adams, "Fbi major case 216 https://www.ideals.illinois.edu/items/105968," 2008.