Toward Effective Security/Reliability Situational Awareness via Concurrent Security-or-Fault Analytics (SoFA)

Mingyan Li (lim3@ornl.gov), Robert A. Bridges (bridgesra@ornl.gov), Pablo Moriano (moriano@ornl.gov), Christian Engelmann (englmannc@ornl.gov), Feiyi Wang (fwang2Wornl.gov), Ryan Adamson (adamsonrm@ornl.gov)

Oak Ridge National Laboratory, Oak Ridge, TN, USA

Topic: Adversarial modeling & simulation. Integrity & provenance. Secure data architecture. Graph algorithms.

Challenge: Modern critical infrastructures (CI) and scientific computing ecosystems (SCE) are complex and vulnerable [1]. The complexity of CI/SCE, such as the distributed workload found across ASCR scientific computing facilities, does not allow for easy differentiation between emerging cyber security and reliability threats. It is also not easy to correctly identify the misbehaving systems. Sometimes, system failures are just caused by unintentional user misbehavior or actual hardware/software reliability issues, but it may take some significant amount of time and effort to develop that understanding through root-cause analysis. On the security front, CI/SCE are vital assets. They are prime targets of, and are vulnerable to, malicious cyber-attacks. Within DoE, inter-disciplinary and cross-facility collaboration (e.g., ORNL INTERSECT initiative, next-gen supercomputing OLCF6), traditional perimeter-based defense and demarcation line between malicious cyber-attacks and non-malicious system faults are blurring. Amidst realistic reliability and security threats, the ability to effectively distinguish between non-malicious faults and malicious attacks is critical not only in root cause identification but also in countermeasures generation.

Today's segregated fault diagnosis and attack detection approaches leave much room for improvement. Anomalies during attack detection could very well point to system faults (thus the problematic high false-positives with root-cause identification failure). Similarly, fault diagnosis anomalies could be from attacks (dangerous false negatives). Realistic situational awareness requires a wholistic analytics approach, counting in both security and faults contexts, and concurrently, to achieve effective situational awareness and realistic root-cause identification.

Opportunity: Concurrent & parallel security/fault analytics improves attack/failure root-cause analysis quality. There are two implications – proactive/adaptive analytics and countermeasure generation. There are no guarantees that when anomalies were first detected, sufficient evidence has already been collected and analyzed, particularly in the real-time scenarios where attacks/faults are progressing. Analyzing, and adaptively collecting additional relevant data pertinent to scenario developments, on both malicious and non-malicious fronts, helps to properly exonerates/confirms suspected attack/fault developments. Examples of additional data include increased-fidelity telemetry streams, directed endpoint detection and response (EDR) queries, or even information normally kept by third party organizations relevant to a suspicious CI/SCE workflow. With accurate root-cause identification, effective countermeasures (repair or defend) can then be devised and applied. Such an improved fault/attack paradigm could serve to improve and protect DoE CI/SCE where detection and defense of a compromised workflow component at one facility could prevent the spread of a malicious payload to another. One potential approach is to model system's behavior with (temporal) graphs and using the interactions between its different states to uncover complex modes of operation, including faults and attacks [2]. We can further employ model-driven checkpoints based on scientific and data workflow to proactively monitor, collect and detect anomalies, in both data privacy and federated learning contexts.

Figure-1 conceptually illustrates concurrent attack/fault analytics. On the left is an example fault tree [3], or it could be a state graph with nodes represent behavioral states and edges represent transitions between states. Observed events are mapped into perspective nodes (colored dots). Scenarios are analyzed horizontally across, denoted by the directed curve. On the right, a conceptual adversarial model (an attack strategy goal/subgoal model here – could also be ML-based models) [4], also with horizontal scenario analytics curves. Here, concurrent analytics takes place. If called for, in real-time environment, the detection system can adaptively adjust auditing behavior to proactively look-ahead to monitor/search for relevant data per suspicious scenario developments (e.g., dual arrows on the right). Otherwise, in offline mode, retroactively retrieve required data locally or externally for assessment. Such concurrent attack/fault analytics improves root-cause attribution effectiveness thus leads to improved situational awareness and countermeasure generation [5].

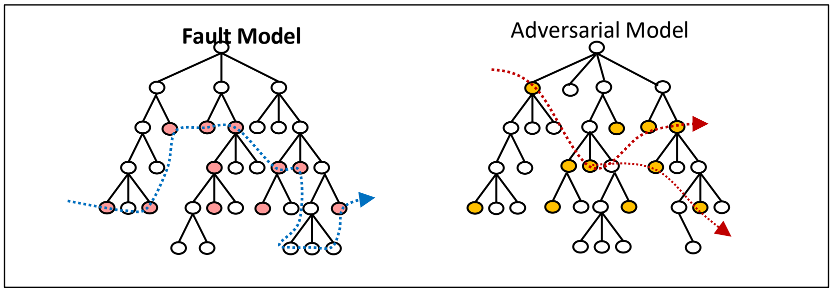

Figure-1 Concurrent fault & attack analytics conceptual illustration

Proactive adaptive auditing, i.e., dynamic data ingestion, further helps to alleviate the “information impoverished” problem in big data collection. Adaptively audit/collect only contextual relevant data, per scenario development, reduces the “needle in the haystack” scalability challenge in real-life applications.

Timeliness/Maturity: Both failure/fault diagnostic and security analytics (e.g., Host-based/network-based Intrusion Detection Systems, and ML-based anomaly detection systems) technologies have been investigated in the past, but two stove pipes did not really come together to realistically address the real-life situational awareness challenge. Past attempts have been made to build/utilize declarative security attack as well as fault models, but they were severely limited by scalability in terms of declarative knowledge capturability. Recently, work on streaming telemetry and log events has matured within DoE supercomputing facilities, and platforms exist to enable dynamic software-defined data collection. With today’s machine learning based or graph-based data science approaches, there lies the potential of realization of such integrated multi-modal approach.

References:

[1] National Strategic Computing Reserve, https://www.whitehouse.gov/wp-content/uploads/2021/10/National-Strategic-Computing-Reserve-Blueprint-Oct2021.pdf

[2] B. Bowman, C. Laprade, Y.Ji, and H. Huang. "Detecting Lateral Movement in Enterprise Computer Networks with Unsupervised Graph {AI}." In 23rd International Symposium on Research in Attacks, Intrusions and Defenses ({RAID} 2020), pp. 257-268. 2020.

[3] Z. Gao, C. Cecati, and S. X. Ding, “A Survey of Fault Diagnosis and Fault-Tolerant Techniques—Part I: Fault diagnosis with model-based and signal-based approaches && Part II: Fault Diagnosis with Knowledge-Based and Hybrid/Active Approaches”, IEEE Trans. Ind. Electron., vol. 62, no. 6, Jun 2015

[4] Y. Deldjoo, T. D, NOIA, and F. A. Merra. “A survey on Adversarial Recommender Systems: from Attack/Defense strategies to Generative Adversarial Networks”, ACM Comput. Surv., Vol. 54, No. 2, Mar 2021

[5] G. Tertytchny, N. Nicolaou, and M. Michael “Classifying network abnormalities into faults and attacks in IoT-based cyber physical systems using machine learning”, Microprocessors and Microsystems, Sept 2020.