CICI:UCSS: Blockchain Based Assured Open Scientific Data Sharing and Governance

Murat Kantarcioglu PI, muratk@vt.edu

VIRGINIA TECH

Objectives:

- Scientific data's integrity need to be protected while sharing data
  - No unauthorized changes to data
  - Track any changes to data (i.e., capture provenance info)
  - Make sure provenance information is captured securely
  - Analyze the provenance data for potential attacks and issues

Approach:

Past work:

- Develop hybrid blockchain solutions for capturing and sharing provenance information
- Integrate blockchain with federated learning (FL) to enable accountable, secure and privacy preserving scientific data sharing. (E.g., see our ACM CODASPY 2021, BlockFla paper)
- Develop ML models to show how the captured provenance data could be leveraged for detecting malicious behavior both in FL and Intrusion Detection settings. (E.g., see our Usenix Sec 2023, Provenance Detector paper)
- Develop attacks against ML models used to detect malicious behavior through the analysis of provenance graphs. (E.g., see our ICLR 2025 GOttack paper)

This year's work:

- Understand the impact of domain constraints in ML based attacks (e.g., what if attacker cannot change certain parts of the graph ?)
- Leveraging agentic AI for provenance graph analysis.

Key Results (last 12 month):

- Developed a new adversarial attack that leverages domain constraints (to appear at ICML 2026 conference)
  - Our results show that existing attacks against graph neural networks do not perform well if the attacker have limits (e.g., certain edges are harder to manipulate in the provenance graph)
  - Our novel attack can work even under domain constraints
- Developed a new agentic framework (ProvSeek) for automatically identifying malicious events from high-level threat information. (arXiv, https://arxiv.org/abs/2508.21323) See below for overview.

SCAN ME

Agentic Workflow of ProvSEEK