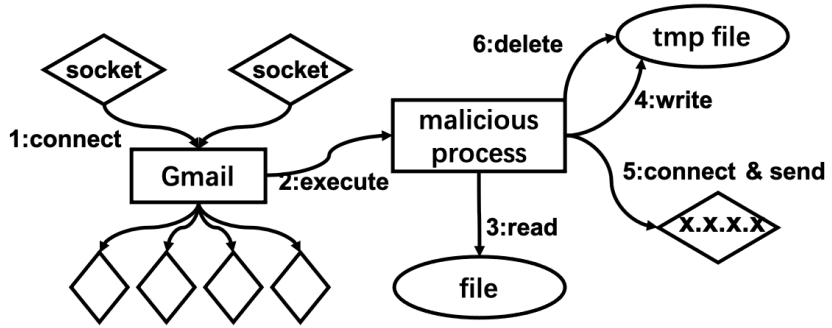
Advanced Persistent Threat Detection via Mining Long-Term Features in Provenance Graphs

Fan Xu, Qinxin Zhao, Xiaoxiao Liu, Nan Wang, Meiqi Gao, Xuezhi Wen, Dalin Zhang

Frontiers of Computer Science, DOI: 10.1007/s11704-024-40610-8

Problems & Ideas

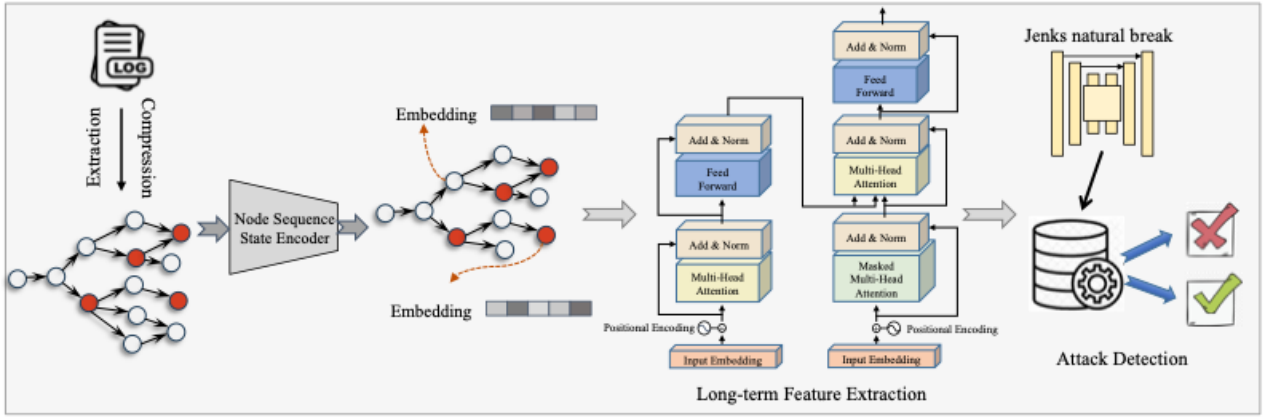
- Problems of conventional stereo matching approaches:
  - With the continuous increase in the size of the generated provenance graph, the actual computational power limits the analysis capability.
  - Due to the considerable duration of APT attacks, existing models can mostly observe information within a small window, rather than focusing on long-term correlated features in the provenance graph.
  - Abnormal attack data are hard to obtain, thus existing methods can hardly generalize to unseen attack type.
  - Ideas: An end-to-end APT attack detection model within provenance graphs that takes both long-term contextual features and fine-grained node embeddings into account.



A simple case of provenance graph.

Main Contributions

- Contributions:
  - A novel and effective provenance graph feature representation method is proposed, which retains contextual information from large-scale provenance graphs and efficiently encodes a large amount of data to obtain information-rich intermediate representations.
  - In order to effectively discover potential associations generated by long-term persistent APT attacks, an end-to-end APT attack detection model based on long-term feature extraction methods is proposed.
  - In situations where obtaining attacking data is difficult, APT attack detection can be realized without any predefined attack features, while remaining high level of detection accuracy and efficiency.



The overview of long-term feature association provenance graph detector.