

Nomination Statement

Provenance analysis based on system auditing logs produces a provenance graph that shows system activities (e.g., processes reading and writing files), which has emerged as a highly effective solution for cyber threat investigation, especially for Advanced Persistent Threat (APT). However, existing approaches generally produce a large provenance graph (containing more than 1 million edges) and makes it difficult for security experts to recover attack steps from the graph. While recent research proposes techniques to filter out irrelevant edges, these techniques still generate a relatively large graph (more than 200 edges) and may incorrectly filter out important edges. To address this important problem and make provenance analysis more practical, this paper presents a much-needed approach, DepComm, that summarizes the produced provenance graph as communities, performs aggressive compression inside communities, and identifies representative information flows as each community's activity summary. That is, without filtering any edge, DepComm summarizes large provenance graphs as a small number of communities (10-20), which is practical for security experts to trace the attack-related activities and identify root causes of the attacks. This is the first paper that tackles this difficult problem from a completely new angle, and kicks off a new direction for this line of research.

Zhuotao Liu