Intro

Company users require access to the internet to do their job. Although, the internet has a vast amount of threats looming. With the size of corporations and how much data is produced, it is impossible to search through every event or log to judge its threat level. Companies emphasize security so being able to identify malware threats is important. Therefore, I am going to analyze how the researchers at HP Labs applied a large-scale graph inference by constructing a host-domain access graph to identify malware threats.

HP labs

Researchers at HP Labs modeled detection as an inference problem by creating very large graphs with event logs companies already record. They seeded the graph with minimal truth; a small fraction of domains is labeled as malicious and benign and the rest are unknown. Then a belief propagation is adapted and if a domain is over a threshold it is marked as malicious. Belief propagation was used because it is fast and approximate that scales to larger graphs.

Inference Problem

Host-domain access graph

Malware communication structure

Belief Propagation