IoT Device Anomaly Detection using Provenance Graphs

Ebelechukwu Nwafor
Howard University
P.O. Box 1212
Washington, D.C 43017-6221
ebelechukwu.nwafor@bison.howard.edu

Gedare Bloom
Howard University
P.O. Box 1212
Washington, D.C 43017-6221
gbloom@howard.edu

ABSTRACT

Over the years, the Internet of Things (IoT) has revolutionized the way we interact with devices. From smart grids, to healthcare and home automation This paradigm shift inadvertently allows for ease of device access. Unfortunately, this technological advancement has been met with unforeseen security challenges. One major security challenge is malicious intrusions. A common way of detecting a malicious attack is by treating an attack as an anomaly and using anomaly detection techniques to pinpoint the source of an intrusion. In a given IoT device, provenance graphs which denotes causality between system events offers immense benefit for device anomaly detection. Provenance provides a comprehensive history of activities performed on a system which indirectly ensures trust. Given a provenance graph, how can we determine if anomalous activities exists in a memory constrained IoT device? This paper seeks to address this issue. In this paper, we present a lightweight approach to providing anomaly detection of IoT devices using provenance graphs. We introduce an error tolerant graph embedding technique using local frequencies of nodes and edges in which provenance graphs are converted into a vector space representation.This vector space representation can further be used as an input parameter for clustering or classification algorithms. We compare our method to two known anomaly detection algorithms.

KEYWORDS

IoT Security, Data Provenance, Anomaly Detection

ACM Reference format:
Ebelechukwu Nwafor and Gedare Bloom. 2016. IoT Device Anomaly Detection using Provenance Graphs. In *Proceedings of ACM Conference, Washington, DC, USA, July 2017 (Conference'17),* 6 pages.
DOI: 10.475/1234

1 INTRODUCTION

Over the years, there has been a surge in the level of data breach on computing devices around the world.

An anomaly, also referred to as an outlier, is defined as data that deviates from the normal system behavior. Anomaly detection has applications in various domains such as intrusion detection, fraud detection, medical health devices, sensor fault detection, web spam. For instance, in fraud detection, an anomalous event could indicate the presence of a suspicious credit card transaction which might cause the card to be temporarily disabled until proper identification can be determined. Additionally, in health devices such as a pace maker or a glucometer, an anomalous event could be detrimental which might lead to catastrophic events.

Due to the nature of safety critical systems, anomaly detection has been an area of active research. This is also motivated by the unprecedented level of data breaches seen all over the world. Detecting current and future malicious attacks is of utmost importance to the security of data in IoT devices. Intrusion detection, an application of anomaly detection offers a method of detecting malicious attacks based on system events. Malicious intrusions could have disastrous financial consequences to an organization. For example, a data breach on a consumer website could lead to the theft of sensitive financial and personal information. IDS systems such as snort, have been deployed in numerous consumer and enterprise systems all around the world. Most of these systems provide real-time detection of intrusions that might occur in a host system or a network.

Anomaly detection looks for patterns that deviates from the normal system behavior. This enables the detection of known and unknown malicious attacks. An anomalous system event could indicate a system is being used as a botnet in a distributed denial of service attack (DDOS). It could also indicate a system fault.

The challenge in anomaly detection is providing the right features from a dataset to use in detecting intrusions. Another challenge exists in defining what constitutes as normal system behavior. There often is a thin line between what is considered normal system behavior and what is considered an anomaly. In addition, what is considered normal system behavior is constantly evolving. The issue of generating training or test dataset which classifies anomalous and normal system behavior is a major challenge since not all known anomalous system behavior can be recreated.There are some advantages of using text categorization techniques for detecting anomalies in provenance graphs. Anomaly detection is a binary event (abnormal or normal). This makes using text categorization approach feasible. Since anomaly detection is a binary event (abnormal or normal). This makes using text categorization approach seems feasible and straightforward.

Ensuring data trust in internet of things device is a challenge. How do we provide an effective means of detecting malicious intrusion in a given system? Provenance data provides a history of system events which can be used to detect system faults or anomalous system behaviors. For example, sensors deployed in a oil rig, provenance can be used to detect when there is a device leak. Also, in an IoT enabled home, containing smart devices such as a smart thermostat, smart fridge provenance can be used to detect a point of intrusion in an event of a system hack.

Conference'17, Washington, DC, USA

DOI: 10.475/1234