

Project Description: Behavioral Models and Exploratory Data Analysis

Company: Interset

Supervisors: Shaun Pilkington and Maria Pospelova, Data Science, Interset

The Company

Interset provides a highly intelligent and accurate insider and targeted outsider threat detection solution that uses data science, behavioral analytics, machine learning and big data to protect critical data.

The Science

Our approach is to identify risk within the organization by ingesting activity data from various sources (such as desktops, servers and data repositories) and then detecting abnormal, suspicious and high risk behaviors of user accounts, machines, and intellectual property assets (such as files). We do this by learning normal, baseline patterns from the data (using machine learning and mathematical models), and then quantify abnormal behavioral through probabilistic models that look for differences between observed behavior and baseline.

The Project

We incorporate new data sets by doing exploratory data analysis (EDA) on actual log datasets (provided by our customers) that describe human activity surrounding valuable data repositories to protect, such as source code audit logs, cloud repository activity information, and database query logs.

The dataset(s) that will be analyzed for model design, development and validation include:

- Cloud activity sharing (such as Dropbox)
- Cloud authentication and access logs (such as Amazon Web Services)
- Network traffic (such as proxy logs)

As part of the EDA project, you will gain experience in:

- ingesting, sampling, visualizing and summarizing extremely large amounts of data using tools such as Python, R and RStudio
- applying intuition and domain knowledge to identify, evaluate and engineer features of interest
- experimentally designing, applying and evaluating predictive models (primarily unsupervised machine learning) to identify high risk behaviors and indicators of compromise
- documenting, visualizing and communicating results using tools such as R Markdown

We are excited about the possibilities of using principled and powerful mathematics and machine learning to detect data theft. Join and help us do real life data science, and help make the world a safer place.

The Team

We catch bad guys with math. Our mission is to enable our clients to protect their data using new insights and advanced analytics. We are smart and creative. We are empowered to make a difference.