# Project Description: Bayesian Network-Based Log Generation

Company: Interset
Supervisors: Shaun Pilkington, 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

One major issue facing analytic companies is the difficulty in generating realistic datasets in order to test their system. To be practical the data generation process will have to use randomness, but in order to be realistic this must have a specific structure. One very useful framework for representing such stochastic structures is with Bayesian networks. These allow one to explicitly model the parameters and their dependencies in a comprehensible way. Given such a model it is trivial to generate data from it. There also exist methods to learn the parameters or even the structure of networks from data.

The project is to create a framework (using Bayesian networks) that can represent and generate synthetic log data in a way that can be easily modified.

A prototypical example is log data with a set of users who are active at certain times and perform certain actions against some set of resources.

Given a Bayesian network model of the above, it is easy to add more users, change user parameters, or add in anomalous behavior. There is much software that assists in designing and implementing Bayesian networks, which could be leveraged to improve the result.

# 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.