HIRAM
COLLEGE
INTIMATE LEARNING. GLOBAL REACH.

Genetic Optimization of an Intruder Detection System

By: Stefan Maurer

References:
•Kumara Sastry, David Goldberg, Graham Kendall. "Genetic Algorithms." 2005.

Abstract:

Modeled after the evolutionary process that modifies all organisms over thousands of years, a genetic algorithm is designed to improve an algorithm's effectiveness in a specific environment. For this project, we will be taking advantage of such a genetic algorithm in an attempt to optimize the parameters of an intrusion detection system. Just as an organism's genome decides how an organism behaves, the parameters of an algorithm control how the algorithm behaves in its environment. Starting with a random sample of all possible properties for the system, we can evaluate the effectiveness of each set of parameters in the environment, and then allow the more effective sets to "reproduce". During reproduction, two selected parent parameters, also known as chromosomes, create two children. Each child's chromosomes are a mixture of each parent's chromosomes. After each child is created, there is a chance that a portion of their chromosomes will be mutated. By repeating this process thousands of times, we can model the evolution of organisms, and attempt to find an optimal state for our intrusion detection system.

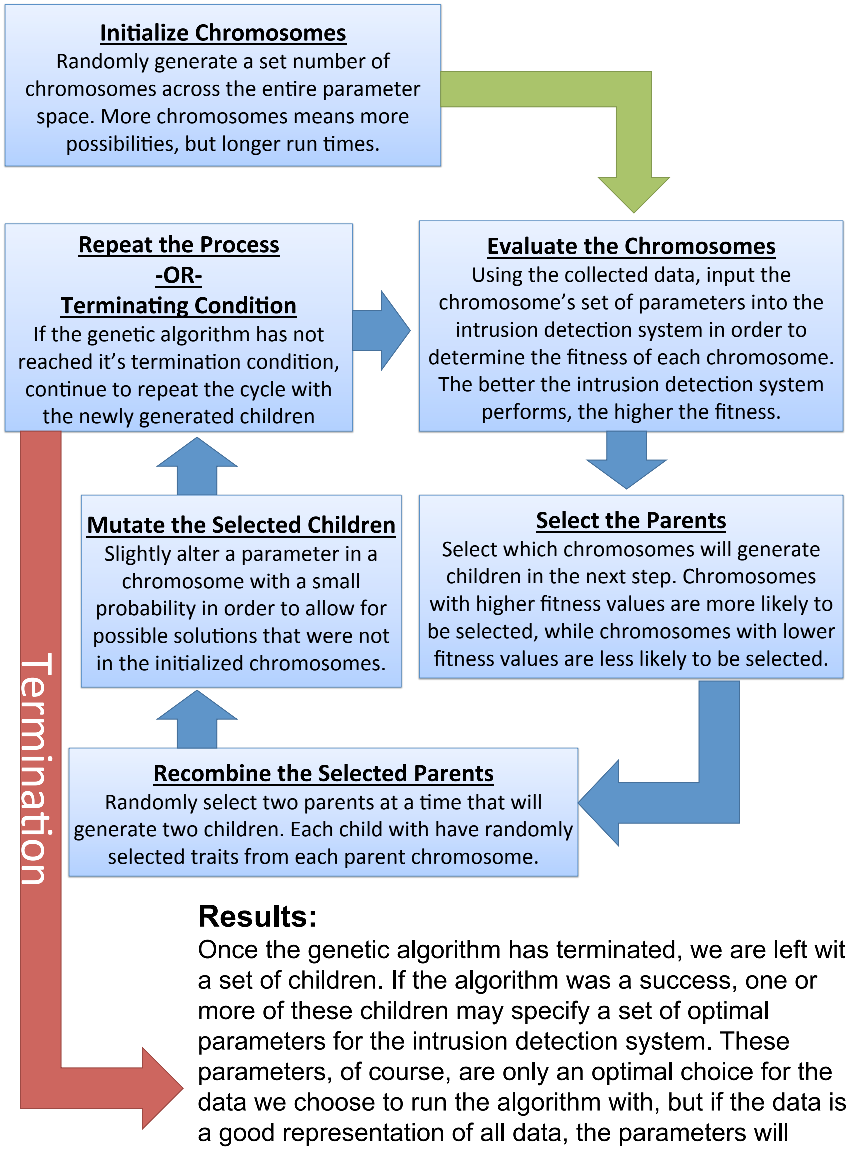
The Intrusion Detection System:

Developed last year for my second IRC in Artificial Intelligence, the intrusion detection system uses a bioinformatics technique, the Smith-Waterman local alignment algorithm, to compare a user's behavior to a user's past behavior. The purpose of the intrusion detection system is to identify when the current user's behavior does not correspond to the user's past behavior. However, due partially to the fact that the intrusion detection system uses an alignment algorithm to differentiate between a user and an intruder, there are several parameters that must be carefully determined. Our goal for this project is to optimize these parameters as to maximize the efficiency of the intrusion detection algorithm.

Gathering Data:

In order to determine the effectiveness of a particular set of parameters, we needed to test a user's past behavior against their own behavior and an intruder's behavior. By testing in this manner, we can attempt to minimize how often the legitimate user is wrongly detected as an intruder, as well as how often an intruder is not flagged as such However, in order to accomplish this, raw data was needed. Using an process recorder, we were able to record accurate behavioral patterns of several volunteers. Each person's data was then compared to all other behavior in order to determine the accuracy of the system.

Process of the Genetic Algorithm



Initialize Chromosomes
Randomly generate a set number of chromosomes across the entire parameter space. More chromosomes means more possibilities, but longer run times.

Evaluate the Chromosomes
Using the collected data, input the chromosome's set of parameters into the intrusion detection system in order to determine the fitness of each chromosome. The better the intrusion detection system performs, the higher the fitness.

Select the Parents
Select which chromosomes will generate children in the next step. Chromosomes with higher fitness values are more likely to be selected, while chromosomes with lower fitness values are less likely to be selected.

Recombine the Selected Parents
Randomly select two parents at a time that will generate two children. Each child with have randomly selected traits from each parent chromosome.

Mutate the Selected Children
Slightly alter a parameter in a chromosome with a small probability in order to allow for possible solutions that were not in the initialized chromosomes.

Repeat the Process
-OR-
Terminating Condition
If the genetic algorithm has not reached it's termination condition, continue to repeat the cycle with the newly generated children

Results:

Once the genetic algorithm has terminated, we are left with a set of children. If the algorithm was a success, one or more of these children may specify a set of optimal parameters for the intrusion detection system. These parameters, of course, are only an optimal choice for the data we choose to run the algorithm with, but if the data is a good representation of all data, the parameters will perform well on data that was not given originally!