

A Case Study in Genetic Algorithms applied to Off-Line Intrusion Detection Systems

Pedro A. Diaz-Gomez	Dean F. Hougen
Robotics, Evolution, Adaptation, and Learning Laboratory (REAL Lab) School of Computer Science, University of Oklahoma Norman, OK, USA pdiazg@ou.edu	Robotics, Evolution, Adaptation, and Learning Laboratory (REAL Lab) School of Computer Science, University of Oklahoma Norman, OK, USA hougen@ou.edu

Off-line intrusion detection systems are computer security mechanisms that search audit trail registries, looking for user activities that match patterns of events known as attacks. Because such search is NP-complete, heuristic methods will need to be employed as databases of events and attacks grow. Genetic Algorithms (GAs) have been widely used as appropriate heuristic search methods. However, balancing the need to detect all possible attacks in an audit trail with the need to avoid warnings of attacks that do not exist is a challenge, given the scalar fitness values required by GAs.

We present a case study of *GASSATA*—a Genetic Algorithm as an Alternative Tool for Security Audit Trail Analysis (Mé, 1998), a previously proposed GA-based IDS that shows this difficulty with respect to its fitness function, and we propose a new method to overcome it.

GASSATA is an off-line intrusion detection system (Mé, 1998) with fitness function $F(I) = \alpha + \sum_{i=1}^{N_a} W_i \cdot I_i - \beta * T^2$, where I is the hypothesis vector, α maintains $F(I) > 0$ in order to retain diversity in the population (using proportional probability selection), N_a is the number of known attacks, W is the weighted vector that reflects the risk of each attack, β provides a slope for the penalty function, and T is the number of times for which $(AE \cdot I)_i > OV_i$, where AE is the attack-event matrix that shows which events are required for each attack, and OV is the observed vector of events.

(Mé, 1998) reports good results with *GASSATA* but our experience has been that the system often generates false positives and negatives (Diaz-Gomez & Hougen, 2005). For this reason, based on experimental results we proposed a new fitness function: $F(I) = N_e - T'$, where N_e corresponds to the number of events, and T' corresponds to the number of times that $(AE \cdot I)_i > OV_i$, for each attack I_i . That is, if a hypothesized attack I_i considered alone, would cause $(AE \cdot I)_i > OV_i$ for some i , and another hypothesized attack I_j considered alone, would also cause $(AE \cdot I)_i > OV_i$, then T' would have a value of 2 (whereas T would have a value of 1). To avoid false negatives, we added a mechanism that takes the union of all newly hypothesized attacks that are consistent with the existing aggregate solution set. With the new fitness function and mechanism suggested there are no false positives and the number of false negatives decreases dramatically compared to the results we saw previously (Diaz-Gomez & Hougen, 2005).

This research shows some difficulties in providing accurate values to parameters in the fitness function suggested in *GASSATA* (Mé, 1998) and proposes a solution independent of variable parameters making the fitness function to solve this particular problem quite general and independent of the audit trail data. Our solution has proved more effective than the original and a variant of the original using tournament selection.

References

Diaz-Gomez, P., and Hougen, D. 2005. Improved off-line intrusion detection using a genetic algorithm. To appear in *Proceedings of the Seventh International Conference on Enterprise Information Systems*.

Mé, L. 1998. GASSATA, a genetic algorithm as an alternative tool for security audit trail analysis. In *First International Workshop on the Recent Advances in Intrusion Detection*.