



THE WHITE ROSE GRID e-Science Centre

Dynamic Policy Learning

Motivation and Objective

Traditional top-down computer security policies can be too rigid to cope with changes in operational environments. Recent work has shown that machine learning techniques can be used to infer security models, where the learning is directed by high level objectives or characterised by previously seen decision examples.

As the operational environment may change, a fixed policy will eventually become sub-optimal and hence a policy needs to be continually updated, especially in a highly dynamic operational environment. In this paper, we show that how Multi Objective Genetic Programming (MOGP), a kind of evolutionary algorithms, can be used to learn dynamic security policies from decision examples.

Time variant Security Policy

Since there is not any dynamic security policy model, we first design a time-varying risk budget based security policy model. This model is used to generate training examples for MOGP and serves as the benchmark against which the models learnt by MOGP are evaluated.

In a system using a risk budget based, each user is given an amount of risk tokens that represents how much risk the system is willing to take with that user. To access an information object, a subject raises a request that is characterized by the profiles of the subject, object, communication channel, environment, and an offer to pay some amount of risk tokens for the

access. Then, the system will compute a quantified risk estimate of the request from the profile according to a hidden risk model. The estimate is expressed in number of risk tokens. The request is granted iff $offer \geq (1+\beta)$ (risk estimate), where β is a safety margin that varies with time and in the range $[0, \beta_{max}]$.

Experiment Settings and Results

This model is used to generate random decision examples, $\langle Request(profiles, offer), decision \rangle$. β changes every 1000 examples. Every 200 examples (with the same β) are grouped into a chunk and fed into the policy inference framework.

The learning target is the policy that best approximates the correct mapping of $\langle Request(profiles, offer), decision \rangle$, i.e., the one with the highest accuracy (the percentage of correct decisions made when fed with a set of requests).

However, there are problem in optimising dynamic problem with evolutionary algorithms. The evolving population will eventually converge. This loss of diversity may result the population trapped in a local optima. Theoretically, mutation is the way out. However, mutation may require too many generations to produce the desired changes.

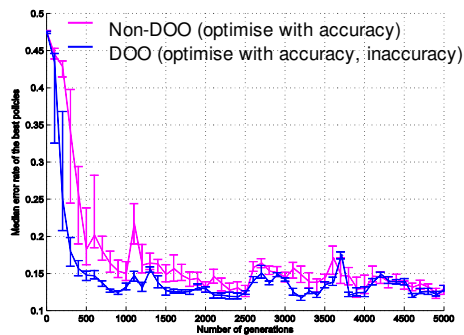
We propose here the Diversity via Opposing Objectives (DOO) framework). DOO changes each objective of the problem in question into a pair of opposing objectives. For

example, in our problem, the accuracy of security policy becomes two objectives: the accuracy and inaccuracy of a security policy. Then, the converted problem is optimised using MOEA. In such a setting, no solution can be dominated by another.

Each individual has a fair chance to survive and pass on its genes. Diversity is preserved and encouraged. 2 pairs of objectives are used here:

- (accuracy, inaccuracy) evaluated against the current chunk of examples.
- (accuracy, inaccuracy) evaluated against all (old) chunks of examples the learning process has seen.

The output policy is the one with the highest bias evaluated against the latest chunk.



The following shows the median error rates of the best policies learnt after every 100 generations of training time along with their 95% confidence intervals. The best policies learnt using DOO generally has a lower median error rates. The heights of the error rate “spikes” are also much lower in DOO.

Conclusions

This paper shows that dynamic policy can be learned from examples using evolutionary algorithms. A novel

dynamic learning framework – diversity via opposite objectives (DOO) is proposed. DOO treats an N objective optimisation problem as a $2N$ objective optimisation problem by adding an opposing objective for each of the original objectives. With such a setting DOO is able to maintain the diversity among the individuals in the population whilst optimising the intended objectives. Diversity among the individuals can aid in avoiding premature convergence and coping with concept drift in dynamic learning.

References

Yow Tzu Lim, Pau-Chen Cheng, Pankaj Rohatgi, John A. Clark, “Dynamic Security Policy Learning”, Proceedings of the first ACM workshop on Information Security Governance 2009, Chicago, Illinois, USA November 13 - 13, 2009

Acknowledgement

Research was sponsored by the US Army Research laboratory and the UK Ministry of Defence and was accomplished under Agreement Number W911NF-06-03-0001. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the US Army Research Laboratory, the U.S. Government, the UK Ministry of Defence or the UK Government.

Further Information

Yow Tzu Lim
yowtzu@cs.york.ac.uk
<http://www.cs.york.ac.uk>