

Autonomic Computer Network Defence using Reinforcement Learning and Risk States

Luc.Beaudoin@drdc-rddc.gc.ca
Defence Research and Development Canada

Computer Network Defence (CND) is concerned with the active protection of information technology (IT) infrastructure against malicious and accidental incidents. Given the growing complexity of IT systems and the speed at which automated attacks can be launched, implementing timely and efficient network incident mitigating actions, whether proactive or reactive, is a great challenge. A human is simply not able to handle the combination of the complexity and speed with which the analysis and response must be performed to limit risk and avoid self-inflicted damages. Therefore, some believe that in order to efficiently protect critical IT networks, CND actions selection and implementation should be automated. We refer to this area of research as *Autonomic CND*.

A significant technical challenge in Autonomic CND is finding a strategy to efficiently generate, trial and compare decision policies and retain the best performing one. For this purpose, we use Reinforcement Learning (RL) to explore CND action and state spaces, and learn which policy optimally reduces risk. A simulated CND environment is implemented using discrete event scheduling and a novel CND graph model. An asset valuation technique based on business needs, and a dynamic risk assessment algorithm are proposed to provide evaluation metrics (Figure 1).

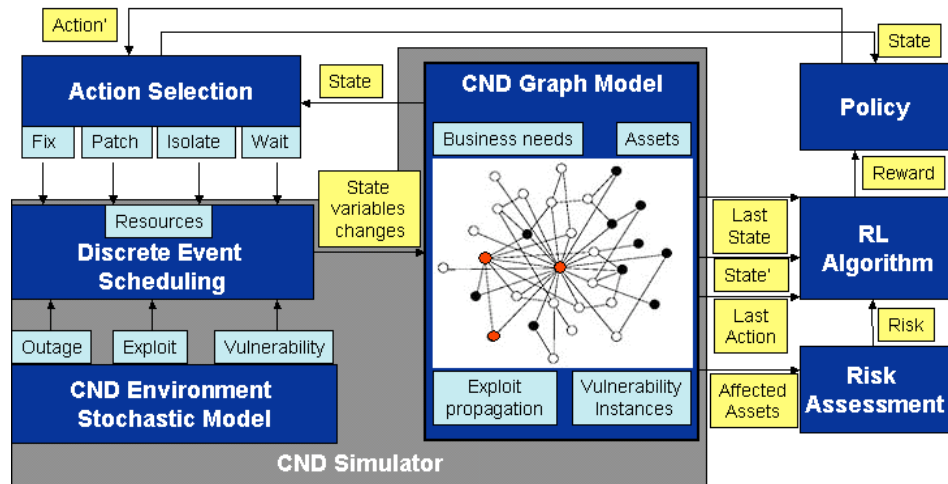


Figure 1 Architecture of an Autonomic CND system using RL and risk states.

This experimental framework is used to train two Reinforcement Learning agents and compare their policy risk performances against three simple heuristic-based policies. These different policies are evaluated in various CND scenarios, including determining optimal repair and patch sequences of multiple affected assets, and choosing between isolating vulnerable services and assuming the potential damages from exploits spreading on the infrastructure.

We found that both Reinforcement Learning policies can improve the overall risk in a significant manner. We also found that these RL policies can converge to the same globally optimum policy, in a limited state space scenario. We show that *addressing the affected asset with the highest value first* is a simple policy yielding superior risk results in many cases. Finally, we show that the difference between policies becomes less significant as the response resources are increased.