Chapter 25
Reinforcement Learning's Contribution to the Cyber Security of Distributed Systems: Systematization of Knowledge

Christophe Feltus
https://orcid.org/0000-0002-7182-8185
Luxembourg Institute of Science and Technology, Luxembourg

ABSTRACT

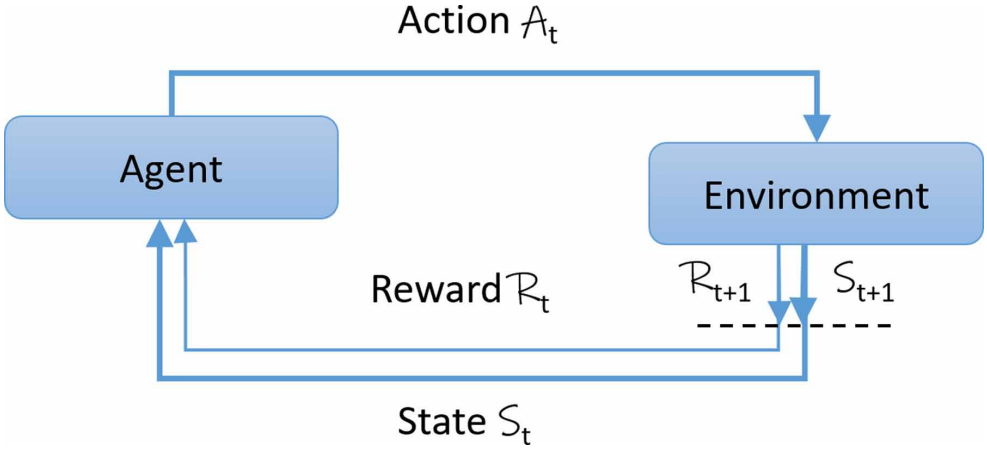
Reinforcement learning (RL) is a machine learning paradigm, like supervised or unsupervised learning, which learns the best actions an agent needs to perform to maximize its rewards in a particular environment. Research into RL has been proven to have made a real contribution to the protection of cyberphysical distributed systems. In this paper, the authors propose an analytic framework constituted of five security fields and eight industrial areas. This framework allows structuring a systematic review of the research in artificial intelligence that contributes to cybersecurity. In this contribution, the framework is used to analyse the trends and future fields of interest for the RL-based research in information system security.

1. INTRODUCTION

The contribution of artificial intelligence to cyber-security is paramount, given that it has the potential to increase the security level of the defended distributed system (Feltus et al., 2007) up to the state-of-the-art level generally reached by the attackers. In the field of machine learning, the approaches by which the computer program learns to generate output from experiments are classified into three paradigms: supervised, unsupervised and reinforcement learning (RL). In supervised learning, the model is trained using the input data labels, in unsupervised learning, the model is trained using patterns discovered in the input data, and in RL, a software agent learns to react on its own to an environment that it does not yet know (Van Otterlo & Wiering, 2012).

DOI: 10.4018/978-1-6684-7132-6.ch025

Figure 1. RL's mechanism schema



Reinforcement learning involves agents, states (S), and actions per state (A). Agents evolves from state to state when they perform actions. In order to learn how to react, agents make decisions and take action at time t A_t – (Fig.1) with the objective of accumulating rewards (R_t) while avoiding errors. As RL algorithms mostly use dynamic programming techniques, this reward-based environment is typically represented in the of Markov decision processes. These processes reflect a straightforward description of the problem in order to learn to reach a desired goal. In practice, agents continually select actions while the form environment in which they behave responds and presents new situations (Fig. 1)

In contrast to classical dynamic programming methods, RL algorithms have no knowledge of the exact Markov decision processes. Q-Learning [38] is an RL algorithm, whose purpose is to learn the policy that informs agents of the action they have to achieve in determined situations. This policy is optimized and gives all the successive steps necessary to achieve a goal while maximizing the gain of the rewards. Agents that learn the environment must continuously choose between exploiting the knowledge learned and exploring new potential actions to perform. Hence, an important parameter to be considered while defining RL algorithms is the e-greedy, which represents the proportion of exploration vs. exploitation actions (e.g., Li et al., 2018).

Reinforcement learning has already proven to be worthwhile for many fields, such as operations research, multi-agent systems, genetic algorithm or game theory. For some years, it has also been regarded as a strong potential contributor to the security and cyber-security domains (Feltus et al., 2009). However, although reviews of the contributions of machine learning and deep learning to computer security have already been undertaken for very specific fields, like biometry (e.g., Sundararajan & Woodard, 2018), to our knowledge, no systematic deep analysis of the contributions of reinforcement learning to the different fields of cyber-security has ever been completed. This is the aim of this paper. Elaborated from the strategic literature review method (Petersen et al., 2015), the paper will successively answer three knowledge questions: