Towards Autonomic Intrusion Response Systems

Stefano Iannucci
Mississippi State University
Email: stefano@dasi.msstate.edu

Sherif Abdelwahed
Mississippi State University
Email: sherif@ece.msstate.edu

***Abstract*—Intrusion Response Systems (IRSs) have been a major research topic in the last decade. At the core of an IRS is the response selection algorithm, which selects the best response action to counter the currently detected attack. This work advances the state of the art by proposing a meta-model based on Multi-Agent Markov Decision Processes which can be used to model a system and to plan for multi-objective, optimal, long-term, eventually proactive response policies. Experimental results show that long-term policies always outperform short-term ones and a thorough performance assessment shows that the proposed approach can be adopted to secure large systems.**

I. Introduction

The amount of attacks directed to computer systems is increasing year by year, and security mechanisms, such as firewalls, encryption and properly configured access control policies have quickly shifted from being *the* defense mechanisms to being just the *first line* of defense. Signature-based or anomaly-based Network Intrusion Detection Systems (IDSs) are usually used as a second line of defense. However, the huge number of cyber-attacks makes it infeasible for the system administrators to manually handle all the alerts generated by the IDSs. Intrusion Response Systems (IRSs) try to address this problem by automatically selecting the responses to the attacks detected by the IDSs. Most of the works proposed so far (e.g. [5], [6]), either try to model the behavior of the attacker or to model the dependencies between the system components, but none of them introduces a comprehensive model able to describe the attacker behavior, the defender (IRS) behavior and the actual system dynamics. In this work we model a system controlled by an IRS: unlike other approaches, we do not select a single short-term optimal response action, rather we produce an optimal long-term policy, that is, an optimal sequence of response actions able to drive the system from its initial (under attack) state to a set of target (desired) states.

This work advances the state of the art in dynamic IRS by proposing a meta-model, a prototype, and an extensive effectiveness and performance evaluation of a IRS supporting: (i) optimal long-term response policies; (ii) proactive policies; (iii) multi-objective optimization; (iv) IDS uncertainty in attack detection. The proposed meta-model is based on the Markov Decision Process (MDP) [3] framework and supports modeling: (i) the defended system behavior and (ii) the attacker behavior. When only the former is taken into account, a Single-Agent (SA)-MDP is used, while when both of them are considered, a Multiple-Agent (MA)-MDP is employed [3].

II. IRS Meta-Model

In this section we describe the MA-MDP meta-model, being the SA-MDP easily obtainable by removing any reference to the attacker and to its characterization.

In the MA-MDP meta-model, states are composed by joining three macro-attributes: (i) the current attack vector **p**, (ii) the system variables **v** and (iii) the attack belief vector **b**. The first contains as many variables as the number of attacks detectable by the IDSs and each variable $p_i \in \mathbf{p}$ represents the probability value that the system is currently under attack i. The second represent the current system status; the third contains as many variables as the number of attacks executable by the attacker and each variable $b_j \in \mathbf{b}$ represents the probability that the attacker is going to execute the attack j in a future stage of the game. The dynamics of the system are represented by joint actions $(x, y)_k$, where $x \in A_{irs}$ represents the defense action chosen by the IRS and $y \in A_{attacker}$ represents the attack action chosen by the attacker at stage k. Both defense and attack actions are characterized by pre-conditions and post-conditions: the former are boolean expressions on the state attributes, that is, they identify a subset of the state space in which the actions are executable; the latter define instead a probability distribution over the possible next states reachable by the agents after the execution of the joint action.

Upon the execution of an action, the agents get a joint reward $R_k = (R_{k_{irs}}, R_{k_{attacker}})$. We define a negative reward R_{irs} for the IRS agent and a positive reward $R_{attacker}$ for the attacker, being not necessarily $R_{irs} = -R_{attacker}$. The former is a weighted sum over three pre-defined criteria (response time, cost and impact), while the latter is instead defined as a static mapping with the attack actions. The game between the attacker and the IRS, which can contain proactive defense actions due to the attack beliefs contained in **b**, concludes when the IRS drives the system in one of the states belonging to the subset of the target states $S_{tgt} = \{s|F = true\}$, where F is a termination function expressed as a boolean condition on the state attributes.

III. Effectiveness Evaluation

The parameter γ is the MDP discount factor and can be used to specify how much short-term rewards should be preferred over long-term rewards: setting $\gamma = 0$ results in planning optimal short-term policies, while setting it to a value close to 1 results in considering long-term rewards. In Figure 1 we show that the policies generated by solving the MDP with $\gamma = 0.9$ always outperform policies planned with $\gamma = 0$ in

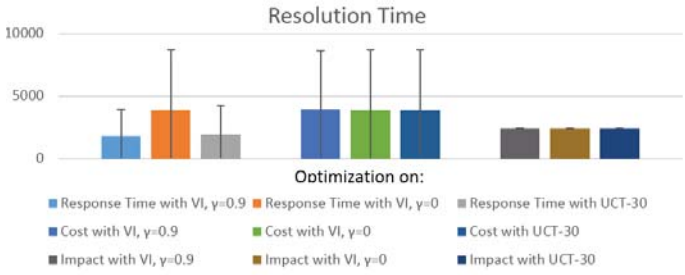

Fig. 1. Response Policies Comparison

a SA-MDP setting, with a system composed of 17 attributes and with 18 response actions available. The metric used in the comparison is the response time, evaluated for plans generated with both an optimal and a sub-optimal solver.

Furthermore, in the following, we report an example short game played between the attacker (on the left hand side) and the IRS (on the right hand side), showing that the IRS is able to prevent future attacks:

1) portScanAttack / generateAlert
2) attackVsftpd / increaseLogVerbosity
3) noOp / quarantineSystem
4) noOp / manualResolution
5) noOp / softwareUpdate

The game starts with the attacker performing a port-scan and the protected system immediately reacts with an alert generation; at the second stage, the attacker exploits a vulnerability on the ftp server. Increasing the log verbosity is not a direct response to this attack, rather it is still due to the initial port-scan. The reward function is set up so that the manual resolution, which requires a quarantined system, is the best choice to counter the attack in this case and thus it is executed afterwards. However, since the software remains vulnerable even after the manual resolution, the attacker could be able to exploit again the vulnerability. Therefore, as a final proactive action, the IRS triggers a software update.

IV. Performance Evaluation

One of the mostly used algorithm to solve MDPs is Value Iteration (VI) [2]. Unfortunately, its execution time is exponential in the number of the state attributes. However, only the attributes that directly or indirectly help in facing the threat can be considered. To this end, we designed a feature selection algorithm that is able to instantiate the minimal MDP needed to compute optimal response policies to counter the currently detected attack. Should this shrewdness not be enough, the sub-optimal algorithm UCT [4] applied to long-term planning can provide better results than optimal short-term planning, as shown in Figure 1. Figure 2 compares the planning time of the VI algorithm implemented in [1] for solving SA- and MA-MDPs with the planning time of our parallel implementation named Parallel-VI. The latter exhibits an almost linear speedup and, therefore, when executed with 10 thread spread across 10 cores, the planning time is reduced by one order of magnitude. Finally, the comparison is also made with the sub-optimal

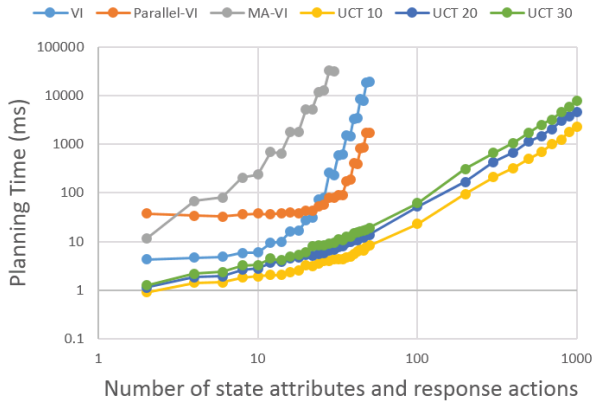

Fig. 2. Planning Time Comparison

planner, configured with a look-ahead of 10, 20 and 30 steps. The performance gap between the optimal and sub-optimal planners is 10% in average in the considered experiment.

V. Conclusions and Future Works

None of the IRS proposed so far provide a full framework supporting multi-objective optimization, long-term policies and attacker/defender modeling. In this work we described the meta-model based on MA-MDP. The effectiveness evaluation shows that long-term policies always outperform short-term ones and the performance evaluation shows that the proposed approach can be used to defend large systems.

As a future work we plan to establish a feedback loop between the protected system and the IRS model, in order to continuously update the initial parametrization of the actions postconditions with the real state transition probabilities. Furthermore, we plan to offload the execution of the IRS to Intel Xeon Phi coprocessor as well as GPUs, in order to (i) potentially obtain better performances and to (ii) not impact on the available CPU resources.

Acknowledgement

Funding for this work was (partially) provided by the Pacific Northwest National Laboratory, under U.S. Department of Energy Contract DE-AC05-76RL01830.

References

[1] Brown-umbc reinforcement learning and planning (burlap). http://burlap.cs.brown.edu/.

[2] R. Bellman. Dynamic programming. princeton, nj: Princeton university-press. *BellmanDynamic Programming1957*, 1957.

[3] L. Busoniu, R. Babuska, and B. De Schutter. A comprehensive survey of multiagent reinforcement learning. *IEEE Trans. on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 38(2):156–172, 2008.

[4] L. Kocsis and C. Szepesvári. Bandit based monte-carlo planning. In *Machine Learning: ECML 2006*, pages 282–293. Springer, 2006.

[5] S. Ossenbuhl, J. Steinberger, and H. Baier. Towards automated incident handling: How to select an appropriate response against a network-based attack? In *IT Security Incident Management & IT Forensics (IMF), 2015 Ninth International Conference on*, pages 51–67. IEEE, 2015.

[6] A. Shameli-Sendi and M. Dagenais. Orcef: Online response cost evaluation framework for intrusion response system. *Journal of Network and Computer Applications*, 2015.