An Intrusion Detection Game with Limited Observations

Tansu Alpcan[1] and Tamer Başar[2]

Abstract

We present a 2-player zero-sum stochastic (Markov) security game which models the interaction between malicious attackers to a system and the IDS who allocates system resources for detection and response. We capture the operation of a sensor network observing and reporting the attack information to the IDS as a finite Markov chain. Thus, we extend the game theoretic framework in [1] to a stochastic and dynamic one. We analyze the outcomes and evolution of an example game numerically for various game parameters. Furthermore, we study limited information cases where players optimize their strategies offline or online depending on the type of information available, using methods based on Markov decision process and Q-learning.

I. Introduction

Intrusion detection systems (IDSs) monitor various events in a networked system and analyze them for signs of security compromises [2]. By extending the information security paradigm beyond traditional protective (e.g. firewalls) and reactive measures (e.g. virus and malware detection), they increase the ability of the system administrator to control the system, and help him or her better manage its security [3]. In recent years an increasing number of security related problems in networked systems have resulted in a surge of interest and research in this area. However, the majority of the earlier literature on intrusion detection (ID) relies on ad-hoc schemes and experimental work. We believe that a quantitative decision framework is needed in order to address issues like attack modeling, allocation of limited system resources, and decision on response actions.

Game theory provides a rich set of tools to study problems where multiple players with different objectives interact and compete with each other on the same system. Therefore, game theory is a strong candidate to provide the much needed mathematical framework for analysis, modeling, decision, and control processes for information security and intrusion detection. Such a mathematical abstraction is useful for generalization of problems, combining the existing ad-hoc schemes under a quantitative umbrella, and future research [1]. Consequently, game theory has been recently proposed by several studies for a theoretical analysis of ID [1], [3]–[7].

We have recently investigated a game theoretic approach to network security where the interaction between malicious attackers and the IDS is modeled using 2-player static and repeated noncooperative games [1], [3]. In this setting, the attacker(s) who attempt to gain unauthorized access to a networked system or render it incapacitated through denial of service are represented by one player and the second player is the IDS which allocates system resources to collect information for detection and decides on a response. Furthermore, we have introduced a third *sensor (network) player* in [1], similar to the *nature* player in game theory. This player imperfectly observes the actions of the attackers and conveys this information to the IDS. It has a fixed "strategy" representing probabilistically how well the conveyed information matches the real attack information.

In this paper, we consider a stochastic and dynamic extension of the game theoretic framework introduced in [1] and model the (network of) sensors observing and reporting the attacks to the IDS as a

[0]Research supported in part by a grant from the Boeing Company.

[1]Tansu Alpcan is with the Deutsche Telekom Laboratories, Technische Universität Berlin, Ernst-Reuter-Platz 7, 10587 Berlin, Germany. E-mail: *tansu.alpcan@telekom.de*, Web: http://decision.csl.uiuc.edu/˜alpcan

[2]Tamer Başar is with the Coordinated Science Laboratory, University of Illinois, 1308 West Main Street, Urbana, IL 61801 USA. E-mail: *tbasar@control.csl.uiuc.edu*, Web: http://decision.csl.uiuc.edu/˜tbasar

finite-state Markov chain.[1] As a result we obtain a 2-player Markov (stochastic) game that is investigated under various assumptions on the nature of the information available to individual players. The approach in this paper departs from the model we had introduced in [1] in several aspects. The stochastic Markov model considered here enables us to further capture the complexities of the underlying system. By removing in some cases the full information assumption from the game in [1] we obtain a more realistic depiction of the attacker vs. IDS interaction. When limitations are imposed on information available to players then the outcome and evolution of the game may depend on the success of players' estimations on the system and their learning rate. Another major difference is the dynamic nature of the current model. The static games in [1] are played repeatedly over time in a myopic manner. On the other hand, the players we study here optimize their strategies taking future (discounted) costs into account using the ongoing attacker-IDS interaction as well as deploying dynamic learning methods. They can refine their own strategies offline or online by continuously learning more about the system and their adversaries. (An underlying assumption is that the game's characteristics such as system parameters and player preferences remain stationary for the duration of the game). Thus, we investigate various learning schemes where the players base their decisions on either prior information (on the system and costs) or observations obtained during game play, or a combination thereof. Markov decision processes (MDPs) and Q-learning methods [8] provide part of the theoretical foundation for the development and analysis of player strategies.

Markov games have been studied extensively by the research community in recent years [9]–[14]. Littman [9] has investigated 2-player zero-sum Markov games and Q-learning based methods for solving them dynamically. An extension to this approach has been suggested in [12] again for zero-sum games. Other studies have focused on learning in non-zero-sum Markov games [11], [13], [14]. In such cases non-uniqueness of Nash equilibria at each stage of the game poses a difficult challenge in terms of convergence of solutions. Various methods have been suggested to overcome these convergence problems such as generalizations of Nash equilibrium to correlated equilibria [11], choosing asymmetric learning rates for players to prevent synchronization [10], and cooperation schemes [13].

In security games the players (attacker(s) and IDS) are direct adversaries which can be modeled using zero-sum games in most cases. Hence, we consider in this paper zero-sum Markov games exclusively, which improves the tractability of the problems at hand. We will specifically utilize multi-agent Markov decision processes and Q-learning methods similar to the ones in [8], [9] to solve these security games. The games considered are not ones of full information due to the fact that each player observes the other players' moves and the evolution of the underlying system sometimes only partially and indirectly. Therefore, the players have to base their decisions on their own occurring costs and limited observations of the system and other players' actions. Specifically, we study the cases where the players optimize their strategies with decreasing information available to them using methods such as value iteration to solve MDPs, minimax-Q [9], and naive Q-learning.

The organization of the paper is as follows: we will present in the next section a 2-player zero-sum stochastic security game formulation which is a dynamic and stochastic variation of the framework introduced in [1]. In Section III we analyze some example cases of this game numerically under different information assumptions and learning methods. The paper will conclude with remarks in Section IV.

II. The Stochastic Security Game

We consider a 2-player (one representing the attacker(s) and one the IDS) zero-sum finite Markov game model, where each player has a finite number of actions to choose from. The attacker's action space is defined as $A := \{a_1, a_2, \ldots, a_{Amax}\}$ and constitutes of the various possible attack types. The IDS's action space is denoted by $R = \{r_1, r_2, \ldots, r_{Rmax}\}$ and includes both passive actions such as setting an alert,

[1]This sensor network may constitute, for example, of virtual sensors on a computer network implemented through software agents or of motion or chemical sensors in a physical building setting.

and active ones like taking precautions or gathering further information. In this security game, similar to the one in [1], a network of sensors convey information to the IDS regarding the attacker's actions. Here, it is modeled as the stochastic underlying system on which the players interact. The output of the sensor network is captured by a finite number of environment states, $S = \{s_1, s_2, \ldots, s_{Smax}\}$, where each state may represent detection of a specific type of attack or correspond to "no detection". Different from our earlier formulation, we model the sensor network as a finite-state Markov chain which enables us to utilize well-established analytical tools to study the problem.

The probability of the sensor network's output being in a specific state is given by the vector $\mathbf{x} := [x_1, \ldots, x_{Smax}]$, where $0 \leq x_i \leq 1 \; \forall i$ and $\sum_{i=1}^{Smax} x_i = 1$. The transition probabilities between environment states are then described by the transition matrix M. If we ignore the effect of player decisions on the system, then we have $\mathbf{x}(n+1) = \mathbf{x}(n)M$, where $n \geq 1$ denotes the stage of the game. However, this assumption clearly does not hold as the operation and output of the sensor network depend on whether there is a specific type of an attack or not. Likewise the actions of the IDS may also affect the transition probabilities. Then, we enumerate the transition matrices by $M(k)$ such that $k \in A \times R$, where $A \times R$ is the cross product of the action spaces of the two players.

Each player is associated with a set of costs that is not only a function of the other players' actions but also the state of the system. The IDS's and the attacker's costs are $-c(s_l, r_i, a_j)$ and $c(s_l, r_i, a_j)$, respectively, where $s_l \in S$, $r_i \in R$, and $a_j \in A$. We assume that each player knows its own cost at each stage of the game.

One of the most interesting aspects of this security game is the amount of information each player has about the sensor network's characteristics and the actions of its opponent. In this paper, we focus on three different information structures: full information, no information about sensor network characteristics (transition probabilities, M), and having only information about own costs, past actions, and past states. In the full information case each player knows everything about the sensor network as well as the preferences and past actions of its opponent. Hence, players may utilize well-known MDP methods such as value iteration to calculate their own optimal mixed strategy solutions to the zero-sum game. When we relax the assumption of perfect knowledge on sensor network for the attacker, then he or she has to choose its strategy without knowing the characteristics of the sensors represented by the transition probabilities. In this case, the attacker can calculate its optimal strategy online (i.e. while playing the game) using a technique called minimax-Q, which is a variation of the standard Q-learning technique [9]. In both of these cases, we assumed that the players know each other's costs and observe their opponent's past actions. We consider next the situation where a player only observes the sensor network's output and keeps track of its own actions and costs. In this third case, we study single agent "naive" Q-learning (ignoring the other player's actions) as a possible approach. However, calculating an optimal strategy for this zero-sum Markov game under extremely limited information continues to be an interesting research question.

III. Illustrative Example and Numerical Analysis

We further clarify and study the stochastic security game presented in Section II with the numerical analysis of an illustrative example. For simplicity let us consider a single attack type being available to the attacker(s), i.e., $A = [a, \; na]$, where na denotes no attack. We also limit the possible actions of the IDS to "response" or "no response", $R = [r, \; nr]$. The states of the sensor network is then $S = [d, \; nd]$, corresponding to "attack detected" and "no detection". To simplify the analysis we assume that the IDS's response is passive in this example and does not affect the operation of the sensor network. Hence, there are only two transition (probability) matrices of size 2×2 indexed by the actions of the attacker: $M(a)$ and $M(na)$. The payoff or benefit values for the IDS and the attacker are enumerated by $[c_1^I, \ldots, c_8^I]$ and $[c_1^A, \ldots, c_8^A]$, respectively. A graphical depiction of this game is shown in Figure 1.

Let us further explain the game in Figure 1 by describing a specific scenario step by step, which corresponds to following a path from left to right in accordance with the order of players' actions. The

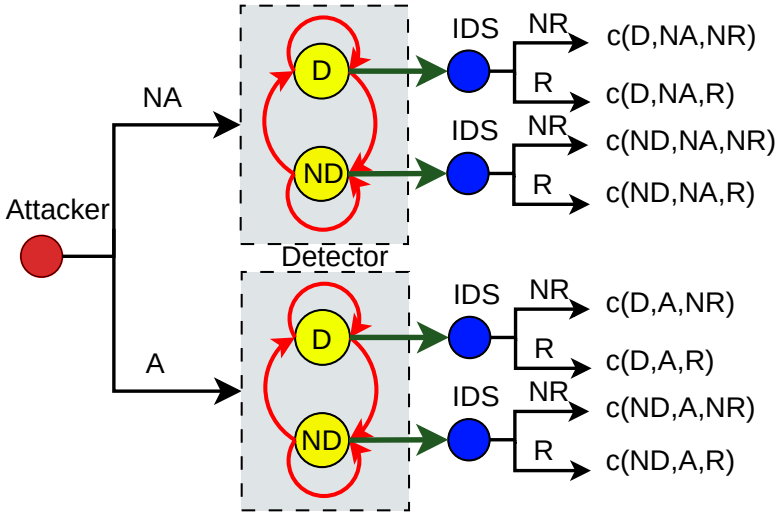


Fig. 1. An illustrative example of the stochastic security game

lower left branch in the figure, labeled A, indicates an attack by the attacker(s) to the system. The sensor network is represented with the lower box containing the Markov chain representing the detection process. Given the information from the sensor network, say D, the IDS decides in branch R to take a predefined response action. The outcome of this scenario is quantified by a cost of $c(D, A, R)$ to the attacker for a single game stage and $-c(D, A, R)$ (same amount of benefit) to the IDS.

We analyze the game numerically for a set of predefined parameters. First, we choose the transition probability matrices between sensor network states, $M(a)$ and $M(na)$, as given by (1). Notice that these transition probabilities model a well functioning sensor network.

$$M(a) = \begin{bmatrix} 0.9 & 0.1 \\ 0.9 & 0.1 \end{bmatrix}, \quad M(na) = \begin{bmatrix} 0.1 & 0.9 \\ 0.1 & 0.9 \end{bmatrix} \tag{1}$$

The cost matrix $C(S \times A \times R)$ of the attacker is:

$$C(S \times A \times r) = \begin{bmatrix} 5 & -1 \\ 5 & -5 \end{bmatrix}, \quad C(S \times A \times nr) = \begin{bmatrix} -10 & 0 \\ -10 & 0 \end{bmatrix} \tag{2}$$

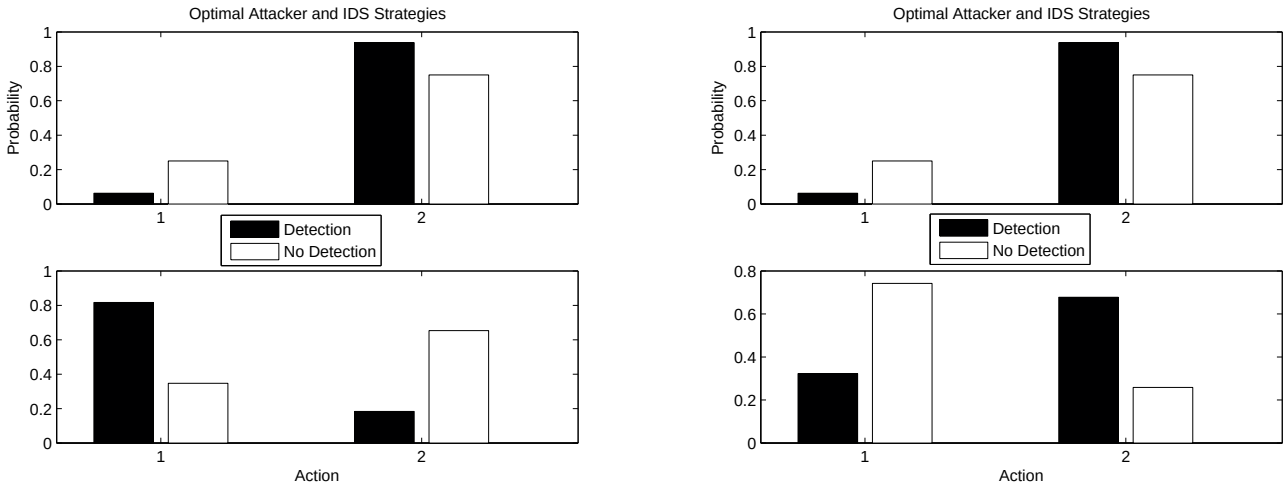
Then, the benefit matrix of the IDS is simply $-C(S \times A \times R)$ due to the zero-sum game formulation. We adopt in this paper the convention that the attacker is the row player (of the corresponding matrix game) and minimizer (of its cost) whereas the IDS is the column player and maximizer (of its benefit). For example, if there is an attack which is detected and responded by the IDS, then the cost to the attacker is $c(d, a, r) = 5$ and the IDS's benefit is 5. Similarly, if there is no attack, detection, or response, then both parties get $c(nd, na, nr) = 0$ benefit.

Under the full information assumption, the players obtain their optimal mixed strategies by solving an infinite horizon MDP with a discount factor of 0.6 where the saddle point of each stage game is calculated separately solving a linear program. Here, we replace the minmax-Q scheme described in [9] with a standard value iteration algorithm where both players perfectly know the transition probabilities M. The resulting optimal policies for the attacker $P^A(S \times A)$ and the IDS $P^I(S \times R)$ are shown in Figure 2(a), where actions 1 and 2 of the attacker and the IDS correspond to $[a, na]$ and $[r, nr]$, respectively. The corresponding final costs of the IDS are $[J(d),\ J(nd)] = [1.2,\ 5.2]$. We observe that the general structure of the results obtained are in agreement with the ones of the static game in [1].

We next investigate the case of defective sensors by redefining the transition probabilities in (1) as

$$M(a) = \begin{bmatrix} 0.1 & 0.9 \\ 0.1 & 0.9 \end{bmatrix}, \quad M(na) = \begin{bmatrix} 0.9 & 0.1 \\ 0.9 & 0.1 \end{bmatrix} \tag{3}$$

Here, the sensors report almost the opposite of the actual attack information. The resulting final costs for the IDS, $[J(d),\ J(nd)] = [5.8,\ 4.3]$, are higher than the previous ones. This is expected as the IDS relies on the output of the sensors to make its decisions as depicted in Figure 2(b). Hence, sensor network behavior affects the outcome of the game. For example, the IDS responds more aggressively even in the cases when there is no detection reported to compensate for the inadequacies of the sensors.



(a) (b)

Fig. 2. The optimal strategies of the attacker and the IDS for (a) reliable and (b) defective sensors. Actions 1 and 2 correspond to $[a, na]$ and $[r, nr]$, respectively.

As another variation on the sensor behavior, we consider a set of sensors which report incidents with a uniform random probability at the steady state, i.e., all transition probabilities are chosen to be equal to 0.5. The final strategies of the players for this case are shown in Figure 3(a). It is at first counter intuitive to note that sensor behavior does not affect the attacker strategies. However, considering that the attacker bases its decisions on previous actions of the IDS and sensor output states as visualized in Figure 1, this result is not surprising.

For the remainder of the simulations, we vary player costs as:

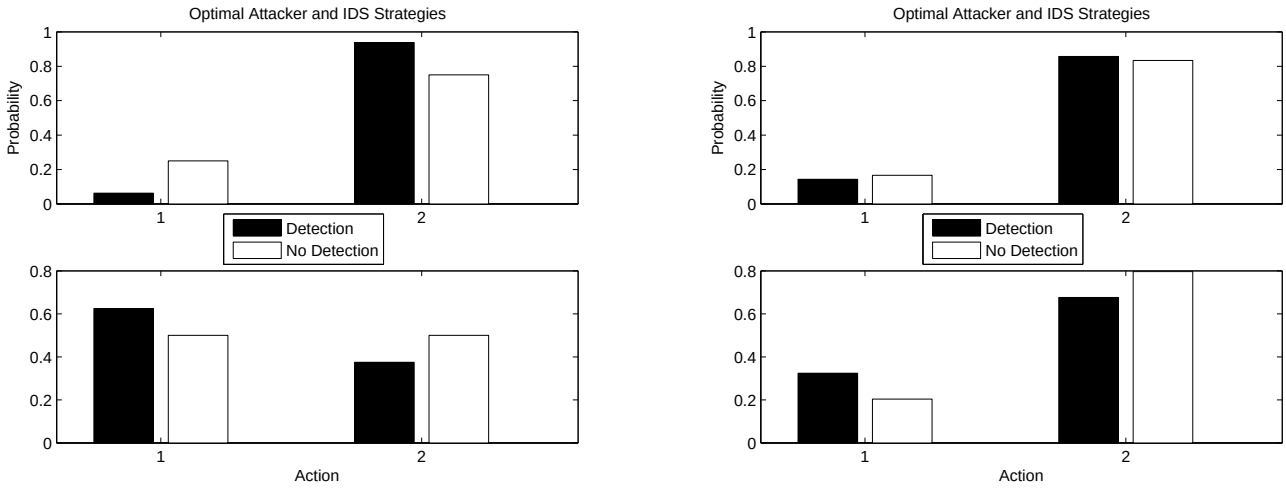
$$C(S \times A \times r) = \begin{bmatrix} 20 & -5 \\ 10 & -3 \end{bmatrix}, \quad C(S \times A \times nr) = \begin{bmatrix} -10 & 0 \\ -5 & 0 \end{bmatrix} \tag{4}$$

The costs in (4) reflect a further distinction between actions taken by the IDS in response to a detection and others. Repeating previous simulations with this set of costs, we observe a similar pattern in terms of player strategies and final costs. In conjunction with the new cost, we also modify the sensor behavior to

$$M(a) = \begin{bmatrix} 0.6 & 0.4 \\ 0.6 & 0.4 \end{bmatrix}, \quad M(na) = \begin{bmatrix} 0.2 & 0.8 \\ 0.2 & 0.8 \end{bmatrix}. \tag{5}$$

In (5), unlike previous cases of (1) and (3), the probability of missing an attack and probability of false alarms are asymmetric. The resulting player strategies are shown in Figure 3(b).

Until now, we have assumed that both players know the sensor behavior (i.e. the transition probabilities described in (1), (3), and (5), perfectly. Therefore, players can calculate their strategies using an iterative



(a) (b)

Fig. 3. The optimal strategies of the attacker and the IDS for sensors (a) reporting events randomly and (b) having asymmetric transition probabilities.

algorithm or other methods offline as they already have full knowledge on the game. We next relax this assumption and investigate a version of the stochastic game where a player learns about these transition probabilities online, i.e., while playing the game. Then, the value iteration algorithm is replaced by minmax-Q for the learning player. We specifically consider the case where the attacker learns. Figure 4 depicts the convergence of the player cost values. We observe, as to be expected, that the cost values of the player deploying the minimax-Q scheme takes longer to converge to their saddle-point values. However, it is important to note that in the long run lack of knowledge on the sensor behavior (or underlying system) does not prevent the player from exploring and eventually locking into the optimal solution.

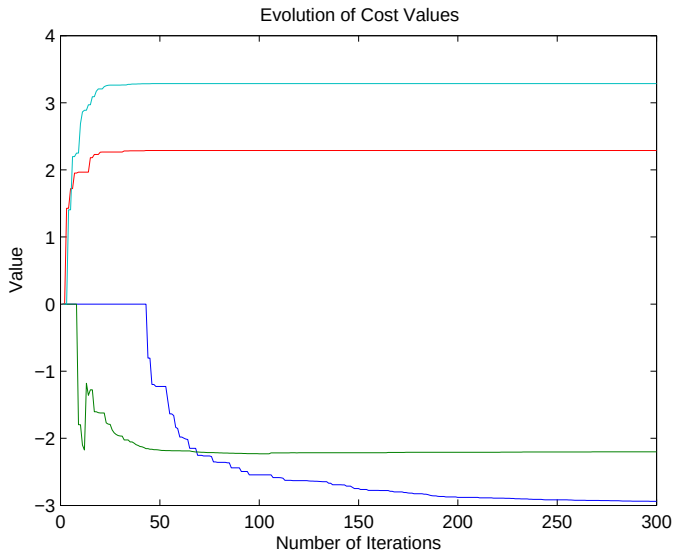


Fig. 4. The evolution of cost values when the attacker deploys the minimax-Q learning algorithm.

Finally, we study the very limited information case where a player knows only his or her own actions and the corresponding costs, ignoring both the sensor behavior and the opponent. Unlike the previous two

cases, in this case there are no algorithms available to players that guarantee convergence to an optimal solution. Instead, we let in this paper the players deploy a naive single agent Q-learning scheme to optimize their strategies online. We then investigate the results numerically for different sets of parameters. The cost values for the players are again given by (4) and we initially assume well functioning sensors as in (1). In the first simulation, the attacker is the one without any information and uses the naive Q-learning scheme while the IDS deploys the minmax Q-learning method to update its strategy. The resulting strategy for the attacker is a deterministic one since in the single agent formulation of the problem deterministic strategies yield an "optimal" solution. The attacker's policy converges to $[P^A(d),\ P^A(nd)] = [na,\ na]$ whereas the IDS's one is $[P^I(d,r),\ P^I(d,nr)] = [0.29,\ 0.71]$, $[P^I(nd,r),\ P^I(nd,nr)] = [0.28,\ 0.72]$. The evolution of attacker and IDS cost values are depicted in Figure 5, where we observe that these policies clearly do not lead to a saddle-point solution. However, it is interesting to note the convergence of cost values and strategies.

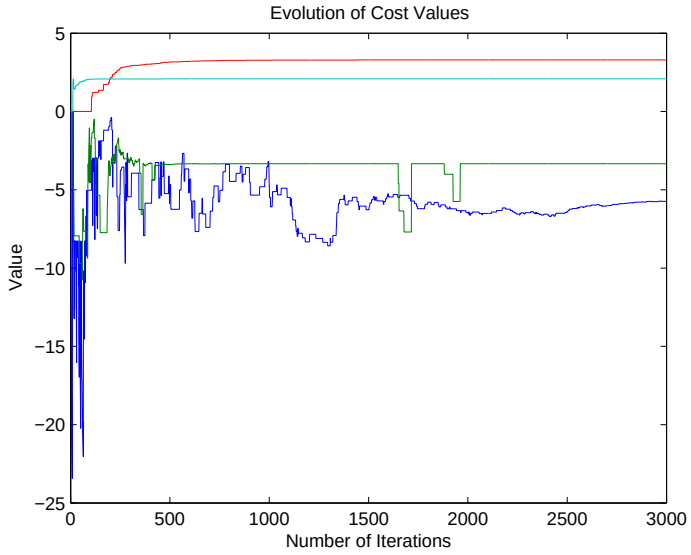
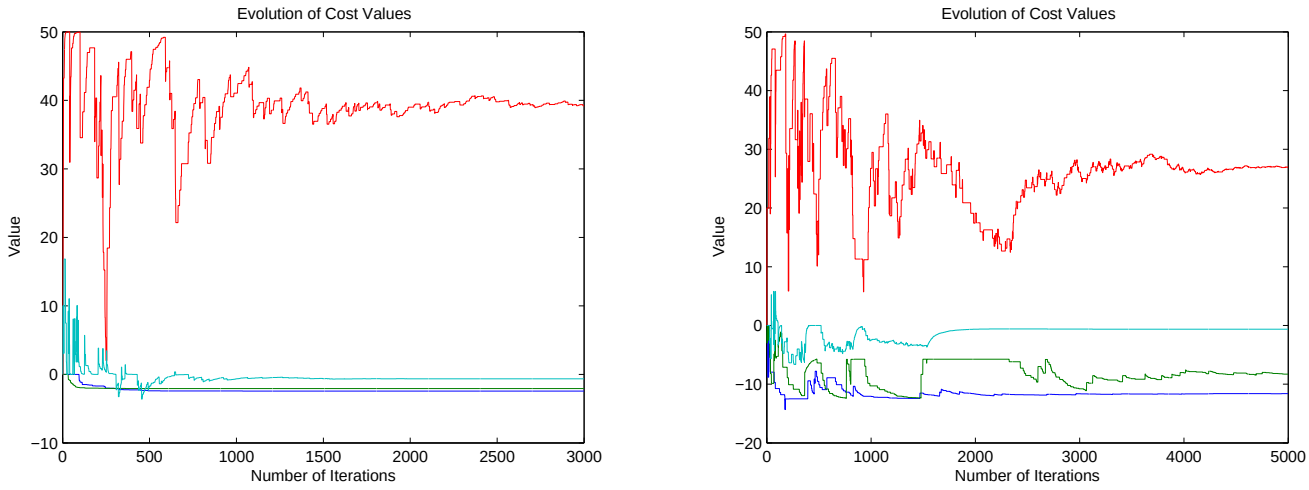


Fig. 5. The evolution of cost values when the attacker deploys single agent Q-learning and the IDS minmax-Q learning scheme.

We next switch the roles of the attacker and the IDS, and repeat the simulation. The IDS's policy converges to $[P^I(d),\ P^I(nd)] = [r,\ nr]$ whereas the attacker's one is $[P^A(d,a),\ P^A(d,na)] = [0.16,\ 0.84]$, $[P^A(nd,a),\ P^A(nd,na)] = [0.17,\ 0.83]$. Figure 6(a) shows the evolution of attacker and IDS cost values. In both of these simulations the results match our intuitive expectations despite limitations and naivete of the method used. Due to well-functioning sensors, the attacker is deterred from attacking in the first one and the IDS responds only when an attack is reported in the second one.

The natural next step is to study the case when both players deploy the naive Q-learning scheme. In this case, the resulting strategies are $[P^A(d),\ P^A(nd)] = [na,\ a]$ and $[P^I(d),\ P^I(nd)] = [r,\ nr]$. Again these policies can be justified with the reliability of sensors in reporting attacks. The evolution of attacker and IDS cost values are depicted in Figure 6(b). It is interesting to observe that the cost values of players exhibit convergent behavior. This may be due to our choice of asymmetric learning-rate parameters in the algorithms in accordance with the analysis in [15] Finally, we rerun this simulation after updating the sensor behavior to the one in (1), i.e., with defective sensors. The player policies are then $[P^A(d),\ P^A(nd)] = [a\ na]$ and $[P^I(d),\ P^I(nd)] = [nr,\ r]$, which are the opposite of the previous ones. This expected result (due to the opposite sensor behavior) indicates that player strategies do not converge to random values when the naive Q-learning scheme is used.



(a) (b)

Fig. 6. The evolution of cost values when the IDS deploys single agent Q-learning and the attacker minmax Q-learning scheme.

IV. Conclusion

We have presented a 2-player zero-sum stochastic (Markov) security game which models the interaction between malicious attackers to a system and the IDS who allocates system resources to collect information for detection and decides on a response. By capturing the operation of a sensor network observing and reporting the attack information to the IDS as a finite-state Markov chain, we have extended the game theoretic framework in [1] to a stochastic and dynamic one. Through the numerical analysis of an illustrative example we have studied the optimal mixed strategy solutions as well as evolution of player costs under various game parameters.

The games considered are not ones of full information due to the fact that each player observes the other players' moves and the evolution of the underlying system sometimes only partially and indirectly. Therefore, we have studied the cases where players optimize their strategies with limited information available, using methods such as MDP value iteration, minimax-Q [9], and naive Q-learning.

Our study -despite admittedly being small in scale- demonstrates the versatile nature of the framework we have proposed. Future research directions include various applications of the model introduced to real world ID problems as well as further study of other possible algorithms in the limited information case.

References

[1] T. Alpcan and T. Başar, "A game theoretic analysis of intrusion detection in access control systems," in *Proc. of the 43rd IEEE Conference on Decision and Control*, Paradise Island, Bahamas, December 2004, pp. 1568–1573.

[2] R. Bace and P. Mell, "Intrusion detection systems," NIST Special Publication on Intrusion Detection Systems, http://www.snort.org/docs/nist-ids.pdf.

[3] T. Alpcan and T. Başar, "A game theoretic approach to decision and analysis in network intrusion detection," in *Proc. of the 42nd IEEE Conference on Decision and Control*, Maui, HI, December 2003, pp. 2595–2600.

[4] D. A. Burke, "Towards a game theoretic model of information warfare," Master's thesis, Air Force Institute of Technology, Air University, November 1999.

[5] K.-W. Lye and J. Wing, "Game strategies in network security," in *Foundations of Computer Security Workshop in FLoC'02*, Copenhagen, Denmark, July 2002.

[6] P. Liu and W. Zang, "Incentive-based modeling and inference of attacker intent, objectives, and strategies," in *Proc. of the 10th ACM Computer and Communications Security Conference (CCS'03)*, Washington, DC, October 2003, pp. 179–189.

[7] A. Agah, S. Das, K. Basu, and M. Asadi, "Intrusion detection in sensor networks: A non-cooperative game approach," in *3rd IEEE International Symposium on Network Computing and Applications, (NCA 2004)*, Boston, MA, August 2004, pp. 343–346.

[8] D. Bertsekas, *Dynamic Programming and Optimal Control*, 2nd ed. Belmont, MA: Athena Scientific, 2001, vol. 2.

[9] M. L. Littman, "Markov games as a framework for multi-agent reinforcement learning," in *Proc. of the Eleventh International Conference on Machine Learning*, San Francisco, CA, 1994, pp. 157–163.

[10] M. Zinkevich, A. Greenwald, and M. L. Littman, "Cyclic equilibria in markov games," in *Proc. of Neural Information Processing Systems, NIPS 2005*, Vancouver, BC, Canada, December 2005.
[11] A. Greenwald and K. Hall, "Correlated q-learning," in *Proc. of the Twentieth International Conference on Machine Learning*, Washington, DC, 2003, pp. 242–249.
[12] M. G. Lagoudakis and R. Parr, "Learning in zero-sum team markov games using factored value functions," in *Proc. of Neural Information Processing Systems, NIPS 2002*, Vancouver, BC, Canada, December 2002, pp. 1659–1666.
[13] R. Aras, A. Dutech, and F. Charpillet, "Cooperation through communication in decentralized markov games," in *Proc. of IEEE Int. Conf. on Advances in Intelligent Systems - Theory and Applications - AISTA'2004*, Kirchberg, Luxembourg, 2004.
[14] J. Hu and M. P. Wellman, "Nash Q-learning for general-sum stochastic games," *Journal of Machine Learning Research*, vol. 4, no. 1, pp. 1039–1069, November 2003.
[15] D. S. Leslie and E. J. Collins, "Individual Q-learning in normal form games," *SIAM Journal on Control and Optimization*, vol. 44, no. 2, pp. 495–514, 2005.