Using Game Theory with Intrinsic Motivation to Examine Anti-Hacking Policies for Autonomous Systems

Kathryn Merrick, Medria Hardhienata
School of Engineering and Information Technology
University of New South Wales, Canberra
ACT, Australia
k.merrick@adfa.edu.au

Kamran Shafi, Jiankun Hu
School of Engineering and Information Technology
University of New South Wales, Canberra
ACT, Australia
[k.shafi|j.hu]@adfa.edu.au

***Abstract*—With the increasing excitement about the emergence of autonomous systems, there is also a level of caution associated with the possible vulnerabilities of such systems. As systems become more independent, it may become more difficult for humans to trust them, particularly in settings where they may be vulnerable to malicious behavior by others. This paper presents a game theoretic model for autonomous systems subject to hacking. We propose a hacking game that models the interaction between an autonomous system and a hacker. Simulations were conducted to examine the outcomes for the victims (rational autonomous agents) when attacked by humans with different motives. The implications of the results for the decision-making process of the players are discussed. This paper also explores how game theory can be used by policy makers to choose a policy when autonomous systems have been attacked.**

Keywords—game theory, cyber war, intrinsic motivation, autonomous systems

I. AUTONOMOUS SYSTEMS

In this paper, we consider an autonomous system as a system that does not require direct human involvement to perform its tasks. In other words, the system operates without direct human oversight or input. Autonomous systems provide many compelling advantages over traditional systems. Automobile manufacturers, for example, are currently developing more autonomous features for cars [1] to reduce accidents. Such advantages are likely to make such systems a part of our everyday life in the coming years.

It is believed that there will be a strong market demand for cognitive systems when the technology is ready. Thus, establishing laws and policy in relation to the use of autonomous system becomes critical. This includes exploring and understanding the capabilities and limits of the autonomous technology. Such an understanding is needed to eliminate fears that 'robots may take the control over humans' and allow us to further explore the opportunities that these systems offer [2]. Furthermore, study on the ability of autonomous systems is needed as a guideline for policy makers to establish clear guidelines for future development of autonomous systems [2].

At the international level, autonomous technology in military operations has been partly addressed by the United Nations. In May 2014, experts on Certain Conventional Weapons (CCW) convened in Geneva to overcome the issue of lethal autonomous systems [2]. However, autonomous systems are likely to enter both military and civilian lives. Threats to autonomous systems may come from civilian sources such as hackers [3] and hactivists [4, 5], as well as from state-based attacks. Autonomous systems require a way to understand and respond to these threats autonomously. Likewise, future policy makers require ways to interpret these threats and inform the designers of autonomous systems.

While the use of autonomous systems is appealing, a dilemma may also occur from the accountability aspect [6]. If autonomous agents take over the role of humans, the question that arises is who will be responsible if something goes wrong? If front-line operators cannot be blamed for the actions of autonomous systems, some believe that accountability could be attributed to the creator of the systems, such as the engineers, computer programmers, and designers. However, even if a human operator is still engaged in the loop, some argue that it is not necessarily clear that she/he can always control the behavior of the autonomous systems [6]. Thus it might be unfair to expect that a human should always be responsible for system failures [6]. This paper considers ways that the system can reason about its own possible failure modes that result from malicious attacks.

One potential problem arising from the use of autonomous systems is the risk of their communication structures being hacked or hijacked. Adversaries can hack into computers, implant viruses and worms, shut down systems, or order fake commands and actions [7]. High-volume data transfers are vulnerable to interference of noises, wiretap, and interception of credentials [8]. It is therefore probable that hostile forces will launch attacks to disable autonomous systems by hacking their communication infrastructure [8]. Hacked autonomous systems could be dangerous not only to the military, but also to civilians. Furthermore, autonomous systems, such as unmanned air vehicles, can be reconstructed quite easily. The availability, stability, and low cost of the components as well as the modularity of the systems make it feasible to rebuild surveillance drones [8]. These drones can also be equipped with destructive weapons that may harm civilians. In summary, there is a need for autonomous systems to be able to detect

imminent failure and respond autonomously to minimise the impact.

This paper proposes a hacking game as a model of the interaction that could occur between an autonomous system (the victim) and a hacker [3], hactivist [4, 5], or state sponsored attacker [9] (the attacker). We use an intrinsically motivated game theoretic framework [10, 11] to capture the different motives that may lie behind different types of attackers and examine the responses of a rational autonomous agent to each type of attacker.

The remainder of this paper is organised as follows: Section II presents our game. Section III examines a number of simulations of different variations of this game and discusses the implications for the design of autonomous systems or policies concerning autonomous systems. We conclude in Section IV.

II. A Game-Theoretic Model for Decision Making by Compromised Autonomous Systems

To minimize the risks posed by autonomous systems when they are hacked, it is important that these systems have the ability to evaluate and choose responses rapidly when they are under threat. This implies that the system may need to make the decision autonomously. To address this issue, this section presents a game theoretic approach to modelling such decision making, assuming that machines may need to react to a range of differently motivated attackers.

In this study, we consider scenarios where an attacker tries to launch a cyber-attack on an autonomous system. As discussed above, such an attack may include actions such as stealing critical information, sending a virus, installing malicious software, taking control of the autonomous systems, and so on. In real life, some examples of these attacks include infecting drone plane command centre with a computer virus [12], intercepting unencrypted live video feed from military drones [13] and UAV hijacking [14].

The scenarios we use here were partly inspired by the evildoer game [15]. In the evildoer game, a hacker tries to cause damage to a victim's network. It is assumed that a primary attack has been launched initially by the attacker and has succeeded at the beginning. To detect the attack afterwards, the victim analyses the attack based on the available alerts. However different from the evildoer game, we focus our study where the victim is an autonomous system and thus we propose different scenarios for the attacker and the victim as described below.

We assume that an initial event X has already been performed by a human attacker and has been noticed, but not yet analysed, by the victim (the autonomous system). At this point, the autonomous system does not yet know if the attacker has launched their entire attack—which we refer to as the main or primary attack—or whether they will launch a secondary attack (e.g. a flood attack) to overwhelm the victim with both primary and secondary attacks.

Given that an initial event X has occurred, the victim has an option to repair only the main system that is being attacked, and let the rest of their subsystems work as usual, or shut down all their subsystems to prevent further propagation of the attack. The remainder of this section will present the game-theoretic model for this scenario. We will then examine decision making in this setting in Section III.

A. Modelling the Game

Previously we have reviewed and classified information warfare games [16]. However, we did not find existing games that focus on modelling conflict between an autonomous system and an attacker. Here, we propose a novel game as a starting point for modelling such conflict. The game that we propose is a dynamic game with complete information. The game has two types of players: A^1 as the attacker and A^2 as the victim. In this case the victim is the autonomous system and the attacker is human, who may have one of a varying set of motives.

1) Behaviors: Let B_t^α be the behavior of player A^α at time t. Each player has the choice of two behaviors. Player A^1 can choose between B^P (primary attack only) and B^F (flood attack). Player A^2 can choose between B^S (shut down all subsystems) and B^R (repair the main system).

A list of possible advantages (+) and disadvantages (-) of each behavior in this game is presented in Table I. This list is further used to define the payoff matrix and parameters of the game in the next section.

TABLE I. Advantages and Disadvantages of the Behaviors Available to Each Player.

		A^1 (attacker)	
		B^P	B^F
A^2 (victim)	B^S	For A^1: (+) Disrupt the main system (–) No further attacks For A^2: (+) Prevent propagation of attack (–) Complete deactivation of all subsystems	For A^1: (-) Cost of secondary attacks For A^2: (+) Prevent propagation of attack (–) Deactivation of subsystems that were not attacked
	B^R	For A^1: (+) Disrupt the main system (–) No further attacks For A^2: (+) Prevent propagation of attack (+) Learn about the attacker and how to repair. (–) Cost of repairs	For A^1: (+) Disrupt multiple subsystems (+) Make isolation of primary attack more difficult (-) Cost of secondary attacks For A^2: (–) Unprotected subsystems (–) Possibility of more damage

TABLE II. Payoff Matrix For The Human VS Machine Hacking Game.

		A^1 (attacker)	
		B^P	B^F
A^2 (victim)	B^S	$(0, -C_2)$	$(-C_1, 0)$
	B^R	(G, G)	$(V - C_1, -V)$

2) Payoffs and Parameters: Table II shows a general payoff matrix for this game. By considering the advantages and disadvantages of each behavior shown in Table I, we define the parameters for the game as follows:

- G: Payoff for launching the main attack for A^1 and for repairing the main system for A^2. We assume that A^1 gains from the attack, but no more or less than A^2 gains (learns) from the process of being attacked and executing a successful repair.
- V: Value for A^1 launching a secondary attack. In contrast, a negative value of this parameter captures the disadvantages for A^2 of being attacked by flood attacks.
- C_1: The cost for A^1 to launch secondary flood attacks.
- C_2: The cost for A^2 to shut down (deactivate) all their subsystems.

We denote the payoff to agent A^α at time t as U_t^α. The payoff to player A^1 is the first mentioned value in brackets in Table II. The payoff to player A^2 is the second mentioned value.

3) Assumptions: The following assumptions are used in this study:

- $G \geq 0$. Specifically, we consider two cases for the parameter G. In the first case, we consider $G > 0$. This captures the case where the attackers get benefit from launching the main attack on the autonomous systems. At the same time, the autonomous systems also gain the benefit from knowing their systems' vulnerability and possibly planning to retaliate against the attackers. In the second case, we consider that $G = 0$. This corresponds to the case where the attackers only launch an attack to reveal the vulnerability of the autonomous system, without planning on a large scale attack. In this case the victim only repairs the main system and does not plan to retaliate against the attacker. In such a case, both players receive payoff 0.
- $V, C_1, C_2 > 0$ I.e, values and costs are always positive.
- $V > C_1$ We assume that the benefit for launching flood attacks is always higher than the cost.

These assumptions and conditions imply three different variations that we can consider for this game:

- Scenario 1: $0 < G < V - C_1$. This scenario implies that a flood attack is more beneficial than the primary attack alone.
- Scenario 2: $0 < G$ and $G > V - C_1$. This scenario implies that a flood attack provides no explicit benefit over the primary attack alone (note that we will see that this does not mean some won't execute such an attack!).
- Scenario 3: $G = 0$. This scenario implies that there is no benefit or loss to either party as a result of the primary attack (but a secondary attack may still be beneficial to the attacker).

III. Simulating Game Outcomes Using Motivated Learning Agents

A. Motivated Learning Agents

The distinct feature in this study compared to other game theory approaches for cyber-security [16], is that we consider that the attacker's decision-making is influenced by where they sit on a spectrum of hacker [3], hactivist [4, 5], or state sponsored attacker [9], and how this might influence their decision making. We use computational motivation to model the differences in decision making.

Different motivations in humans have long been shown to have a significant impact on humans' behaviors in social dilemma games [17]. Power-motivated individuals, for example, are characterized by a tendency to favor conflict, a desire to control and influence others, and a preference for high incentive goals. This matches the concept where certain individuals, influenced by the culture of the age, tend to prefer competition (conflict). In contrast, affiliation-motivated agents are characterized by a tendency to belong to a group, to avoid conflict, and a preference for goals of lower payoff, to avoid competition with others. Recently, these concepts have been implemented in artificial agents, which allows the agents to make decisions based on both explicit payoff and their own motivation [10, 11]. In this paper we use these models to represent the variation that can occur in human decision making, so that we can examine the response of a rational autonomous agent in each case. We assume that the attackers are human decision makers and thus they may perceive the payoff of the game differently if they are hackers attacking 'for fun' or state sponsored attackers attacking as an act of war.In line with the motivational psychology theory, we adopt the concept that although individuals have both the needs for affiliation and power, each individual exhibits different characteristics that depends on their dominant motives. This means that one of the motives will have a stronger impact on decision-making. We thus assume that the motive profiles of the agents do not change during the game. In this study, we consider four types of attackers which may perceive the payoff of a game differently based on their motivation. Each type of attacker agent has different optimally motivating incentive (OMI), Ω. An OMI is used to compute an agent's subjective value $\hat{I}_t^\alpha$ fora given payoff U_t^α as follows [11]:

$$\hat{I}_t^\alpha = U^{max} - |U_t^\alpha - \Omega| \quad (1)$$

where U^{max} is the maximum payoff available in the game. By embedding the agents with different OMIs, the agents may subjectively perceive the same payoff differently. The four types of agents considered are:

- Pow(1): strong power-motivated agents. This is the most aggressive agent, with the highest preference for the greatest explicit payoff. We consider this the 'state-based attacker' end of the spectrum.
- Pow(2): weak power-motivated agents.
- Aff(1): weak affiliation-motivated agents; and
- Aff(2): strong affiliation motivated agents. This is the least aggressive agent, but has the greatest tendency to misperceive the payoff of the original game. This means that they may pursue a low explicit payoff, because it best satisfies their implicit motives. We consider this the 'hacker for fun' end of the spectrum.

We adapt the motivated learning algorithm proposed in [11] as summarized in Algorithm 1. The outcomes of the algorithm at each time step are the probabilities of choosing each behavior B^i, $P(B_t^\alpha = B^i)$.

Algorithm 1. A motivated learning agent, adapted from [11].

1. Initialize game world **W** of form in Table II, and identify U^{max}
2. Initialize agent's optimally motivating incentive Ω and initialize probabilities $P(B_0^\alpha = B^i)$ for all *i*.
3. Repeat:
4. If $t > 0$
5. Receive payoff U_t^α for previously executed behavior B_{t-1}
6. Compute subjective incentive $\hat{I}_t$ using Equation 1
7. Compute new probabilities $P(B_t^\alpha = B^i)$
8. Select next behavior B_t^α probabilistically
9. Store and execute B_t^α

B. Simulations

This section examines the interaction between the agents in the hacking game. For all scenarios, we use Algorithm 1 and examine thirty pairs of agents (A^1 (attacker) and A^2 (victim)) learning during the game. To allow the agents to sufficiently learn and converge to the equilibrium point of their perceived game, the learning rate in the simulations is set to 0.01 and the maximum iterations are set to 3,000. We consider iterations *t* here to represent the ongoing, moment-to-moment decision making processes by the attacker and victim. The results are presented in terms of changes in the probabilities $P(B_t^\alpha = B^i)$ for each agent. The horizontal axis on each result chart corresponds to player A^1 (the attacker) and measures $P(B_t^1 = B^F)$, while the vertical axis corresponds to player A^2 (the victim) and measures $P(B_t^2 = B^S)$. Each corner of the chart has a prescribed meaning as shown in Figure 1. For example, a point in the upper right of the corner indicates that outcome (B^F, B^S) occurs most frequently over time. Whereas a point in the bottom left of the corner indicates outcome (B^P, B^R). The lines in Figure 1 show the trajectory of the learned values of $P(B_t^1 = B^F)$ and $P(B_t^2 = B^S)$ over time. For the simulations, the payoff and OMI values are normalized within the range of [0, 1] to comply with the learning model to the replicator dynamics described by Borgers and Sarin [18]. At *t=0,* the probability that behavior B^i is executed is randomly chosen from a uniform distribution. As shown in Figure 1, we can see that the trajectories start at random positions. As time progresses, the trajectories move towards one or more of the corners. This indicates that the agents tend to approach an equilibrium but still changes their decisions as the learning process continues. In the example shown in Figure 1, we can see that the agents eventually converge on a stable strategy (B^P, B^R) (bottom left corner).

1) Scenario 1(a): We use the game in Table III as an example of a Scenario 1 game. According to the rules proposed by Merrick [11], we select OMIs for the human attacker agents as follows: $\Omega^{Pow(1)} = 0.95$; $\Omega^{Pow(2)} = 0.75$; $\Omega^{Aff(1)} = 0.30$; $\Omega^{Aff(2)} = 0.10$.

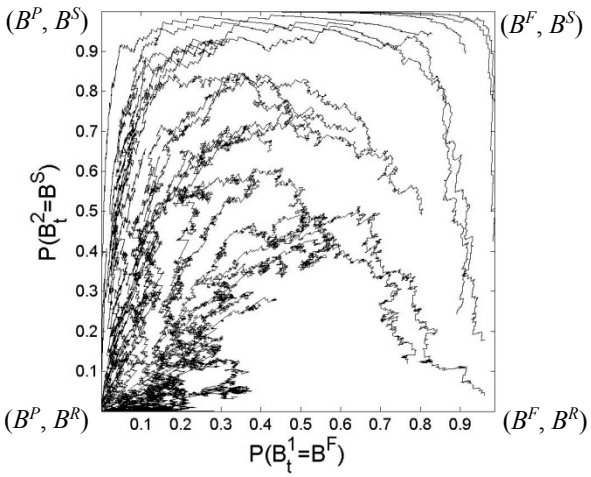

Fig. 1. Sample results with legend. Each corner of the graph represents a different equlibrium as labelled. Player 1 (x-axis) is attacker; player 2 (y-axis) is the victim. $P(B_t^\alpha = B^i)$ is the probability that behavior B^i is executed at time *t* by player α.

TABLE III. SCENARIO 1(A) PAYOFF MATRIX

		A^1 (attacker)	
A^2 (victim)		B^P	B^F
	B^S	(0, –0.3)	(–0.4, 0)
	B^R	(0.45, 0.45)	(0.5, –0.9)

2) Scenario 2: We use the game in Table IV as an example of a Scenario 2 game. The OMIs for the human attacker agents are: $\Omega^{Pow(1)} = 0.95$; $\Omega^{Pow(2)} = 0.85$; $\Omega^{Aff(1)} = 0.35$; $\Omega^{Aff(2)} = 0.10$.

TABLE IV. SCENARIO 2 PAYOFF MATRIX

		A^1 (attacker)	
A^2 (victim)		B^P	B^F
	B^S	(0, –0.6)	(–0.8, 0)
	B^R	(0.45, 0.45)	(0.1, –0.9)

3) Scenario 3: We use the game in Table V as an example of a Scenario 3 game. The OMIs for the human attacker agents are: $\Omega^{Pow(1)} = 0.90$; $\Omega^{Pow(2)} = 0.58$; $\Omega^{Aff(1)} = 0.08$; $\Omega^{Aff(2)} = 0.01$.

TABLE V. SCENARIO 3 PAYOFF MATRIX

		A^1 (attacker)	
A^2 (victim)		B^P	B^F
	B^S	(0, –0.1)	(–0.1, 0)
	B^R	(0, 0)	(0.5, –0.6)

4) Scenario 1(b): We use the game in Table VI as a second example of a Scenario 1 game, but this time consider decision making when both the attacker and the victim are assumed to be human. The 'victim' in this case is a policy maker who needs to either make policies in advance of an attack on an autonomous system, or respond to an attack on autonomous systems' infrastructure. We select OMIs for eight types of agents as follows:

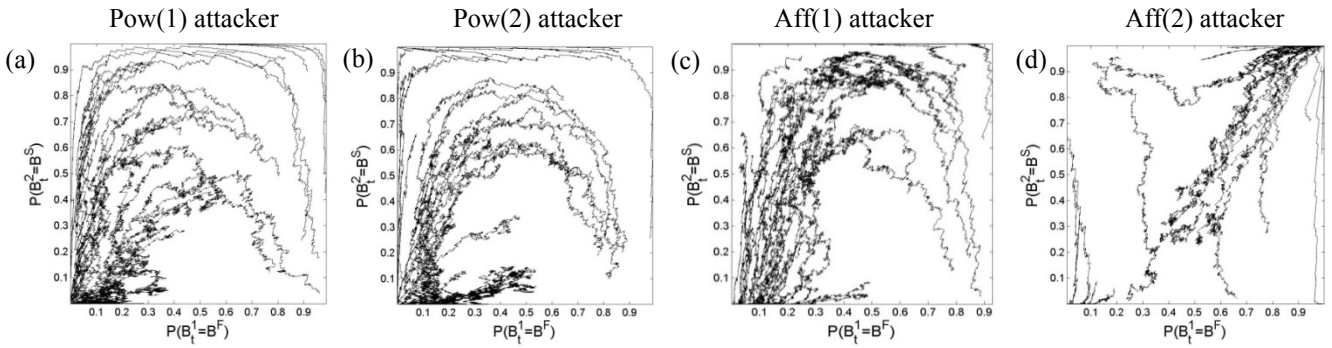

Fig. 2. Results for Scenarios 1(a) in which a rational autonomous agent is attacked by a (a) Pow(1) agent, (b) Pow(2) agent,(c) Aff(1) agent, (d) Aff(2) agent.

- For the human attacker: $\Omega^{\text{Pow(1)}} = 0.90$; $\Omega^{\text{Pow(2)}} = 0.70$; $\Omega^{\text{Aff(1)}} = 0.35$; $\Omega^{\text{Aff(2)}} = 0.05$.
- For human the policy maker: $\Omega^{\text{Pow(1)}} = 0.90$; $\Omega^{\text{Pow(2)}} = 0.40$; $\Omega^{\text{Aff(1)}} = 0.30$; $\Omega^{\text{Aff(2)}} = 0.01$.

We take the view that differences in the OMIs of different policy makers may be influenced by national *Zeitgeist*. *Zeitgeist* is defined as 'as spirit of the times' [19]. The effect of *Zeitgeist* on the policy maker is to influence their tendency to favour or avoid escalation of conflict.

TABLE VI. SCENARIO 1(B) PAYOFF MATRIX

		A^1 (attacker)	
		B^P	B^F
A^2 (victim)	B^S	(0, –0.9)	(–0.9, 0)
	B^R	(0.5, 0.5)	(0.1, –1)

C. Results

1) Scenario 1(a): The simulation results for Scenario 1 are shown in Figure 2. In this scenario, a flood attack is more beneficial than the primary attack alone; see Section II.This figure shows the learning trajectories of the autonomous system and different attackers as they engage in decision making under this condition.

Figure 2 shows that all of the different types of attackers will maintain an initially high probability of executing a flood attack (B^F) for a lengthy period of time. Note that the line trajectories maintain to be on the right side of the figure for some interations. Likewise, they will tend to maintain an initially high probability of executing only a primary attack (B^P). In all cases where the attacker exibits a tendency towards flood attacks, a rational autonomous agent will increase its preference for shutting down all of its subsystems (B^S). This is indicated by the upward trends of their trajectories in Figure 2. The effect on Pow(1), Pow(2) and Aff(1) attackers is then to reduce their preference for executing a flood attack. The response of a rational autonomous system is then to repair the main system.

These results suggest that the use of a rational strategy in an autonomous system produces two desirable features: first adaptation to a changing strategy from the attacker, and secondly sensible preferences for the (B^P, B^R) and (B^F, B^S) equilibria.

When the attackers are strong affiliation-motivated (Aff (2)), different behavior is observed; see Figure 2(d). For this case, the majority of the agents converge to (B^F, B^S) and some others converge to (B^P, B^R). This result shows that different types of attackers may change the equilibrium of the game.

2) Scenario 2: Figure 3 shows the simulation results for Scenario 2. This scenario implies that a flood attack provides no explicit benefit over the primary attack alone.

Figure 3 shows that, like in Scenario 1(a), all of the different types of attackers will maintain their initial preferences for either executing a flood attack (B^F) or executing only the primary attack (B^P) for a lengthy period of time, until the response of the victim becomes clear. However, unlike Scenario 1 we now see more difference in the behavior of attackers with different motives. Aff(1) and Aff(2) attackers (our hactivists and 'hackers for fun') may prefer to execute a flood attack, even though there is no explicit benefit of doing so. This is indicated by their eventual convergence in the upper right corner of the figure.

As we saw previously, in all cases where the attacker displays a preference for flood attack, a rational autonomous agent will increase its preference for shutting down all of its subsystems (B^S). This corresponds to an initial upward trend of the graph. The effect on Pow(1) and Pow(2) attackers is to reduce their preference for executing a flood attack. The response of a rational autonomous system is then to repair the main system.

In the case of Aff(1) and Aff(2) attackers, we see that there is a tendency for them to adapt their B^P strategy if they perceive the victim is only repairing the main system. However, we see that a rational autonomous system can respond to this change forcing the emergence of the (B^F, B^S) equilibrium.

3) Scenario 3: This scenario implies that there is no benefit or loss to either party as a result of the primary attack, but a secondary attack may be beneficial to the attacker. Results are shown in Figure 4.

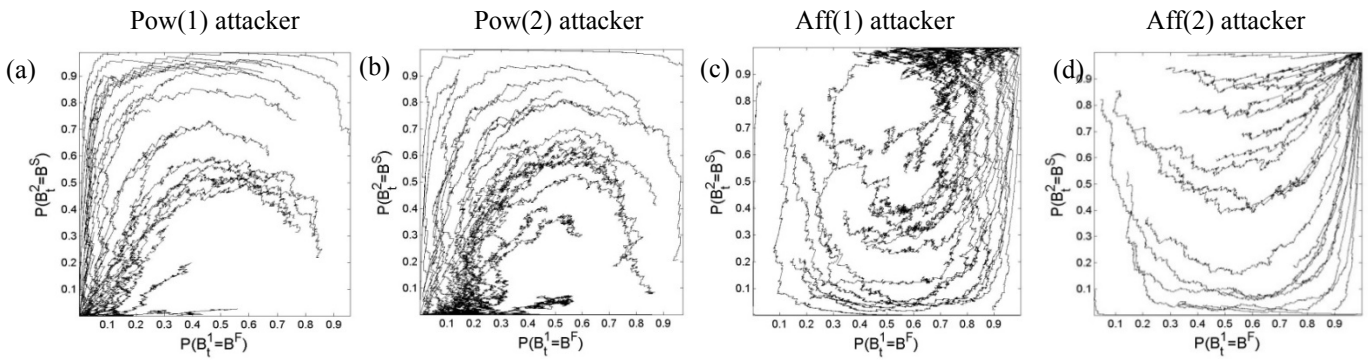

Fig. 3. Results for Scenarios 2 in which a rational autonomous agent is attacked by a (a) Pow(1) agent, (b) Pow(2) agent,(c) Aff(1) agent, (d) Aff(2) agent.

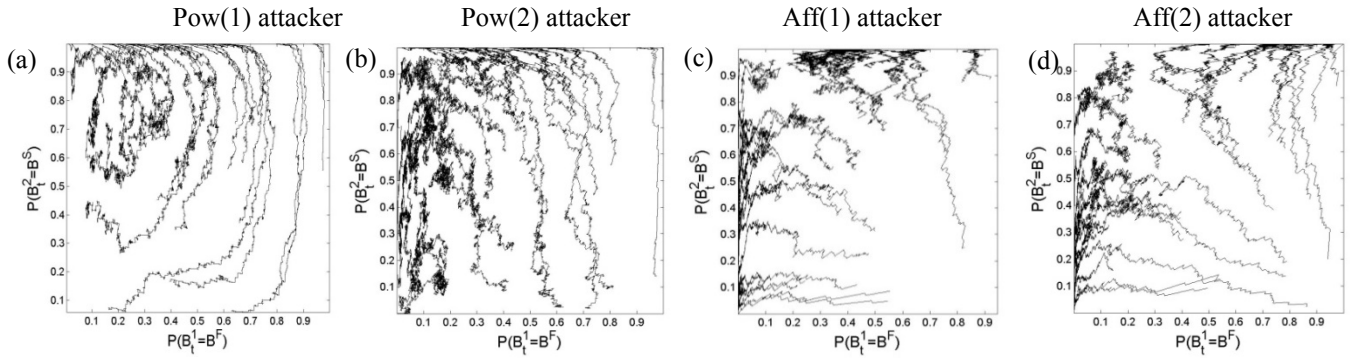

Fig. 4. Results for Scenarios 3 in which a rational autonomous agent is attacked by a (a) Pow(1) agent, (b) Pow(2) agent,(c) Aff(1) agent, (d) Aff(2) agent.

In this experiment, we see that rational agents become more defensive, with a greater tendency to shut down the entire system. This is because there is no value from the primary attack, meaning that a flood attack may be imminent. Aff(1) and Aff(2) agents develop a preference for the flood attack in roughly half of the thirty trials, with their final preference being determined by their initial probability of doing so. Pow(1) and Pow(2) agents (representing our state based attackers) tend to have much less variation in their probablity of executing a flood attack, suggestive of adherence to a predefined plan.

4) Scenario 1(b): In this experiement, we return to a Scenario 1 setup, but consider humans making policies about autonomous systems, rather than direct responses by an autonomous system. We will now assume that the policy maker (victim) is human and the attack more threatening (i.e. beyond a single system, and having potential ramifications at a national level). To model a more threatening situation, we assign greater values for *V, G,* C_1 and C_2 compared to the three previous scenarios described above (see Table VI). To model decision making by human policy makers, we examine motivated agents along both axes in Figure 5.

For the case where both the attackers and the human policy maker are strong power-motivated agents (Figure 5(a)), we can see that when the attackers have an initial preference for launching only the main attack (B^P), the policy maker responds by shutting the systems down. However, this leads to a loss $-C_2$ for the policy maker, requiring a change to behavior B^R. This leads to an eventual convergence to the (B^P, B^R) equilibrium.

This result suggests that the policy maker should prefer to play aggressively when dealing with attackers that are aggressive.

When the attackers are weak power-motivated, an ongoing cycle emerges between the four pure strategy equilibria as shown in Figure 5(b). This result indicates that the policy maker must continue to adapt to the actions of the attacker.

Now suppose the attackers are weak affiliation-motivated, then different equilibria emerge. Figure 5(c) shows that these agents are initially aggressive by choosing behavior (B^P, B^R) leading to the downward moving of the learning trajectory. However, the attackers learn that the policy maker tends to repair only the main system, and thus, the attackers change their strategy to launch flood attacks B^F. This corresponds to the learning trajectory that moves from the left to the right of the graph. Realizing that the opponent has changed their behavior to launch flood attacks, the policy maker adapts by playing behavior B^S to avoid greater damage to the autonomous systems. Eventually, the agents converge on the top right corner of the graph. Similar behavior is observed when the attackers are strong affiliation-motivated as shown in Figure 5(d).

We will now discuss the case where the policy maker is weak power-motivated. Under the case when the opponents are strong power-motivated, we can see in Figure 5(e) that the learning trajectories move directly to (B^P, B^S). This represents a defensive policy.

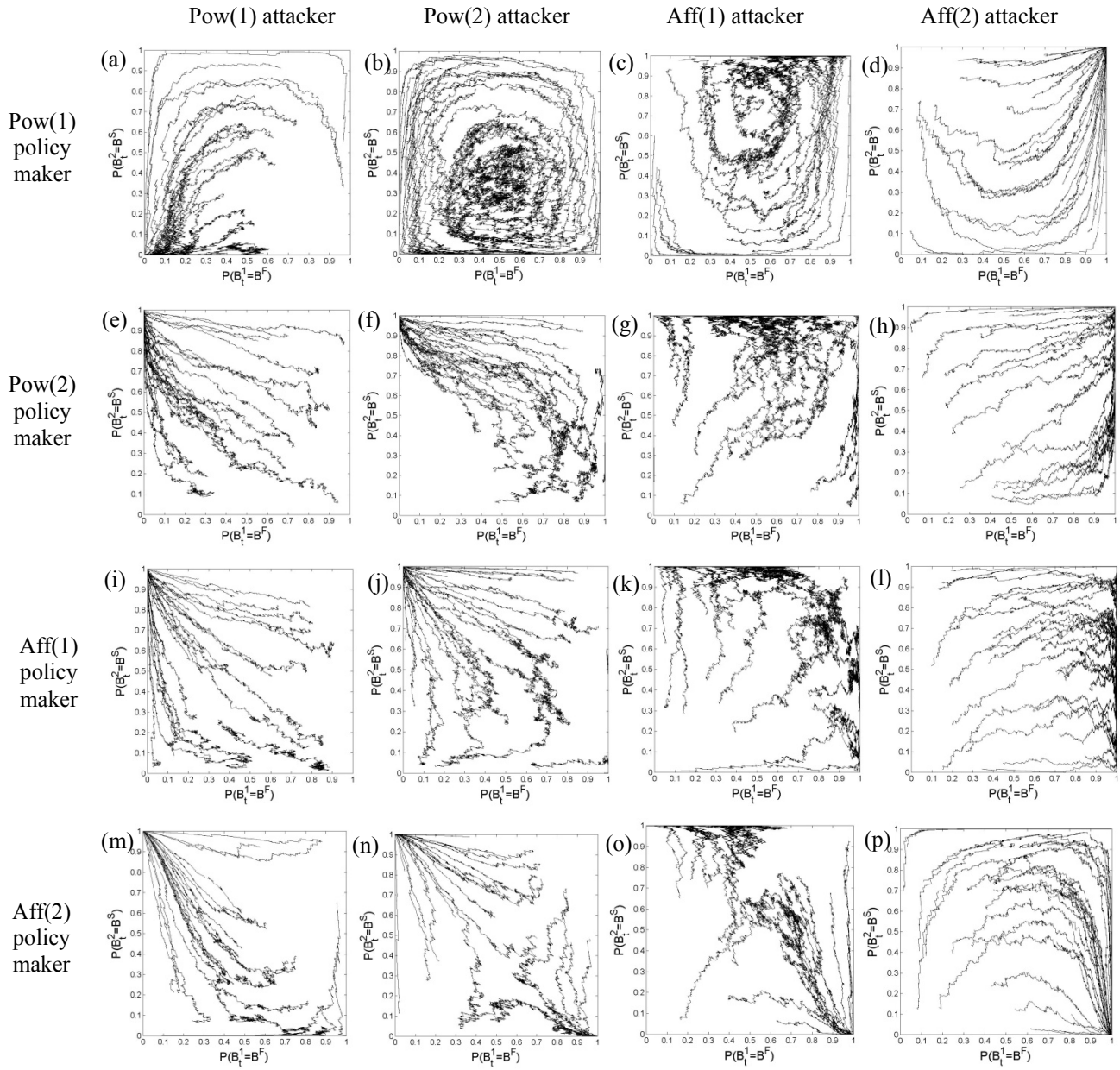

Fig. 5. Scenario 1(b) results for (a-d) Pow(1) policy maker; (e-h) Pow(2) policy maker; (i-l) Aff(1) policy maker; (m-p) Aff(2) policy maker.

When the attackers are weak power-motivated, we can see in Figure 5(f) that initially the agents show quite similar behavior to the case when the attackers are strong power-motivated. There is, however, a slight difference in that the learning trajectories initially move to point (B^F, B^S), but then converge on the upper left corner of the graph (B^P, B^S). In the case where the attackers are weak power-motivated, the policy maker prefers to behave defensively by shutting the systems down. When dealing with attackers that are weak affiliation-motivated quite similar behavior is observed; see Figure 5(g). The agents converge on different equilibrium when the attackers are strong affiliation-motivated. This can be seen in Figure 5(h) where the agents now converge on (B^F, B^S).

Next we consider what happens if the policy maker is weak affiliation-motivated. When the attackers is strong power-motivated and the center is weak-affiliation motivated, we can see from Figure 5(i) that the policy maker tends to behave defensively by selecting behavior B^S while the attackers tend to launch the main attack. This corresponds to the eventual convergence to (B^P, B^S). Now suppose the attackers are weak power-motivated, we can see that the majority of the agents still maintain a preference for behavior (B^P, B^S). Some others, however, choose to converge on the bottom right corner of the graph as shown in Figure 5(j). We observed similar behavior when the attackers are weak affiliation-motivated as shown in Figure 5(k). When the attackers are strong affiliation-motivated, they tend to

maintain a high preference for flood attacks as shown in Figure 5(l). Notice that the learning trajectories move from left to right of the graph. In this case, the agents eventually end up along the line between (B^F, B^S) and (B^F, B^R).

We will now observe the case where the policy maker is strong affiliation-motivated. Under this case, when the attackers are strong power-motivated, the agents maintain a preference for behavior (B^P, B^S) as can be seen in Figure 5(m). Now when the attackers are weak power-motivated, the agents will converge on either (B^P, B^S) or (B^F, B^R) outcome; see Figure 5(n). Suppose the attackers are weak affiliation-motivated quite similar behavior can be observed. There is however a small difference in that the learning trajectories now tend to slightly move to the upper right corner of the graph; see Figure 5(o). In the case where the attackers are strong affiliation-motivated, the agents initially try to reach equilibrium (B^P, B^S). The attackers then adapt by changing their strategy to launch flood attacks. Notice that the learning trajectories move from left to right of the graph. The agents eventually reach an equilibrium point on the bottom right corner of the graph; see Figure 5(p).

From the above discussion we can see that agents that are embedded with different motive profiles may perceive the game differently. This permits us to model attackers with different motives, as well as the effects of different policies. This permits the policy maker to evaluate a range of possible actions which can be useful to establish anti-hacking policies for autonomous systems.

IV. Conclusion

This paper has presented a game theoretic model for autonomous systems that become victims of hacking. To model individual differences in decision-making by attackers, we used computational models of motivations which allow some of the agents (players) to choose the decisions based on both the explicit payoff of the game and their implicit motives.

Simulations were conducted to examine the behavior of the agents during our hacking game. Results show that autonomous systems that take a rational approach to decision making can respond sensibly to decision making by attackers with different motives, even when these motives mean that the attacker acts in a way that is 'objective irrational'.

We also examined how game theory can be used by a cyber-security policy maker to evaluate a range of policy options for national level threats.

In future work we will examine the ability of these findings to be transferred to other games representing alternative scenarios that may be encountered by future autonomous systems.

Acknowledgment

This work was part of a project funded by the Australian Centre for Cyber Security.

References

[1] N. Winton, "Autonomous cars like the google may be vable in less than 10 years," http://www.forbes.com/sites/neilwinton/2014/06/06/autonomous-cars-like-the-google-may-be-viable-in-less-than-10-years, vol. (accessed December 17, 2015), 2014.

[2] J. Jay, "Autonomous military technology: opportunities and challenges for policy and law," http://www.heritage.org/research/reports/2014/08/autonomous-military-technology-opportunities-and-challenges-for-policy-and-law, vol. (accessed December 17, 2015), 2014.

[3] D. Ventre, *Information Warfare*: John Wiley and Sons, 2012.

[4] P. William, "Information warfare: time for a redefinition," presented at the Eleventh Australian Informatino Warfare and Security Conference, Perth, WA, 2010.

[5] K. Hearn, R. Mahncke, and P. William, "Culture jamming: from activism to hactivism," presented at the Tenth Australian Information Warfare and Security Conference 2009.

[6] C. Grut, "The challenge of autonomous lethal robotics in international humanitarian law," *Journal of Conflict and Security Law,* vol. 18, pp. 5-23, 2013.

[7] P. Meilinger, "The mutable nature of ware," *Air and Space Power Journal,* vol. 24, pp. 24-30, 2010.

[8] J. Weber, "Robotic warfare, human rightrs and the rhetorics of ethical machines," *Ethics and Robotics,* pp. 83-103, 2009.

[9] M. Robinson, K. Jones, and H. Janicke, "Cyber warfare: issues and challenges," *Computers and Security,* vol. 49, pp. 70-94, 2015.

[10] K. Merrick and K. Shafi, "A game theoretic framework for incentive-based models of intrinsic motivation in artificial systems," *Frontiers in Psychology - Cognitive Science, Special Issue on Intrinsic Motivations and Open-Ended Development in Animals, Humans and Robots,* vol. 4, 2013.

[11] K. Merrick, "The role of implicit motives in strategic decision making: computational models of motivated learning and the evolution of motivated agents " *GAMES,* vol. 6, pp. 604-636, 2015.

[12] Associated Press, "Computer virus infects drone plane command center in US", Guardian, 9 October 2011.

[13] C. Arthur, "SkyGrabber: the $26 software used by insurgents to hack into US drones", Guardian, 17 December 2009.

[14] A.Rawnsley, "Iran's Alleged Drone Hack: Tough, but Possible.", Wired, December 2011.

[15] J. Jormakka and J.V.E. Molsa," Modelling information warfare as a game,"*Journal of information warfare*, *4*(2), pp.12-25, 2005.

[16] K. Merrick, M. Hardhienata, J. Hu, and K. Shafi, "A survey of game theoretic approaches to modelling decision making in information warfare scenarios," *Future Internet* vol. 8(3):34, 2016.

[17] K. W. Terhune, "Motives, situation and interpersonal conflict within prisoner's dilemma," *Journal of Personality and Social Psychology, Monograph Supplement,* vol. 8, pp. 1-24, 1968.

[18] T. Borgers and R. Sarin, "Learning through reinforcement and replicator dynamics," *Journal of Economic Theory,* vol. 77, pp. 1-14, 1997.

[19] J. Heckhausen and H. Heckhausen, *Motivation and action*. New York: Cambridge University Press, 2010.