Proceedings of the 2004 IEEE
Workshop on Information Assurance and Security
United States Military Academy, West Point, NY, 10-11 June 2004

A Game Theoretic Approach to Modeling Intrusion Detection in Mobile Ad Hoc Networks

Animesh Patcha[†], *Student Member IEEE* and Jung-Min Park[†], *Member IEEE*
{apatcha, jungmin}@vt.edu

Abstract— **Nodes in a mobile ad hoc network need to come up with counter measures against malicious activity. This is more true for the ad hoc environment where there is a total lack of centralized or third party authentication and security architectures. This paper presents a game-theoretic method to analyze intrusion detection in mobile ad hoc networks. We use game theory to model the interactions between the nodes of an ad hoc network. We view the interaction between an attacker and an individual node as a two player non-cooperative game, and construct models for such a game.**

I. Introduction

Mobile Ad hoc Networks (MANET's) are autonomous distributed systems that comprise a number of mobile nodes connected by wireless links, forming arbitrary time-varying wireless network topologies. Mobile nodes function both as hosts and routers. As hosts, they represent source and destination nodes in the network while as routers, they represent intermediate nodes between a source and a destination providing store-and-forward services to neighboring nodes. Store-and-forward services are needed due to the limited range of each individual mobile host's wireless transmissions. Nodes that constitute the wireless network infrastructure are free to move randomly and organize themselves arbitrarily. Applications such as military exercises and disaster relief will benefit from ad hoc networking, but secure and reliable communication is a necessary prerequisite for such applications.

Security in mobile ad-hoc networks is particularly difficult to achieve, notably because of the limited physical protection to each of the nodes, the sporadic nature of connectivity, the absence of a certification authority, and the lack of a centralized monitoring or management unit. Intrusion prevention is not guaranteed to work all the time, and this clearly underscores the need for intrusion detection as a frontline security research area under the umbrella of ad hoc network security. In traditional wireless networks, mobile devices associate themselves with an access point which is in turn connected to other wired machines such as a gateway or a name server which handle network management functions. Ad-hoc networks, on the other hand, do not use such access points, and form a completely distributed architecture. The absence of infrastructure and subsequently, the absence of authorization facilities impede the usual practice of establishing a line of defense—distinguishing nodes as trusted and non-trusted. There may be no ground for an *a priori* classification, since all nodes are required to cooperate in supporting the network operation, while no prior security association (SA) can be assumed for all the network nodes. Freely roaming nodes form transient associations with their neighbors: they join and leave sub-domains independently with and without notice.

An additional problem related to compromised nodes is the potential Byzantine failures encountered within MANET routing protocols. In a Byzantine failure, a set of the nodes could be compromised in such a way that incorrect and malicious behavior cannot be directly noted at all. Malicious nodes can inflict a Byzantine failure on a system by creating new routing messages, advertising non-existent links and providing incorrect link state information. It can therefore be seen that intrusion prevention measures in MANET's might prevent some attacks but such measures are not enough.

Intrusion detection (aptly called "the second line of defense") techniques are widely used in wired networks to protect networked systems once an intrusion is detected. However, techniques geared towards wired networks would not suffice for an ad hoc environment because of differences such as lack of fixed infrastructure, mobility, the vulnerability of wireless transmissions to eavesdropping and the lack of a clear separation between normal and abnormal behavior in ad hoc networks. In addition, the ad hoc networking paradigm does not allow for the presence of traffic concentration points in the network, whereas most conventional intrusion detection systems (IDS) geared towards wired networks depend on such an architecture.

In this paper, we concentrate on providing a mathematical framework for intrusion detection in mobile ad hoc networks. We will describe how game theory can be used to find strategies for both the malicious node and the administrator of the target node.

The organization of the paper is as follows. In section 2

[†] Bradley Department of Electrical and Computer Engineering, Virginia Polytechnic Institute and State University, Blacksburg VA 24061

0-7803-8572-1/04/$20.00 ©2004 IEEE

we briefly describe some of the related work that has been done in this area. Section 3 introduces the formal model for non-cooperative games and relates the elements in the model to the problem at hand. In Section 4 we present our model for intrusion detection in a MANET. Section 5 concludes the paper

2. Related Work

Game theory has been used extensively in computer and communication networks to model a variety of problems. This includes the work of Shenker for modeling service disciplines [1], Akella et al. for TCP performance [2],Baser et al [3] for modeling power control in a multicell wireless network while Bencsáth et al. [4] applied game theory and client puzzles in order to devise a defense against denial of service (DoS) attacks. In the area of MANET's, Michiardi et al. [5] used cooperative and non-cooperative game theoretic constructs to develop a reputation based architecture for enforcing co-operation. Modeling intrusion detection using game theory is ,however, a relatively new approach. Kodialam et al. [6] used a game theoretic framework to model intrusion detection via sampling in communications networks, and developed sampling schemes that are optimal in the game theoretic setting. Our work is more closely related to the model proposed by Alpcan et al. [7]. We have extended the model proposed in [7] to include MANET's, and have analyzed the interaction between an attacker and a host based IDS as a dynamic two player non-cooperative game.

3. Modeling Intrusion Detection in Mobile Ad Hoc Networks as Non-Cooperative Games

Game theory provides us with tools to study the interaction between multiple players in a society, in our case a MANET. With the application of game theory, we can address problems where players/nodes with different, and in many cases, unique objectives interact. A specific branch of game theory that has been often used is non-cooperative game theory. Non-cooperative game theory studies situations in which a number of nodes/players are involved in an interactive process whose outcome is determined by the node's individual decisions and affects the well-being of each agent in a possibly different way.

In this paper, we model the interaction between an attacker and an intrusion detection system as a *basic signaling game* which falls under the gambit of *multi-stage dynamic non-cooperative game with incomplete information.*

A game can be defined as a *dynamic game* if the order in which the decisions are made is important. In a non-cooperative game *with incomplete information*, we model situations in which some players have some private information before the beginning of a game. This initial private information is called the *type* of the player and it fully describes any information the player has, which is not common knowledge. A player may have several types, one for each possible state of his/her private information. It is also assumed that each player knows his/her own type with complete certainty.

Basic Signaling Game [8]: A basic signaling game, in its simplest form has two players, Player 1 who is the sender and Player 2 who is the receiver. For sake of convenience we treat Player 1 as masculine and Player 2 as feminine. Nature draws the type of the sender from a type set Θ, whose typical elements are θ. The type information is private to each sender. Player 1 observes information about his type θ and chooses an action a_1 from his action space A_1. Player 2, whose type is known to everyone observes a_1 and chooses an action a_2 from her action space A_2. Player 2 on the other hand has prior beliefs , before the start of the game, about Player 1's type. In other words, before observing the senders message, the receiver believes that the probability that the sender is some type $\theta \in \Theta$ is $p(\theta)$. The spaces of mixed actions are A_1 and A_2 with elements α_1 and α_2 respectively.

Player i's payoff is denoted by $u_i(\alpha_1, \alpha_2, \theta)$. Player 1's strategy is a probability distribution $\sigma_1(\cdot|\theta)$ over actions a_1 for each type θ . A strategy for Player 2 is a probability distribution $\sigma_2(\cdot|\alpha_1)$ over actions a_2 for each action a_1

After both the players have taken their actions, the payoffs are awarded according to the *message* sent by the sender, the action taken by the receiver in response and the type θ of the sender chosen by Nature.

Type θ's payoff to strategy $\sigma_1(\cdot|\theta)$ when Player 2 plays $\sigma_2(\cdot|a_1)$ is

$$u_1(\sigma_1, \sigma_2, \theta) = \sum_{a_1} \sum_{a_2} \sigma_1(a_1|\theta)\sigma_2(a_2|a_1)u_1(a_1, a_2, \theta). \quad (1)$$

Player 2's payoff to strategy $\sigma_2(\cdot|a_1)$ when Player 1 plays $\sigma_1(\cdot|\theta)$ is

$$\sum_{\theta} p(\theta)(\sum_{a_1} \sum_{a_2} \sigma_1(a_1|\theta)\sigma_2(a_2|a_1)u_2(a_1, a_2, \theta)). \quad (2)$$

Player 2 updates her beliefs about θ and bases his choice of action a_2 on the posterior distribution[1] $\mu(.|a_1)$ over Θ. Bayesian Equilibrium dictates that Player 1's action will depend on his type. Therefore, if $\sigma_1^*(\cdot|\theta)$ denotes this strategy, then knowing $\sigma_1^*(\cdot|\theta)$ and by observing a_1, Player 2 can use Bayes rule to update $p(\cdot)$ and $\mu(\cdot|a_1)$

Definition:*A perfect Bayesian equilibrium(PBE) of a signaling game is a strategy profile σ^* and posterior beliefs $\mu(\cdot|a_1)$ such that*

$$\text{(P1)} \quad \forall\theta, \sigma_1^*(\cdot|\theta) \in \arg\max_{\alpha_1} u_1(\alpha_1, \alpha_2, \theta), \quad (3)$$

[1]In Bayesian inference, when we have performed an experiment we use Bayes' theorem to find a new distribution which reflects the result of the experiment. This new distribution is called the posterior distribution

$$\text{(P2)} \quad \forall a_1, \sigma_2^*(\cdot|a_1) \in \arg\max_{\alpha_2} \sum_{\theta} \mu(\theta|a_1) u_2(a_1, \alpha_2, \theta), \tag{4}$$

and

$$\text{(B)} \quad \mu(\theta|a_1) = p(\theta)\sigma_1^*(a_1|\theta) / \sum_{\theta' \in \Theta} p(\theta')\sigma_1^*(a_1|\theta') \tag{5}$$

$$\text{if} \quad \sum_{\theta' \in \Theta} p(\theta')\sigma_1^*(a_1|\theta') > 0$$

and $\mu(\cdot|a_1)$ is any probability distribution on Θ

$$\text{if} \quad \sum_{\theta' \in \Theta} p(\theta')\sigma_1^*(a_1|\theta') = 0$$

where P_1 says that Player 1 takes into account the affect of a_1 on Player 2's action. P_2 states that Player 2 reacts optimally to Player 1's action given her posterior beliefs about θ and B corresponds to the application of Bayes rule.

Thus, a perfect bayesian equilibrium can be thought of as a set of strategies and beliefs such that at any stage of the game, strategies are optimal given the beliefs. These beliefs are obtained from the equilibrium strategies and the observed actions using Bayes rule.

We believe that intrusion detection in MANET's can be modeled as a basic signaling game for a number of reasons. First and foremost, in a MANET environment, it is very hard to detect a friend from a foe in the absence of security mechanisms like PKI, digital certificates, etc. Therefore, the *type* of a particular node is not easily verifiable by other nodes in the system. Secondly, in most intrusion detection systems, both for wired and wireless networks, the IDS responds to the intrusion after the intrusion has occurred. Therefore, we believe that modeling intrusion detection in a game theoretic framework based on dynamic non-cooperative games is the right direction to take.

4. Game Theoretic Model of Intrusion Detection

The very nature of MANET's, dictates that any IDS designed for such a network has to be distributed in nature. Centralized solutions that have a single point of failure cannot be used. Assuming a host based IDS, we model an intrusion detection game played between a host and an intruder.

With every passing day the popularity and the number of users of the Internet is increasing manyfold. As a result of this increased popularity and usefulness, the Internet has come to contain both interesting targets and enough malicious and ignorant users that DoS attacks are only bound to increase. Current intrusion detection techniques, however, fail to detect malicious activity completely. This is more true for MANET's than for conventional networks.

In this section we present our game theoretic framework to analyze and model the response of an IDS. The IDS response actions typically vary from setting off an alarm to watching suspicious activity before setting off an alarm to a total system reconfiguration.

We model the interaction between an attacker and a host based IDS as a two player signaling game which falls under the gambit of multi-stage dynamic non-cooperative game with incomplete information.

The intrusion detection game is played between an attacker and an IDS. The objective of the attacker is to send a malicious message from some attack node, with the intension of attacking the target node. The intrusion is deemed successful when the malicious message reaches the target machine without being detected by the host IDS. We assume that an intrusion is detected and the intruding node is blocked when a message sent by a probable intruder is intercepted and the host IDS can say with certainty that the message is malicious in nature.

For an IDS, the basic performance criteria is the rate of false alarms in the system. There exists a tradeoff between the reduction in false alarms by decreasing the sensitivity of the system with the increase in rate of undetected intrusions. However, either extremes are undesirable as the IDS becomes totally ineffective in such circumstances. In our system model, we consider the cost associated with an undetected intrusion to be much more severe than the cost associated with false alarms.

To simplify our analysis, we assume that a *malicious node* attacks only one *node* at a time and that collusions between malicious nodes do not occur. The IDS does one of two things, it either sets off an alarm on detection of an intrusion or does nothing.

System Model: In our model of the signaling game, a node is the *sender* and the host based IDS based is the *receiver* to which the message is directed. A sender can be of two *types*. The sender node could be a *regular node* or he could be a *malicious node/attacker*. In other words, the type space of a given sender is given by $\Theta = [Attacker, RegularNode]$. The IDS prior beliefs concerning the probability that any other node in the system is either an *attacker* or a *regular node* can be described by a single number $q \in [0, 1]$. The malicious node's *(attacker)* decision is a choice between exhibiting malicious behavior or exhibiting normal behavior. Let the probability of a particular malicious node exhibiting malicious activity be s, and the probability of the same node exhibiting normal behavior be $1 - s$. The particular choice that the attacker makes is his "message". The IDS "detects" this decision with a probability t and misses it with a probability $1 - t$ depending on his beliefs.

Consider the attacker-IDS game shown in Fig 1. The IDS has a gain of $-\gamma_{success}$ for detecting an attack where as there is a cost involved whenever the IDS misses an attack (γ_{miss}) or when it raises a false alarm (γ_{falarm}). On the other hand, the intruder has a gain of $-\delta_{intrude}$ on a

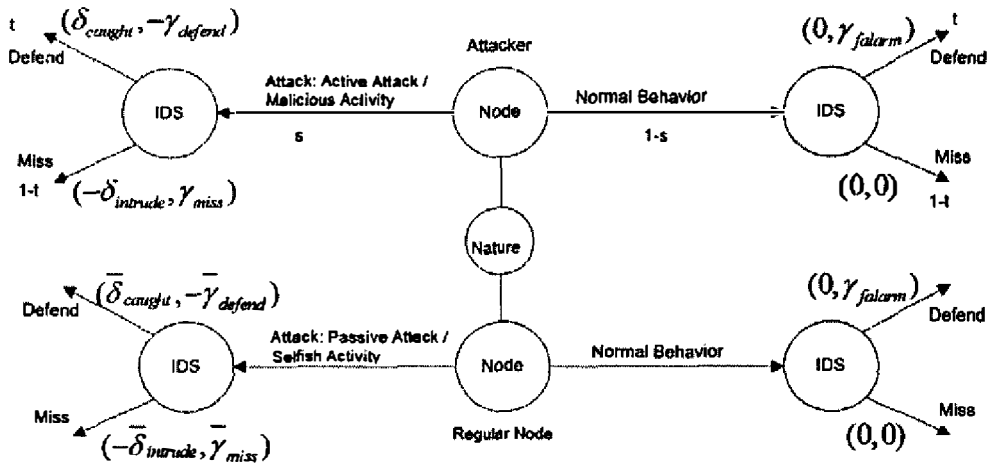

Fig. 1. An Attacker-IDS Basic Signaling Game

successful undetected intrusion and a cost of δ_{caught} on being detected and blocked. False alarms have a zero cost value to the attacker.

For the attacker, in all possible cases, the expected payoff is

$$s[t\delta_{caught} - (1-t)\delta_{intrude}] \tag{6}$$

Similarly, for the IDS in all possible cases the expected payoff is

$$s\gamma_{miss} + t\gamma_{falarm} - st(\gamma_{detect} + \gamma_{falarm} + \gamma_{miss}) \tag{7}$$

A rational node will always try to maximize (7). If the cost of false alarm (γ_{falarm}) is relatively low, then the IDS will always choose to sound an alarm. The Nash Equilibrium for such a signaling game is described by the following condition.

For the attacker/ regular node: Given the strategy of the IDS, each type θ of a node evaluates the utility from sending a message a_1 as $\sum_{a_2} \sigma_2(a_2|a_1)u_1(a_1, a_2, \theta)$ and $p(\theta)$ puts weight on a_1 only if its amongst the maximizing messages in this expected utility.

For the IDS: The IDS will proceed in two steps. First for every message a_1 that is sent with positive probability by some type θ, the IDS uses Bayes rule to compute the posterior probability assessment that a_1 comes from each type θ . According to the Nash equilibrium condition, for all a_1 that are sent by some type θ with positive probability, every response a_2 in support of the IDS's response should be the best response to a_1 given the beliefs that are computed using Bayes rule.

Therefore, we can say that the IDS strategy will be the best response to the sending nodes behavior strategy if and only if it maximizes its expected utility over all possible pure strategies.

$$\forall\theta, \sigma_1^*(.|\theta) \in \arg\max_{\alpha_1} u_1(\alpha_1, \sigma_2^*, \theta), \tag{8}$$

In other words, the strategy of the IDS will be to pick the optimal strategy out of its available set, in response to a message a_1 from the sending node. The choice of strategy must be based on the receivers prior beliefs, such that it is able to maximize the effective payoff by minimizing the cost due to false alarms and missed attacks.

Baye's theorem being recursive in nature, allows each node to periodically update its posterior beliefs about other nodes from its previous posterior distribution based on independent observations. Intuitively, we can see that with time the false alarm rates will decrease. When applied in tandem with other approaches like likelihood evaluation and active intruder profiling, the false alarm rates can be further reduced. A full explanation of these methods is beyond the scope of this paper.

The game theoretic investigation presented in this paper gives us valuable insight into the behavior of the attacker and the IDS. We believe that most of the simplifying assumptions made in this paper can be modified to incorporate more realistic scenarios.

5. Conclusions and Research Issues

Ad hoc network security has come into the lime light of network security research over the past couple of years. However, little has been done in terms of defining the security requirements specific to MANET's. Such security requirements must include countermeasures against node misbehavior in general and denial of service attacks in particular. In this paper, we used the concepts of multi-stage dynamic non-cooperative game with incomplete information to model intrusion detection in a network that uses a host based IDS. We believe that this game-theoretic mod-

eling is more realistic than previous modeling techniques. As part of our future work, we intend to extend our game theoretic approach to include selfish nodes[2].

REFERENCES

[1] S. Shenker, "Making greed work in networks: A game-theoretic analysis of switch service disciplines," in *IEEE/ACM Transactions on Networking*, 1995.

[2] A. Akella, R. Karp, C. Papadimitriou, S. Seshan, and S. Shenker, "Selfish behavior and the stability of the internet: A game theoretic analysis of tcp," in *Proceedings of SIGCOMM 2002*,, 2002.

[3] T. Alpcan, T. Basar, and S. Dey, "A power control game based on outage probabilities for multicell wireless data networks," in *Proc. of American Control Conference (ACC)*, July 2004.

[4] B. Bencsth, L. Buttyn, and I. Vajda, "A game based analysis of the client puzzle approach to defend against dos attacks," in *SoftCOM 2003 11th International conference on software, telecommunications and computer networks*, pp. 763-767, 2003.

[5] P. Michiardi and R. Molva, "Core: A collaborative reputation mechanism to enforce node co-operation in mobile ad hoc networks," in *6th IFIP Communications and Multimedia Security Conference*, September 2002.

[6] M. Kodialam and T. V. Lakshman, "Detecting network intrusions via sampling : A game theoretic approach," in *IEEE INFOCOMM 2003*, April 2003.

[7] T. Alpacan and T. Basar, "A game theoretic approach to decision and analysis in network intrusion detection," in *42nd IEEE Conference on Decision and Control, Maui*, 2003.

[8] D. Fudenberg and J. Tirole, *Game Theory*. Cambridge, MA: MIT Press, 2002.

[9] S. Axelsson, "The base-rate fallacy and its implications for the difficulty of intrusion detection," in *Proceedings of the 6th ACM conference on Computer and communications security,Singapore*, 1999.

[2]Selfish nodes are nodes that use the network but do not cooperate in an attempt to save battery life for their own communications. They do not intend to directly damage other nodes