ARSoS: an Adaptive, Robust and Sub-optimal Strategy for Automated Deployment of Anomaly Detection System in MANETs

Zonghua Zhang, Farid Naït-Abdesselam, and Soufiene Djahel
IRCICA/LIFL - CNRS UMR 8022 - INRIA
University of Sciences and Technologies of Lille, France

***Abstract*—While a variety of AIDS (Anomaly-based Intrusion Detection system) are claimed to be fully distributed, light-weight, and ready for application, their detection cost are not always neglectable, especially considering the fact that MANET nodes have scarce resource, which usually impels the nodes to save their energy from any unnecessary action. It is therefore a significant issue to deploy AIDS sensors in an optimal way for achieving the best tradeoff between detection cost and performance. However, this optimization problem is challenging in essence because of the special characteristics of MANETs. In particular, the deployment strategy must be adaptive to capture nodes mobility and robust to the failure of detection agents resulted from either accidental system error or intentional subversion. In this paper, we propose an adaptive, robust and sub-optimal strategy, which is called ARSoS, for tackling this issue. ARSoS treats each AIDS sensor as an independent agent, and then formulates their cooperative behavior as a decentralized decision problem. Since each AIDS sensor is only aware of partial information about the other sensors and the neighboring nodes, a reward signal integrating both local observation and global detection is introduced to guide their cooperation with one another. An online policy gradient algorithm is then applied to solve the formulated problem. An ARSoS prototype is implemented for simulations, and the results validate its performance in terms adaptability, robustness and optimality.**

I. Introduction

Recent years have witnessed the rapid development and wide-spread application of Mobile ad hoc networks (MANETs), which significantly enhance access to Internet services and provide various means for ubiquitous computing that enables high-speed and high-quality information exchange between mobile/portable devices located anywhere in the globe. Despite the tremendous potential and far-reaching impacts on the revolution of human life, MANETs suffer from security and privacy issues which dramatically impede their applications. To cope with the attacks, a large variety of intrusion detection techniques such as authentication, authorization, cryptographic protocols, key management schemes have been developed. However, most of them have high probabilities of being compromised due to various vulnerabilities resulted by open medium and wireless communication, as well as malicious insiders. As one of the backup measures, intrusion detection and response have paramount significance in the second defense line of MANET security.

While many intrusion detection systems (IDS) have been developed for wired networks, most of them can not de applied directly to MANETs due to the special characteristics of the infrastructure and the communication mode of MANETs that may lead to different vulnerabilities from traditional network paradigms, and it has been shown that anomaly-based IDS (AIDS) rather than misuse-based IDS is more suitable to diagnose anomalous nodes in MANETs. Also, since MANETs are self-policing and nodes are autonomous, fully distributed architectures are required for monitoring, reporting and tracking anomalous events, that is, AIDS sensors must be decentralized. While a variety of AIDS are claimed to be fully distributed, light-weight, and ready for application, their detection cost are not always neglectable, especially considering the fact that MANET nodes have scarce resource, which usually impels the nodes to save their energy from any unnecessary action. It is therefore a significant issue to deploy AIDS sensors in an optimal manner for achieving the best tradeoff between detection cost and performance. However, this optimization problem is challenging in essence due to the special characteristic and communication mode of MANETs. For example, the deployment strategy must be adaptive in that network topology always under changes caused by nodes mobility, and the deployment must be robust against the accidental communication failure due to the signal noise, channel interference, and traffic congestion. More seriously, some sophisticated attackers may manage to undermine AIDS by eavesdropping, capturing and analyzing AIDS sensors.

By tackling the technical challenges, we propose an adaptive, robust and sub-optimal strategy, which is called ARSoS, for automated deployment of AIDS sensors in MANETs. ARSoS is built on a sound theoretical framework, in which each AIDS sensor is treated as independent agent, and their cooperative behavior is formulated as a decentralized decision problem. In particular, since each AIDS sensor is only aware of partial information about the other sensors and the ongoing states of the network, a reward signal integrating both local observation (e.g., location, remaining energy) and global detection (e.g., detection performance, detection cost) is introduced to guide their individual behavior and cooperation with one another. Based on the model formulation, Multi-agent Partially Observable Markov Detection Process (MPO-MDP), we then employ a policy gradient algorithm to practically infer sensor behavior that is essentially controlled by a set of parameters, with the objective for achieving the best tradeoff between detection cost and detection performance. In order to validate the performance of ARSoS, we implement one prototype upon a reputation-based AIDS and conduct simulations to demonstrate its performance.

The rest of this paper is organized as follows. Section II reviews the related works. We put forth the deployment problem in Section III by giving observations, practical assumptions, and formal definition. Section IV addresses the specific design of our deployment strategy ARSoS. Section V reports simulations and discusses the results.

II. Related Work

A number of intrusion detection techniques and architectures for MANETs have been reported in [9], [10], [17], which suggest that the design of an IDS in MANETs mainly contains three elements: collecting and monitoring observations for characterizing node behavior and building normal profiles; designing classification algorithms for discriminating the deviation between ongoing network activity and normal profiles; developing models and architectures for exchanging, correlating and aggregating IDS alerts. In addition to intrusion detection, a set of response must be followed automatically for mitigating the detected anomalies. Intrusion detection and response therefore can be tightly integrated into a single model. For example, reputation-based system [7], [8], [18] introduces *reputation* to characterize the behavior of network nodes in terms of particular performance metrics, such as packet-forwarding rates. The periodical update of reputation allows the system to punish anomalous nodes whose reputation falls below a certain threshold. An alternative is to design fine-grained detection models by specifically analyzing routing protocols [15]. No matter what kind of detection techniques, a common assumption for them is that detection architectures should be fully distributed, and usually the detection (and response) cost is neglectable. Also, the performance evaluation of these systems solely relies on simulations, while sound theoretical analysis attracts much less attention than it deserves.

Game-theoretical framework is used in [2], [12] for analyzing the performance of detection schemes (as well as attack schemes). While the given theoretical bounds may help us to gain insightful understanding on optimal attacker behaviors, the emphasis of these work is not on computational cost consumed by defenders. Based on the assumption that mobile nodes are reluctant to run IDS detection algorithms for saving their energy, a mechanism design-based scheme is proposed in [11] for electing IDS nodes. The objective of this scheme is to balance the detection cost by urging more powerful nodes (which tend to selfish) to run detection algorithms and thus make the weaker nodes live longer. The scheme is built on IDS, while it does no take into account the performance of IDS, neither the adaptability issue. A fundamental difference between our scheme and the existing work is that our scheme treats IDS and underlying network together by integrating both nodes status and detection algorithms into a single model. In particular, our strategy aims at achieving the best tradeoff between detection performance and detection cost, it also deals with network-level factors including nodes mobility and communication errors. In addition, although the current version of our scheme is developed upon a reputation-based AIDS, its independent architecture and friendly interface allow it to be easily extended to the other AIDSs in MANETs.

III. Problem Statement

In this section, we show how to formulate the deployment of AIDS sensors as an optimization problem. Prior to that, we give key observations, design motivation and assumptions, by taking into account the practical implementation issues of AIDS in MANETs.

A. Key observations and motivation

Since MANET is self-policing, each node behaves as an autonomous entity and cooperates each other to fulfil network functions, there is no infrastructure to support centralized AIDS. As such, AIDS must be decentralized, that is, AIDS sensors must be fully distributed upon the nodes to cover the whole network. Furthermore, as MANET nodes are usually resource-constrained, the operational cost associate with detection and response must be minimal, or neglectable compared to the regular computational overhead. We also have a number of significant observations by examining the existing AIDSs, as follows,

- each AIDS has its own detection coverage and blind spot, which relies not only on detection algorithms but also on observations, i.e, the observable subjects under monitor, as well as their data-centric properties.
- since MANETs nodes are resources-constrained, AIDS sensors are usually light-weight and cost-sensitive, while the detection cost and latency are not always neglectable when an AIDS is considered as a whole.
- as MANET is self-policing, infrastructureless, and mobile, AIDS is expected to be self-adaptive to capture the drifts of network normality.
- there are a number of trade offs need to be tuned for achieving better detection performance, such as detection accuracy and false alert, anticipated performance and computational cost.

The key observations motivate us to explore a significant and challenging issue, that is, the optimal deployment of AIDS sensors in MANETs. In particular, we wonder whether it is a compelling need to fully deploy AIDS sensors on all the nodes or a majority of nodes, or is it possible to achieve an acceptable detection performance with fewer AIDS sensors and less detection cost. In addition to the optimality, robustness and adaptability are also key properties need to be examined carefully. We envision a framework which can provide the theoretical foundation for examining the relationship between detection cost and performance, along with a set of adaptive, robust and optimal deployment strategies for achieving the best tradeoff between the two metrics.

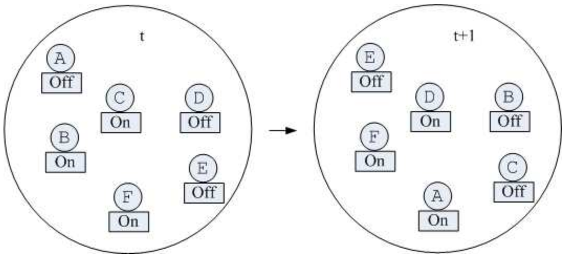

Fig. 1. Deployment of AIDS sensors: node B, C, F turns on their sensors at time t, while A, D, F turns on their sensors at time $t+1$; each node need to make decision whether its sensor should be turned on or turned off by estimating its local operating environment and global detection performance of AIDS on the entire network

B. Assumptions and Problem formulation

We assume that AIDS sensors operate independently from the underlying network, and cooperate each other to detect intrusive events. However, each AIDS sensor is only aware of its local environment, the detection coverage of a particular AIDS sensor is therefore limited, and they must exchange the local observation for obtaining global overviews. In addition, the detection results of each AIDS sensor must be aggregated for achieving accurate results and suppressing false alerts. Henceforth, both the local detection and global aggregation are important for an AIDS. However, while the functionality is always one of the key performance metric of AIDS, the computational cost can never be ignored. For a distributed AIDS with certain detection capability, its deployment strategy has great impact on computational overhead.

We also assume that AIDS sensors are fully distributed in MANET so that each node is able to run detection algorithm. But they can either switch on or off AIDS sensor (Fig. 1 illustrate the operations). Formally, we assume a MANET has N nodes (where the network size can be changed) and M AIDS sensors need to be switched on at time t for monitoring the whole network. In addition, we generally assume that the detection cost associated with an AIDS deployment strategy $p \in P$ is C_p, our objective is therefore simply represented as follows,

$$\arg \min_{0<M\leq N} C_p \tag{1}$$

This equation also implies the relationship between the number of AIDS sensors M and deployment strategy p, where M is usually less than N. Theoretically, there is a set of deployment strategies $P' \subseteq P$ which can minimize C_p. However, in practice, the special characteristics of MANETs make it impossible to explore a deterministic relationship between M and C_s, thus optimal solutions are not readily available. Rather, on the basis of our assumptions and specific considerations, which are treated as practical constraints, we use meta-heuristic algorithms to solve this optimization problem.

IV. Our proposed strategy: ARSoS

This section discusses our specific design, ARSoS. We firstly formulate the general behavior of AIDS sensors as a multiple agent partially observable markov decision process, or MPO-MDP, we then apply an online policy gradient algorithm to practically infer the collective behaviors of AIDS sensors.

A. Design requirements

Our design is primarily from the perspective of AIDS that runs constantly and remains independent or transparent to the network, enhancing the quality of service associated with security. In general, AIDS should be effective in terms of *detection accuracy* and efficient in terms of *computational cost*. While our design aims at improving AIDS performance by balancing the two metrics, the following design requirements must be examined carefully,

- the design should not cause much extra computational cost, and it should not result in performance deterioration.
- the design should not introduce new vulnerabilities.
- the design should be robust and resist subversion, that is, the failure of any AIDS sensor should not result in performance degradation.
- the design should be adaptive to the changing environment that is mainly caused by nodes mobility.

Our design treats AIDS as a whole and as an self-policing system, in which detection sensors are autonomous and behave independently. After a certain period of monitor process and based on the local observations, each sensor makes a decision on the occurrence of an anomalous event. Since each sensor works in its own operating environment, it can only be aware of its local area and thus have no global knowledge about the entire network. The sensors, therefore, have to exchange their personal opinions for achieving a global overview. To do that, a set of efficient and effective aggregation protocols and consensus mechanisms is usually required. However, the explicit communications among AIDS sensors must be avoided because they may cost extra computational overhead, which is against the design requirement listed above.

B. Design rationale

For each AIDS sensor, its estimate on the current state, as well as the potential action, depends on the previous state, so the decision process is a Markov decision process. Also, each sensor can partially observe the ongoing network state, while a general state of the entire network can be only estimated and maintained by the multiple AIDS sensors, by collecting and integrating their observations together. Therefore, the collective behavior of multiple AIDS sensors can be essentially formulated as Multi-agent Partially Observable Markov Decision Process, or MPO-MDP for short. Formally, a POMDP model is structurally characterized by four key elements [1]: a finite state space S, an action space U, an observation space Z, and a (possibly stochastic) reward $r(i) \in R$ for each state $s_i \in S$, i.e., $\mathcal{M} = \{S, U, Z, R\}$.

As a POMDP model, the interaction between an individual AIDS sensor id_x and its operating environment includes a sequence of decision stages, working as follows,

- at stage i, the local network monitored by id_x is in a particular state $s_i \in S$, and the monitored observation is $z_i \in Z$ (we assume the observation is generated with a certain probability distribution $\nu(s_i)$),
- motivated by the observed z_i, sensor id_x takes action in accordance with a randomized policy based on a probability distribution $\mu(z_i)$ over actions,
- the action $u_i \in U$ determines the state transitions from s_i to s_j, with a certain probability $p_{ij}(u_i)$,
- after taking the action and with a certain delay, sensor id_x receives a reward signal $r_i \in R$, while the objective of sensor id_x is to choose a policy for maximizing a predefine reward function.

Therefore, in general, the decision process of each AIDS sensor can be regarded as a Markov chain: $s_i \in S[\nu(s_i)] \rightarrow z_i \in Z[\mu(z_i)] \rightarrow u_i \in U[p_{ij}] \rightarrow s_j \in S$. The principle is that AIDS sensors need to search the policy space which is a mapping from current state to actions.

With the formulation of AIDS sensor's independent behavior, the entire AIDS system can be naturally formulated as a multi-agent POMDP, or MPO-MDP. Formally, the action set of AIDS U contains the cross product of all the individual AIDS sensor's action, that is, $U = \{\hat{u}_i | \hat{u}_i = u_i^1 \times u_i^2 \times \cdots \times u_i^m\}$ (where m is the number of AIDS sensors). At each stage, each AIDS sensor selects its action independently according to an observation vector, which then combines a general action of AIDS. For stochastic policies, the overall action distribution is the joint distribution of actions for each sensor, $\mu(u_1, u_2, ...u_n | z_1, z_2, ...z_n)$. Although we need to consider the practical constraints and specific characteristic of MANETs, the model formulation has a number of promising advantages,

- the cooperation among AIDS sensors does not take into account their explicit inter-communication, since only a reward signal is shared, enabling AIDS to adapt to diverse network situations,
- the distributed architecture allows any AIDS sensor to leave and be incorporated into the detection system without any extra operation, enabling its scalability to the changing network topology,
- since AIDS essentially considers the individual behavior of sensor as a whole, the consensus strategy and agreement protocol may achieve reliable decisions in the presence of sensor's Byzantine behavior,
- since the cooperation manner among AIDS sensors is formulated as an optimization problem, the inferred collective behavior may maximally increases the rewards, eventually leading to a set of optimal (or near-optimal) decision strategies,
- there is a suit of algorithms ready for solving the formulated problem, e.g., a policy-gradient reinforcement learning algorithm can be applied to tackle the delayed reward, partially observable, multi-agent learning problem.

C. Model Construction

Model construction is a process of specifying model-base parameters. Except the reward signal, we consider the other parameters independently for each AIDS sensor. We assume each sensor collects and maintains three observations,

- *link_state*, which reflects the status of network topology;
- *reputation*, which denotes the node reputation;
- *power*, which measures the node's remaining resource.

Let θ=$(\theta_1, \theta_2, \theta_3)$ be a parameter vector serving as thresholds for the three observations. For example, we have $link\_state = 1$ if $\theta_1 > \#neighbor(i)_t / \#neighbor(i)_{t-1}$ which represents the changing ratio of the neighbors of node i from time $t-1$ to t, and node i is supposed to run AIDS sensor in this case; also, we say node i can not serve as an AIDS sensor if its reputation less than threshold θ_2, or its power is not sufficient to achieve the lower bound θ_3 for running AIDS sensor. The observation space of the model is therefore $Z = \{link\_state : 0, link\_state : 1, reputation : 0, reputation : 1, power : 0, power : 1\}$, and the state space S is directly derived from Z. The action space is defined as $U = \{0, 1, 2, ...u_{max}\}$, where $u_i = 0$ means AIDS sensor is off, $u_i = k$ means sensor AIDS may keep running for k time windows.

To make the model complete, we need to define the reward signal r. Considering the behavior of AIDS sensor and its detection results, two items are absorbed into our reward function: the computational cost C for running detection engine, and the gratitude tokens G from the other nodes for the provided AIDS service. The service can be further classified into two cases, *positive* G_{pos} (good service) and *negative* G_{neg} (bad service). Formally, we define reward function as follows,

$$r = G - C = G_{pos} - G_{neg} - C \tag{2}$$

So Eq. (1) can be replaced by the following one,

$$\max_{p \in P} \{ \lim_{T \rightarrow \infty} \mathbb{E}[\frac{1}{T} \sum_{t=1}^{T} r_t] \} \tag{3}$$

where $\mathbb{E}$ is the expectation operator, and r_t relates reward signal r to the deployment strategy at time t. Obviously, this objective function has two implications,

1) from a long-standing viewpoint, the ultimate goal is to maximize the reward r by taking the optimal deployment strategies,
2) the action of AIDS sensor U is also essentially related to the reward signal via G, so the actions of each AIDS can be optimized simultaneously.

Note that reward signal is a parameter integrating both local and global elements. In particular, in Eq. (2), G is global element, while C is local element. To make it general, we may introduce an impact factor $\alpha \in [0, 1]$ to balance the two elements, that is,

$$r = \alpha \cdot (G_{pos} - G_{neg}) - (1 - \alpha) \cdot C \tag{4}$$

As such, the behavior of a AIDS sensor is impacted by both its available computational resource and evaluation on its

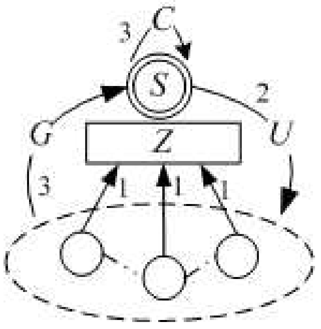

Fig. 2. An AIDS sensor behaves as POMDP model: it collects observations Z from a group of nodes in its vicinity periodically, takes actions according to the detection result (the system estimate S), and then it receives feedback (G and C) as the evaluation of its action consequence; the edge numbers show the decision sequence.

behavior from other nodes, and their coordination may achieve optimal deployment globally. Fig. 2 illustrates the behavior of a typical AIDS sensor acting as the constructed model.

D. A policy gradient algorithm

One problem naturally aries after the model construction, i.e., how can we practically infer the deployment strategy p from the constructed models. To do that, we need to design suitable algorithms for maximizing objective function Eq. (3). While a bunch of reinforcement learning algorithms are available for solving MPO-MDP models [1], the characteristic of MANETs and our design requirements call for a fast and lightwight one. Baxter and Bartlett presented OLPOMDP [5], [6], an algorithm that learns to adjust the parameters θ of a randomized policy with observation z_t, and chooses actions according to $\mu_\theta(z_t)$. Henceforth, let $\eta(\theta) := \lim_{T\to\infty}\mathbb{E}[\frac{1}{T}\sum_{t=1}^{T} r_t]$, Eq. (3) is transformed into such a problem: adjusting policy parameters θ to climb the gradient of $\eta(\theta)$ (the theoretical foundation can be referred in [5]).

As defined previously, in our model, θ is the concatenation of the parameters from an AIDS sensor. So a critical issue here is how to relate parameter θ with AIDS sensor behavior (or policy) $\mu(z_t)$, or how to get $\mu_\theta(z_t)$. As our objective is to save the overall computational cost for running AIDS sensors by seeking their optimal deployment strategies, the occurrence of switching off AIDS sensor for a node is rare but have many opportunities to happen, it is practical and reasonable to assume this probability as a *poisson distribution*. Also, such a parameterized policy structure defines the probability of choosing a particular action which is a continuous differentiable function of θ, supporting the application of OLPOMDP [5]. Moreover, this assumption allows the sensor to be trainable and keep the computations simple,

$$Pr(u_t = 0) = \frac{\exp(-f(\theta, u_t)) f(\theta, u_t)^0}{0!} = \exp(-f(\theta, u_t)) \tag{5}$$

where $f(\theta, u_t)$ further explores the direct relationship between action u_t and parameter θ, varying in the impact of θ_i on u_t. An ideal function is expected to correlate each element of θ with u_t in a fine way so that they can be adjusted simultaneously. However, to keep the analysis and simulation simper, in our model, we only consider the relationship between $\theta_{3,t}$ and u_t (we use $\theta_{3,t}$ instead of θ_3 means that $\theta_{t,3}$ is trainable and different from the other two parameters), while θ_1 and θ_2 are determined in advance. So assuming the observation at time t is $z_t = (z_{t,1}, z_{t,2}, z_{t,3})$, we have the following equation,

$$u_t = \begin{cases} k, & \text{if } z_{t,1} > \theta_1, z_{t,2} > \theta_2, \text{ and } z_{t,3} > \theta_{t,3} \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

where $k = \{1, 2, \cdot\cdot\cdot, u_{max}\}$, and the exact value can be determined by $\theta_{t,3}$ with the positive values of θ_1 and θ_2 that have been predefined as constants. Considering the physical meaning of $\theta_{t,3}$, i.e., the available resource of a node at time t, we give a simple equation for relating $\theta_{t,3}$ and $z_{t,3}$, as follows,

$$u_t = \frac{z_{t,3} - \theta_{t,3} - \delta \cdot z_{t,3}}{\theta_{t,3}} \tag{7}$$

where $\delta \cdot z_{t,3}$ plays as a relaxable coefficient for representing the consumed power other than running AIDS sensor, and the equation only holds when parameters θ_1 and θ_1 are satisfied. By introducing Eq. (7) to Eq. (5), we can derive a parameterizable police $\mu_\theta(z_t)$ for AIDS sensors,

$$\mu_\theta(z_t) = \begin{cases} \exp(-\frac{z_{t,3} - \theta_{t,3} - \delta \cdot z_{t,3}}{\theta_{t,3}}), & \text{if } u_t = 0, \\ 1 - \exp(-\frac{z_{t,3} - \theta_{t,3} - \delta \cdot z_{t,3}}{\theta_{t,3}}), & \text{otherwise} \end{cases} \tag{8}$$

In order to get an optimal policy $\mu_\theta(z_t)$, we need to seek the best controlling parameter θ. A simple method for computing an appropriate direction for update the parameter θ has been proposed and applied in [4], [14], which works as the follows,

$$\theta_t = \theta_{t-1} + \triangle\theta = \theta_{t-1} + \tau_t \cdot r_t \cdot q_t \tag{9}$$

where the long-term average of the updates $\triangle\theta$ lie in the gradient direction $\nabla\eta(\theta)$, r_t is the sum of the rewards that have been received, τ_t is the suitable size of the steps taken in parameter space, and the vector q_t is an eligibility trace of the same dimensionality as θ, and it is used to update the parameter θ and guides the policy $\mu_\theta(z_t)$ to climb the gradient of the average reward. In particular, vector q_t is computed and updated in the following way,

$$q_{t+1} = \rho \cdot q_t + \frac{\nabla\mu_\theta(z_t)}{\mu_\theta(z_t)} \tag{10}$$

where $\rho \in (0, 1)$, $\mu_\theta(z_t)$ is the probability of the action u_t under the current policy, and ∇ denotes the gradient with respect to the parameters θ. By introducing Eq. (8) to Eq. (10) and then to Eq. (9), we can update the parameter step by step and therefore gradually achieve the optimal policy $\mu_\theta^*(z_t)$. The algorithm works as follows,

```
Initialization coefficient $\rho \in [0, 1)$, step size $\tau_0$, initial thresholds $\theta_0$, observations $z_0$
for episode $t = 1$ to $T$ do
    Gets observation $z_t$;
    Takes action $u_t$;
    Reports detection results;
    Receives reward signal $r_t$;
    Updates $q_{t+1}$ according to Eq. (8) and Eq. (10):
    if $u_t = 0$ then
        $q_{t+1} = \rho \cdot q_t - \frac{(1-\delta)\cdot z_t}{\theta_t^2}$
    else
        $q_{t+1} = \rho \cdot q_t + \frac{(1-\delta)\cdot z_t \cdot \exp(1-\frac{(1-\delta)\cdot z_t}{\theta_t})}{\theta_t^2(1-\exp(1-\frac{(1-\delta)\cdot z_t}{\theta_t}))}$
    end
    Updates $\theta_{t+1}^i$ according to Eq. (9):
    $\theta_{t+1} = \theta_t + \tau_t \cdot r_t \cdot q_{t+1}$
end
```

Algorithm 1: AIDS sensor-critic policy gradient algorithm

Note that the parameter θ_t in the algorithm is actually $\theta_{t,3}$ in our model, and the right side of q_t update is the results by introducing Eq. (8) to Eq. (10). The key feature of the algorithm is that the only non-local information each AIDS sensor needs is a global reward signal, and they do not need to know any other information about the system state in order to climb the gradient of the global average reward.

E. Practical considerations

The application of OLPOMDP algorithm to our model requires two implicit assumptions, (1) For every given θ, the system is ergodic and converges to a unique steady state; (2) the update of parameter θ of each AIDS sensor may contribute to a global optimal cooperation between them. The second assumption can be validated by the proof in [3], [5], while the first assumption is often yet not always satisfied. However, although the mobility of nodes in MANETs does not support the occurrence of a steady state, the algorithm anyways tends to converge to such a state, a "near-optimal" state can happen with a certain probability. In general, the application of reinforcement learning algorithm poses three important features,

- *Adaptability*. In our model, reward signal is the only information shared by the sensors, and the cooperation does not take into account the explicit inter-communication between sensors. This allows every sensor to adapt to diverse system conditions and can capture the node mobility.
- *Robustness*. Since AIDS considers all the independent sensors as a whole, it can make correct decisions with a high probability even though in the presence of Byzantine behaving nodes.
- *sub-Optimality*. The cooperation between AIDS sensors is essentially formulated as an optimization problem with the objective function Eq. (1) and some practical constraints. The simple update rule modifies the parameters of each AIDS sensor in the direction that maximally increases the average reward, which leads to parameter values that locally optimize the performance of the independent sensors. The general behavior of AIDS is thus anticipated to be found a set of optimal deployment strategies by turning on and off corresponding sensors. The flexible network topologies determines that only sub-optimal policy can be obtained even though optimal ones are theoretically achievable given a particular topology and sufficient time.

TABLE I
SIMULATION SETTINGS

Network	Area	1000m × 1000m
	Topology	random
	Placement of nodes	uniform
	MAC	802.11b
	Routing protocol	DSR
Node	# of nodes	30
	# of normal nodes	23
	# of malicious nodes	7
	# of pre-trusted nodes	6
	# of anomaly detectors	30
	initial energy	10
Attacks	DoS	Packet dropping
Simulation	# of simulation epochs	20
	running time	5000s
	Length of detection window (l_{DW})	20s
Parameters	θ	$(0.50, 0.65, 0.60)$
	α	0.3

V. PERFORMANCE VALIDATION

In this section, we conduct simulations to validate the performance of our proposed scheme, with the particular concern on its adaptivity, robustness, and optimality.

A. Simulation Settings

Since the objective of our simulation is to examine the deployment strategies of AIDS sensors in MANET, we conduct it with our reputation-based system RARDAR that has been report in [18], where *reputation* is used to characterize the behavior of nodes. The simulation settings follow the one in [18] with minor modifications, as shown in Table I.

1) Network model: We consider a typical MANETs is consisted of mobile nodes. The MAC layer operates IEEE 802.11b, and the network layer runs DSR routing protocol. The application layer uses CBR for generating data packets (6/Sec.). 50 source-destination connections are randomly and periodically (10 seconds) generated among the nodes. The network is randomly deployed in a space of 1000×1000 m^2. Also, we select Random Waypoint Mobility Model to simulate the node mobility, and the node speed is randomly generated between 0 and 20 m/s.

2) Node model: We assume the network to have 30 mobile nodes, among which 23 are benign and 7 are malicious (which launches DDoS attacks by dropping routing packets). We select 6 pre-trusted nodes by setting high reputation values (0.99) and the initial reputation of the rest nodes is 0.50. We also assume that all nodes are location-aware and identified by IP addresses. All the nodes are equipped with AIDS sensors, while at the beginning of the simulation, we only switch on 6 AIDS sensors located at pre-trusted nodes.

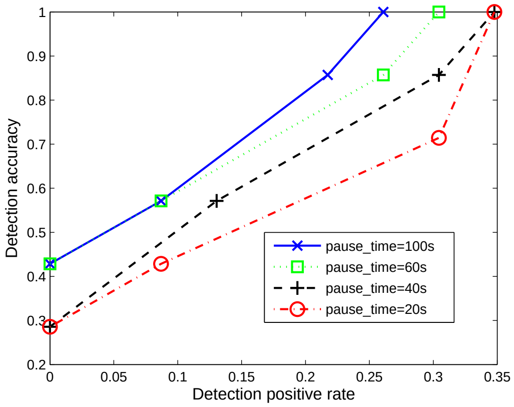

Fig. 3. Adaptability: measuring the detection performance of ARSoS in different scenarios where nodes have different pause time; the longer of the pause time the more stable of network topology and the better of ARSoS performance.

3) Evaluation metrics: While a common criteria for evaluating AIDS is the trade-off between the capability of detecting attacks and the ability of suppressing false alerts, we mainly examine the performance of our deployment strategy of AIDS under particular network scenarios (the detection performance have been validated in our previous work [18]). More specifically, we use adaptability to measure the performance of AIDS in the presence of node mobility; robustness is used to measure the performance of AIDS in the presence of failed sensors (caused either by attacks or system errors); we then use a more general metric, optimality, to observe whether a particular deployment strategy is optimal in sense that the detection cost and performance are balanced. To guide the evolution of AIDS sensor behavior, the reward signal in Eq. (4) is defined by the detection accuracy and false positive rate.

4) Time units: 20 simulation epochs are executed for averaged results, and each of them lasts 5000s. The duration of detection window is set as $l_{DW} = 20s$, so there are totally $|DW| = \frac{5000}{20} = 250$ detection windows available. ARSoS starts running at the first detection window, and the malicious nodes start dropping packets at the 151th detection window.

B. Results and analysis

Firstly, we examined the performance of ARSoS on adaptability by varying the pause time of mobile nodes. The number of nodes running AIDS sensors was initially set as 6 (all are pre-trusted nodes), Fig. 3 shows that the detection performance (represented by ROC) in terms of detection accuracy and false positive rate did not have significant changes as pause time varied, even though slight performance deterioration occurred as the pause time tends to shorter. Careful analysis reveals two reasons leading to this trend: the first one is from ARSoS algorithm itself, since it has less time to achieve a better performance although it could do if time is sufficient; the second one is from reputation-based detection scheme [18], which usually needs a longer time to collect more evidence for calculating trust values, while a shorter time frame may result in higher false positive rate. For example, if the detection accuracy gets to 100%, the false positive rate is $6/23$ with pause time $100s$, $7/23$ with pause time $60s$, and $8/23$ with pause time $40s$ and $20s$. It is obvious that ARSoS is able to achieve a very low false positive rate if the pause time becomes very large, while acceptable false positive rate can be achieved when the pause time lies in a reasonable range.

Secondly, we intentionally set a fraction of nodes to be selfish by manipulating the parameter θ_3 (assume the nodes to be compromised by an attacker so that they did not behave as the algorithm required). This group of nodes always intended to save their energy by refusing to run AIDS sensors. In the simulation, we mainly explored the relationship between the percentage of selfish nodes and the detection performance. Fig. 4 shows us the story: the detection performance does not suffer sharp change when the number of selfish nodes is not so large, e.g., the false positive is $5/23$ and $6/23$ (detection rate is 100%) when the number of selfish nodes is 3 and 7 respectively. However, if the number of selfish node is 9, the false positive rate kept raising to 100% in order to detect all the selfish nodes. So we have to claim there is a unknown threshold determining the robustness of ARSoS. The preliminary simulation and analysis shows that the number of selfish nodes should be kept less than one third of the total nodes (a typical threshold in Byzantine protocols) in order to guarantee the detection performance. Another important observation is that ARSoS has to undergo a longer time to achieve an desirable performance in the presence of selfish nodes, and the detection cost also gets more in this case, as more normal nodes are required to run AIDS sensors for obtaining reliable consensus strategies.

Finally, we examined whether ARSoS can contribute to the saving of detection cost or not. Since in our simulation, malicious nodes launched their attacks at the 150th detection window, the detection windows from 0 to 149 can be viewed as a training process of ARSoS. During the training, we used reward signal r_t as a metric to guide the collective behavior of AIDS sensors. To specify the reward signal defined in Eq. (4), we need to specify each item in this equation, including α, G_{pos}, G_{neg}, and C. In order to make the computation simpler, we set G_{pos} as detection accuracy, G_{neg} as false positive rate, and C as detection overhead ratio, which is defined in [18] as $C{=}\sum_{i=1}^{n} \#msg_d / \sum_{i=1}^{n} \#msg_{all}$ (where $\#msg_d$ is the number of messages related to AIDS, and $\#msg_{all}$ is the number of all the messages travelling among communication links). However, in order to observe the evolution of AIDS sensors collective behavior for examining ARSoS's optimality on detection cost, we only took C as a metric for examination during the detection stage (from detection window 150 to 250), and the pause time was set as $100s$ for easier analysis (a shorter pause time needs more detection windows for ARSoS to get optimal policy). As shown in Fig. 5, the trend of detection overhead ratio is decreasing as time advances, even though a number of local optimal occurred during the evolution. An optimal policy was achieved at the 215th detection window, where detection overhead ratio kept as 0.163.

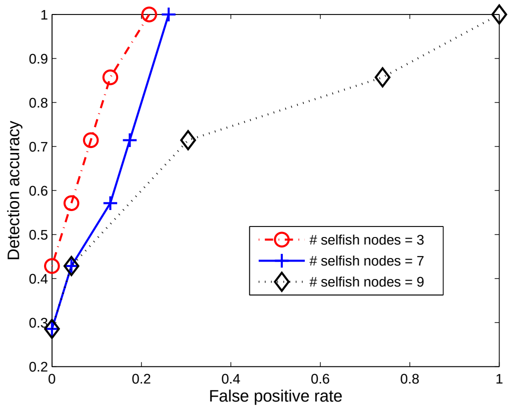

Fig. 4. Robustness: measuring the detection performance of ARSoS in the scenarios where a fraction of selfish nodes refusing to run AIDS sensors even though they are able to do so; the more selfish nodes the worse of ARSoS detection performance and the longer time it needs to get reliable detections.

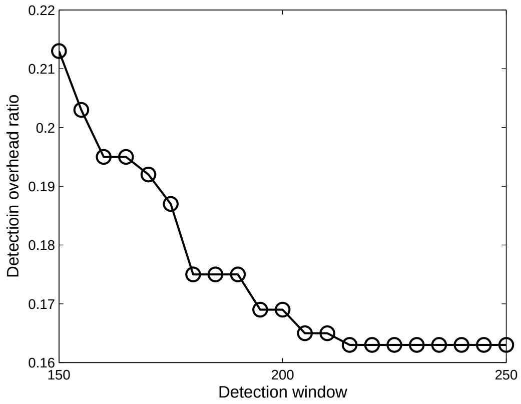

Fig. 5. Optimality: measuring the performance of ARSoS on detection overhead ratio; for a particular network topology at time t, ARSoS tends to maximize reward signal r_t thereby reducing detection overhead

VI. Concluding Remarks and Future Work

In this paper, we proposed an automated strategy, ARSoS, for deploying AIDS sensors in MANETs. Due to the intrinsic characteristics of MANETs, the deployment problem was formulated as an NP optimization problem, where the tradeoff between detection performance and detection cost is viewed as objective function, and practical constrains are drawn from the observations of network nodes and communication links. The problem was then cast in MPO-MDP framework, where each AIDS sensor was treated as an autonomous agent, and its detection was formulated as a Markov decision process. A policy-gradient algorithm was applied to seek a set of optimal parameters controlling AIDS detection policy, so probabilistic rather than deterministic solutions were inferred as deployment strategies. Both the theoretical analysis and simulations validated that the proposed strategy is adaptive, robust and near-optimal. As the subsequent work, we will design more efficient algorithms for searching parameter space, we also need to examine more sensitive parameters that impact AIDS detection performance and optimize them simultaneously. In addition, we will conduct more extensive simulations to validate ARSoS's performance. The adaptability, robustness, and optimality have been examined independently, while the hidden relationships between nodes mobility, number of AIDS sensors, and detection cost are still unclear.

References

[1] D. Aberdeen, "A Survey of Approximate Methods for Solving Partially Observable Markov Decision Processes," *National ICT Australia Report*, Canberra, Australia, Dec. 8, 2003.

[2] J. S. Baras, S. Radosavac, et al., "Intrusion Detection System Resiliency to Byzantine attacks: the case study of wormholes in OLSR," In *Proc. of the IEEE Military Communications Conference 2007*, pp. 1-7, Oct. 2007, Orlando.

[3] Peter L. Barlett and Jonathan Baxter, "Hebbian synaptic modifications in spiking neurons that learn", *Technical report, Computer Sciences Laboratory,* RSISE, ANU, 1999.

[4] Jonathan Baxter and Peter L. Barlett, "Stochastic Optimization of Controlled Partially Observable Markov Decision Processes", *Proceedings of the 39th IEEE Conference on Decision and Control(CDC00).*

[5] Jonathan Baxter and Peter L. Barlett, "Direct Gradient-Based Reinforcement Learning: I.Gradiment Estimation Algorithms", *Technical report*, ANU,1999.

[6] J. Baxter, L. Weaver, and P.L. Bartlett, "Direct Gradient-Based Reinforcement Learning: II.Gradiment Descent Algorithms and Experiments", *Technical report, Research School of Information Sciences and Engineering,* Australian National University, September 1999.

[7] S. Buchegger, and J. -Y. Le Boudec, "Performance analysis of the CONFIDANT protocol," In *Proc. of ACM Mobihoc'02*, pp. 226-236 Lausanne, Switzerland, June 2002.

[8] Q. He, D. Wu, and P. Khosla, "SORI: A secure and objective reputation-based incentive scheme for ad hoc networks," In *Proc. of Wireless Communications and Networking Conference (WCNC2004)*, pp. 825-830, Atlanta, USA, March 2004.

[9] Y. Huang, and W. Lee, "A cooperative intrusion detection system for ad hoc networks," In *Proc. of the ACM workshop on Security in Ad hoc and Sensor Networks (SASN03),* pp. 135-147, Fairfax, Virginia, Oct.,2003.

[10] A. Mishra, K. Nadkarni, and A. Patcha, "Intrusion detection in wireless ad hoc networks," *IEEE Wireless Communications,* pp. 48-60, Feb. 2004.

[11] N. Mohammed, H. Otrok, et al., "A mechanism design-based multi-leader election scheme for intrusion detection in MANET," In *IEEE Proc. of WCNC2008*, Las Vegas, USA, Mar. 2008.

[12] S. Radosavac, A. A. Crdenas, et al., "Detecting IEEE 802.11 MAC Layer Misbehavior in Ad Hoc Networks: Robust strategies against individual and colluding attackers," *Journal of Computer Security*, vol. 15, no. 1, pp. 103128, January 2007.

[13] T. Clausen and P. Jacquet, "Optimized Link State Routing Protocol (OLSR)," *IETF RFC 3626 (Experimental)*, Oct. 2003.

[14] Nigel Tao, Jonathan Baxter, Lex Weaver, "A Multi-Agent, Policy-Gradient approach to Network Routing", *18th International Conference on Machine Learning*, ICML 2001.

[15] C. H. Tseng, S-H. Wang, C. Ko, and K. N. Levitt, "DEMEM: Distributed Evidence-Driven Message Exchange Intrusion Detection Model for MANET," In *Proc. of the 9th Intertional Symposium on Recent Advances in Intrusion Detection (RAID 2006),* pp. 249-271, Hamburg, Germany, Sept. 2006.

[16] S.-H. Wang, C. H. Tseng, K. N. Levitt, and M. Bishop, "Cost-sensitive intrusion responses for Mobile Ad Hoc Networks," In *Proc. of the 10th Intertional Symposium on Recent Advances in Intrusion Detection (RAID 2007),* pp. 127-145, Queensland, Australia, Sept. 2007.

[17] Y. Zhang, W. Lee, and Y. Huang, "Intrusion detection techniques for mobile wireless networks," *ACM Wireless Networks Journal,* 9(5):545-556, September 2003.

[18] Z. Zhang, F. Naït-Abdesselam, P.-H Pin, and X. Lin, "RADAR: a ReputAtion-based scheme for detecting anomalous nodes in wireless mesh networks," In *Proc. of Wireless Communications and Networking Conference (WCNC2008)*, Las Vegas, USA, Mar. 2008.