A Curious Agent for Network Anomaly Detection

(Extended Abstract)

Kamran Shafi
School of Engineering and Information Technology
University of New South Wales @ Australian Defence Force Academy
Canberra, ACT, Australia
k.shafi@adfa.edu.au

Kathryn Merrick
School of Engineering and Information Technology
University of New South Wales @ Australian Defence Force Academy
Canberra, ACT, Australia
k.merrick@adfa.edu.au

ABSTRACT

This paper presents a novel approach to intrusion detection using curious agents to detect anomalies in network data. Curious agents use computational models of novelty-seeking behavior and interest, based on human curiosity, to reason about their experiences in their environment. They are online, single-pass agents that respond to the similarity, frequency and recentness of their experiences. As such, they combine a number of important characteristics required for intrusion detection. This paper presents a generic, curious reflex agent model for network intrusion detection and the results of experiments with a number of variants of this model. Specifically, five different models of curiosity are compared for their ability to detect first instances of attacks in the KDD Cup data set. Results show that our curious agents can achieve high detection rates for intrusions, with moderate false-positive rates.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous

General Terms

Algorithms

Keywords

curious agents, novelty, interest, intrusion detection, anomaly detection.

1. CURIOUS REFLEX AGENTS FOR NETWORK INTRUSION DETECTION

Our curious agent model uses three reasoning processes to monitor the network: sensation, curiosity and activation. These processes are discussed in detail in the following sections.

1.1 Sensation

An agent monitors its environment, in this case a network, using its sensors. In the experiments in this paper, the agent's sensors read simulated network data (connection records) from a comma-separated value file. This raw data is converted into two structures to assist further reasoning. The first is an observation vector and the second an event. An observation vector $O_{(t)} = (o_{1(t)}, o_{2(t)}, \cdots o_{j(t)})$ represents the network data packet at the time t. An event $E_{(t)}$ represents the change in observed network data between time t and time $t-1$

1.2 Curiosity

The curiosity process models the behavior of a network and uses this model to compute a curiosity value $C_{(t)}$ for each observation or event. The curiosity process has up to three layers. The first layer is the clustering layer. In this layer, an unsupervised learning algorithm is used to cluster observations or events. Each time an observation or event is presented to the clustering layer a winning cluster-center $K_{(t)} = (k_{1(t)}, k_{2(t)}, \cdots k_{j(t)})$ is chosen or created to best match the observation or event.

The second layer is the habituating layer [1]. The habituating layer comprises of one neuron for each cluster-center in the clustering layer. The activity of the winning cluster-center (and its neighbors in the case of the SOM) are propagated along the synapse to the habituating layer as a synaptic value $\varsigma_{(t)} = 1$. Losing cluster-centers give an input of $\varsigma_{(t)} = 0$ to the habituating layer. Synaptic efficacy, or novelty, $N_{(t)}$, is then calculated as a stepwise solution to Stanley's model [3] by approximating $N_{(t)}$ as follows:

$$\tau \frac{\mathrm{d}N_{(t)}}{\mathrm{d}t} = \alpha[N_{(0)} - N_{(t)}] - \varsigma_{(t)}$$

$$N_{(t)} = N_{(t-1)} + \frac{\mathrm{d}N_{(t-1)}}{\mathrm{d}t}$$

The habituation function controls the rate of change in novelty values, which permits tuning of the alarm load on the human security supervisor.

The third layer is the interest layer. In this layer, a single interest value is computed using the Wundt curve [4] with the novelty value from the winning habituating neuron as input. The interest function moderates novelty values over time and frequency, providing finer control over the detection versus false-alarm trade-off. Curiosity can thus be considered as a function of the similarity of an observation to previous observations (computed using the clustering layer), its recentness (which impacts its novelty) and the frequency with which it occurs (which impacts its interest). A compar-

Cite as: A Curious Agent for Network Anomaly Detection (Extended Abstract), Kamran Shafi and Kathryn Merrick, *Proc. of 10th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2011)*, Tumer, Yolum, Sonenberg and Stone (eds.), May, 2–6, 2011, Taipei, Taiwan, pp. 1075-1076.

ison of two broad variants of this model is shown in Figure 1. The first models curiosity $C_{(t)}$ as novelty (i.e. $C_{(t)} = N_{(t)}$), while the second models curiosity as interest (i.e. $C_{(t)} = I_{(t)}$).

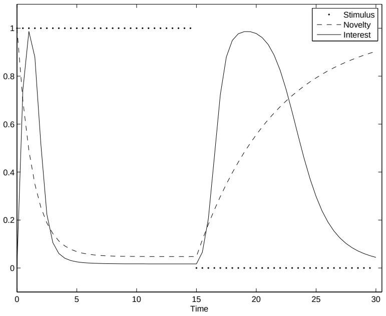

Figure 1: Novelty and interest in response to time and a changing stimulus observation or event.

1.3 Activation

The activation process reflexively raises an alarm when a highly curious, and thus potentially anomalous, observation or event is sensed. The notion of high and low curiosity implies a curiosity threshold Ψ below which network data is ignored and above which an alarm is raised.

2. EXPERIMENTS

This section details an experiment with four variations of the general curious agent model described above. We use the benchmark KDD Cup data set, as the network environment to be inhabited by the agents. We analyze the following variants of curious agents:

SOM-I: A three layer approach reasoning about observations using a SOM clustering layer, a habituating layer to compute novelty and an interest layer. Curiosity is equal to interest using this model.

SOM-N: A two layer approach reasoning about observations using a SOM clustering layer and a habituating layer to compute novelty. There is no interest layer in this model. Curiosity is equal to novelty.

KMEANS-N: A two layer approach reasoning about observations using a K-means clustering layer and a habituating layer to compute novelty.

SART-N: A two layer approach reasoning about observations using a SART clustering layer and a habituating layer to compute novelty.

2.1 Measurement Approach

In this paper we use a weighted measure to identify true-positives. In particular, we are interested in only the first i (for the experiments in this paper, we used $i = 1$) instances of any attack sequence, where an attack sequence may consist of one or more back-to-back connection records belonging to a particular attack type which are disjointed by normal or other types of attack connections. It implies that in a production network an alarm is raised only i times for the network administrator. It is assumed that, for an IDS operating in real-time, the network administrator would take some action to prevent further instances in the attack sequence from occurring at all.

Table 1: Weighted true-positive detection rates (%) for attack categories and unweighted false positive rates for normal data at t=500,000. Only the agents reasoning about observations are shown.

Category	SOM-I	SOM-N	KMEANS-N	SART-N
Probe	44.44	88.89	95.56	97.78
DOS	26	74	76	88
U2R	43.48	91.3	95.65	95.65
R2L	47.62	54.76	69.05	80.95
Normal	53.41	31.85	15.09	36.29

Table 2: Weighted true-positive detection rates (%) for attack categories and unweighted false positive rates for normal data at t=800,000. Only the agents reasoning about observations are shown.

Category	SOM-I	SOM-N	KMEANS-N	SART-N
Probe	55.13	48.86	74.14	89.54
DOS	38.85	42.57	37.16	40.2
U2R	61.71	62.29	69.71	88.57
R2L	48.3	38.92	41.84	65.61
Normal	57.45	37.5	22.39	47.64

2.2 Results and Discussion

Tables 1 and 2 summarize the category-wise results for the four agents at $t = 500,000$ (training data set only) and $t = 800,000$ (training and test data sets). We can conclude that the KMEANS-novelty agent has the best trade-off between true-positive and false-positive rate when the nature of the data being sensed is unchanging.

Almost all of the agents tested in this paper achieved high detection rates on the two rare classes (U2R and R2L) in the KDD Cup data sets. This is in contrast to most published results using traditional machine learning algorithms. For example, the winner the KDD Cup achieved a test accuracy of just 13.16% and 8.40%, on U2R and R2L attacks. Likewise, the runner up achieved a test accuracy of 11.84% and 7.32% on U2R and R2L attacks. Our approaches achieved up to 95% accuracy for detecting first instances of these attack types. This is very encouraging given that the agents are single pass and completely unsupervised.

In summary, the results presented in this paper do show promise for curious agent based anomaly detection approaches to real-time intrusion detection. However, further testing is required to better understand their performance on real traffic data.

3. REFERENCES

[1] S. Marsland, U. Nehmzow, and J. Shapiro. A real-time novelty detector for a mobile robot. In *Proceedings of the European Advanced Robotics Systems Masterclass and Conference (EUREL)*, 2000.

[2] R. Saunders. *Curious Design Agents and Artificial Creativity.* PhD thesis, Faculty of Architecture, Design Science and Planning, University of Sydney, Sydney, 2001.

[3] J. Stanley. Computer simulation of a model of habituation. *Nature*, 261(5556):146–147, 1976.

[4] W. Wundt. *Principles of physiological psychology.* Macmillan, New York, 1910.