Impact of Packet Injection Models on Misbehaviour Detection Performance in Wireless Sensor Networks

Sven Schaust
Institute of Systems Engineering
G. W. Leibniz University of Hannover
Welfengarten 1, 30167 Hannover, Germany.
svs@sim.uni-hannover.de

Martin Drozda
Institute of Systems Engineering
G. W. Leibniz University of Hannover
Welfengarten 1, 30167 Hannover, Germany.
drozda@sim.uni-hannover.de

Helena Szczerbicka
Institute of Systems Engineering
G. W. Leibniz University of Hannover
Welfengarten 1, 30167 Hannover, Germany.
hsz@sim.uni-hannover.de

Abstract

Traffic in wireless sensor networks (WSN) is commonly created by sensor readings which can be modelled by various distributions as the occurrence of events triggers the injection of packets. The Poisson distribution is a common example for a widely used event distribution as many processes (for example the arrival of customers or the dissemination of parasits) are known to be Poisson distributed. The impact of such a packet injection model on sensor networks and especially on the performance of misbehaviour detection systems is therefore of interest. In this paper we investigate the impact of the Poisson model and the constant bit rate model on a misbehaviour detection system, namely an artificial immune system (AIS). We state the hypothesis that both models have no significant effect on the detection rate. We examine the influence of the two models on the detection performance and compare the results. We conclude that the differences between the two models show no statistically significant effects on the detection performance supporting our hypothesis. However, we observe that the AIS had a significantly smaller false positives rate for the Poisson model than for the CBR model.

1 Introduction

Traffic in wireless sensor networks (WSN) is commonly created by sensor readings which can be modelled by various distributions as the occurrence of events triggers the injection of packets. The Poisson distribution is a common example for an event distribution as many processes (for example the arrival of customers or parts for manufacturing) are considered Poisson distributed. The impact of such a packet injection model on the performance of misbehaviour detection systems is therefore an interesting issue, especially as the number of companies using such networks for logistics, production cycle maintenance, indoor security or area surveillance is rising. Companies cannot afford to accept losses due to malfunctioning systems, therefore such networks have to be able to detect any malfunction not only as fast as possible but also with high accuracy, which implies having only a low level ratio of false alarms and a high level ratio of true alarms. Sensor networks typically consist of battery powered devices with limited computational abilities and therefore using strong software cryptography to secure communication channels is almost impossible. Due to the latest developments of wireless communication protocols intended to be used with sensor networks (for example the work of the ZigBee® Alliance [17]) and the development of radio modules with support of hardware based AES-128 cryptography, unauthorised access to sensor networks is becoming more and more difficult. However sensor nodes are not tamper proof, leaving possibilities to gain access even to a secured network. As a result such networks still have to deal with possible node misbehaviour. We believe that artificial immune systems are capable of detecting misbehaviour within sensor networks independent from the current traffic model. AIS are basically a variety of anomaly detection systems derived from the human immune system. The advantage of artificial immune systems in contrast to protocol based security improvements is the universality of

1-4244-1455-5/07/$25.00©2007 IEEE

the approach. An AIS tries to identify anomalies which are not known to exist within the range of regular network behaviour. AIS are based on the *self* and *non-self* discrimination principle [1] and the continuous observation of the system. By using the *negative-selection* principle [12] an immune system is able to produce *detectors* which are able to recognise only *non-self* characteristics. Drozda et al. [6] evaluated the performance and usability of an artificial immune system approach within a simulated ad-hoc network using a constant bit rate (CBR) based traffic and concluded that AIS are indeed useful for sensor networks. In this paper we study the effects of the Poisson model (which better fits event triggered traffic) on an artificial immune system and compare to the results obtained when using the CBR model. We examine the hypothesis that both models will show no significant differences in respect to the AIS detection rate as the system uses only observed traffic information to conclude whether a node is misbehaving or not.

The paper is structured into the following sections: first a brief description of the functionality of an artificial immune system is given, followed by a description of our experimental setup and the parameters used. In section 4 the results obtained from the experiments and a comparison are shown and in section 5 a brief overview of related and prior work is given. Finally conclusions and future work are presented in section 6.

2 Artificial Immune System

Artificial immune systems are derived from the human immune system and therefore several terms and descriptions have been adopted from the biological perspective on immune systems, of which the most important term is *gene*. A *gene* measures network performance and thus is defined by a characteristic based on the data traffic volume, the assumed protocol behaviour or both (for example the number of complete MAC handshakes[1] during a specific time period). An *antigen* is an observation within a time window for a set of genes which can either be interpreted as an *self-antigen* or a *non-self-antigen*. Another important term is *detector*. A *detector* for AIS is a combination of different concepts based on known immune system methods and models which are related to the detection of alien cells. Detectors are defined as bit-sequences which are produced by a negative-selection algorithm [12] and finally should only match non-self antigens. A commonly used method to match antigens with detectors is the *r-contiguous bits matching rule* in which strings are equal if a common substring of length r at the same position p exists [8].

A simple artificial immune system can be divided into a learning phase and a detection phase [9]. During the learning phase detectors are produced (using the available self information) which will be used later in the detection phase to discover misbehaviour indicators. On every node an instance of the AIS is running, using the locally observed traffic to create a self set, a detector set and the antigens necessary to detect misbehaviour at its neighbours. There are several extensions which allow adaptive learning and maturation of detectors, thus avoiding the necessity of an intensive self-set only learning phase. See [10], [2] and [9] for more information on the different mechanisms and the immune system in general.

2.1 Learning phase

Part of the human immune system is the creation process of T-cells in B-macrophages and thymus which is partially adapted to the learning process in artificial immune systems. Using a pseudo-random process T-cells are created and censored by a negative-selection process. During this process T-cells which bind to self are destroyed. Similar to the biological selection detector strings within AIS are produced using a greedy strategy together with a pseudo-random generation process and tested against a self set [12].

It is possible that the selection process of an immune system produces detectors which are able to detect self, due to an incomplete self set during the selection process. Such detectors can cause auto-immune reactions or in terms of AIS *false positives*. A false positive is therefore defined as a node which is detected as misbehaving while it is actually performing within the regular range.

2.2 Detection phase

Using the censored T-cells the immune system is able to detect alien cells. In artificial immune systems antigens are created continuously for a specified time window (by analysing the network traffic) and matched against the complete detector set. If a matching detector is found, an immune response is triggered (for example a timed exclusion of the misbehaving node from routing paths). Detectors which are found to be useful are marked as mature and can cause an intense reaction at a lower activation threshold. In contrast, detectors which are newer or less effective need a higher activation threshold before causing any reactions. AIS offer the possibility to throw away detectors which do not prove to be useful, thus making space for new detectors. This however should only be necessary if the space for detectors is limited and the covering of the non-self set is insufficient. There are several enhancements to the detector maturation process using mutation or other evolutionary mechanisms. See [5] and [7] for more information on advances in artificial immune systems.

[1] A MAC handshake is defined by the sequence of RTS, CTS, DATA and ACK packets.

3 Experimental Setup

The purpose of our experiment was to measure the influence of the two packet injection models on the AIS detection performance, examining the hypothesis that an AIS offers a reasonable detection rate independent from the used packet injection model. The hypothesis is based upon the fact that injection models could only differ, with respect to the matching algorithm, in a total number of faults in an observed time window, thus their distribution on a time scale should therefore be not important. Similar to [9] a bit string representation was chosen for self, non-self and detectors. The matching rule was the *r-contiguous bits matching rule* with $r = 10$. Detectors were produced using a negative-selection strategy and tested against a priorly computed self set. We used two scenarios with 10 fixed but randomly chosen connections performing the CBR or the Poisson injection model. In each scenario we ensured that the average hop count distance between two nodes was about 8 hops. We chose these values as the underlying network topology showed no connection problems with routes of length 8. The number of connections was chosen as a tradeoff between a fast simulation and a reasonable network payload.

3.1 Scenario description

A Poisson and a CBR model using 10 connections were created. The observed network traffic was evaluated according to the described AIS approach. We have chosen the same set of genes as in [6]. They cover a good range of traffic properties which allows the AIS to detect misbehaviour. We captured for every transmitted packet the IP header type (UDP, 802.11 or DSR), the MAC frame type (RTS, CTS, DATA or ACK), the current simulation time, the node address, the next hop address, the global packet source, the global packet destination and the packet size. These values where used to compute the necessary genes as described below in section 3.5. Each scenario was simulated using Glomosim 2.03 (see [4]) 20 times with different seeds for the Glomosim random number generator. We distributed the simulation runs over 30 Linux based PCs (2 GByte RAM, Pentium4 3 GHz).

3.2 Topology

We used a network topology consisting of 1718 nodes which were placed in a 3000 $m \times 3000$ m square plane using a snapshot of a random waypoint walk. Each node was set to have a radio radius of $100m$. We made no restrictions to the graph connectivity, thus allowing isolated subgraphs. See reference [6] for a picture of the topology and the routing paths.

3.3 Misbehaviour

We implemented a simple packet dropping misbehaviour with a 10, 30 and 50% dropping probability (Sink and source nodes were excluded from misbehaviour). We configured 236 of our 1718 nodes to be malicious. Although the number of malicious nodes seems relatively high only one to three of them appeared per route as they were distributed randomly.

3.4 Simulation details

- **Negative selection algorithm:** random generation and testing. Implemented in C++, compiled with GNU g++ v4.0 with -O3 option.
- **Input parameters:** 1. r-contiguous bits matching rule with $r = 10$. 2. Encoding: 5 genes each 10 bits long = 50 bits. 3. Number of detectors $\{500, 1000, 2000\}$. 4. Misbehaviour level $\{10\%, 30\%, 50\%\}$ 5. Window size 500 seconds; 28 complete windows over 4-hours simulation time.
- **CBR Injection rate:** 1 packet/second. 14400 packets per connection were injected. Packet size was 512 bytes.
- **Poisson Injection rate:** $\lambda = 1.0$, meanArrivalExpectation = 1 packet/second. Packet size was 512 bytes.
- **Performance measures:** detection rate, false positives, data traffic rate at nodes; values were produced per simulation run and compared as arithmetic average with 95% confidence intervals over all simulations for each misbehaviour probability.
- **MAC protocol**: IEEE 802.11b DCF.
- **Routing protocol:** DSR.
- **Other parameters:** (i) Propagation path-loss model: two ray (ii) Channel frequency: 2.4 GHz (iii) Topography: Line-of-sight (iv) Radio type: Accnoise (v) Network protocol: IPv4 (vi) Connection type: UDP.

3.5 AIS details

When defining a misbehaviour detection system several observable factors have to be specified. For artificial immune systems these factors are defined by *genes*. We decided to observe two layers of the OSI Stack namely the MAC and Routing Layer using the following set of genes:

MAC Layer:

#1 Ratio of complete MAC layer handshakes between two communicating nodes s_i and s_{i+1} and the RTS packets sent by s_i to s_{i+1}. If there is no traffic between two nodes this ratio is set to ∞ (a large number). This ratio is averaged over a time period. A complete handshake is defined as a completed sequence of RTS, CTS, DATA, ACK packets between s_i and s_{i+1}.

#2 Ratio of data packets sent from s_i to s_{i+1} and then subsequently forwarded to s_{i+2}. If there is no traffic between two nodes this ratio is set to ∞ (a large number). This ratio is computed by s_i in promiscuous mode. This ratio is also averaged over a time period. This gene was adapted from the watchdog idea in [13].

#3 Time delay that a data packet spends at s_{i+1} before being forwarded to s_{i+2}. The time delay is observed by s_i in promiscuous mode. If there is no traffic between two nodes the time delay is set to zero. This measure is averaged over a time period. This gene is a quantitative extension of the previous gene.

Routing Layer:

#4 The same ratio as in #2 but computed separately for RERR routing packets.

#5 The same delay as in #3 but computed separately for RERR routing packets.

Each gene was encoded using an interval representation of size 10 which was adopted from [14]. The correspondending interval was marked by a single 1 within the 10 bit sequence. Antigens were produced by the concatenation of all five genes and always checked against the complete detector set.

4 Results

The task of detecting misbehaviour requires a comparison of all computed detectors with the observed non-self antigens. In our experiments a 500 second time window was used to sample node traffic and to generate one antigen. Thus the resulting number of time windows for 4 hours simulated times was 28 per node. In order to avoid outliers in our analysis we defined a detection threshold of 14 time windows to mark a node as misbehaving. The evaluation of the detection rate requires that the number of packets forwarded by a node exceeds a certain threshold. If a node lacks packets to forward in the learning phase, the AIS's ability to learn is limited. If it lacks packets to forward during the detection phase and at the same time wants to execute misbehaviour, the impact of misbehaviour is weakened. As a result we therefore performed our evaluation using different forwarding threshold values for packet forwarding (minValue = 500, 1000, 2000 and 4000) and considered only those nodes which were above the given thresholds.

Definition: The *detection rate* is defined as $d_r = \frac{n_d}{n_m}$, where n_m = the number of misbehaving nodes to detect, and n_d = the number of correctly detected nodes.

Definition: The *false positives rate* is defined as $fp_r = \frac{n_{fp}}{n_d+n_{fp}}$, where n_{fp} = the number of incorrectly detected nodes, and n_d = the number of correctly detected nodes.

We expected the detection rate and the false positives rate to be similar for both models. The graphs in figure 1 show the average detection[2] results and the deviation ratio for CBR and Poisson packet injection using 500 detectors. While the detection rate is about 10% higher for the Poisson packet injection with a low packet threshold of 500 packets, the ratio for the CBR model seems to be better for higher thresholds. In order to verify the hypothesis that both models show no significant differences we calculated a 95% confidence interval for all three misbehaviour probabilities and evaluated the deviation ratio for both models.

Definition: The *deviation ratio* is defined as $dv_r = \frac{d_v}{d_r}$, were d_v = the deviation value of the 95% confidence interval, and d_r = the detection rate for a specific misbehaviour.

Both CBR and Poisson model show a similar deviation range, with the values for the Poisson model being slightly lower, suggesting that the Poisson model results in a better performance than the CBR model. However the deviation ratio (shown in figure 1 (c) and (d)) for both models is similar. We therefore conclude that the AIS detection performance is indeed not influenced by the chosen traffic model as the detection rate values for the CBR and Poisson model are within in the range of the calculated deviation ratio. We also computed the detection rate and the false positives rate for 2000 detectors (see figure 2). Similar to the results with 500 detectors the Poisson model shows a detection rate which is about 2% to 10% higher than the rate of the CBR model. Both models show again a similar deviation ratio supporting the hypothesis that the packet injection model has no significant impact on the detection rate. However the false positives rate is up to 45% lower (worst case) as in the CBR experiment, which is an interesting result (see figure 2 (c), (d)). We expected the false

[2] We used the arithmetic average over all simulation runs to calculate the detection rate.

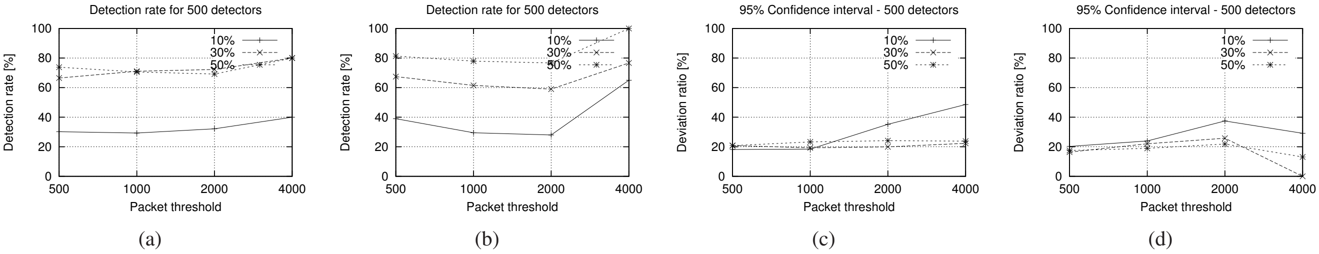

Figure 1. Detection rate for CBR (a) and Poisson (b), deviation ratio for CBR (c) and Poisson (d) for 500 detectors.

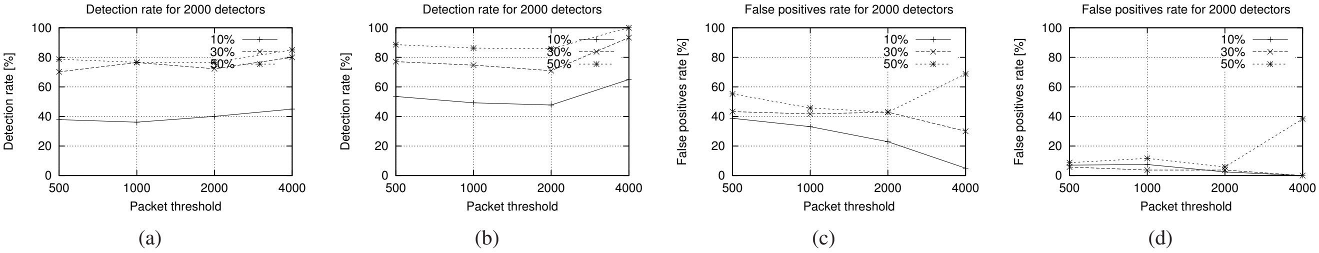

Figure 2. Detection rate for CBR (a) and Poisson (b), false positives rate for CBR (c) and Poisson (d) for 2000 detectors.

positives rate to be at most 10% better than with the CBR model since we expected the models to have a similar influence on both the detection rate and the false positives rate. The results for the Poisson model suggest that due to the different packet injection model and the resulting different arrival time gaps, the observed traffic during the detection phase differed more in cases of misbehaviour and hence looked more similar to the detectors of the detector set. This assumption is based upon the fact that false positives are described by antigens which have to be *similar* to at least one detector and therefore become detected by the r-contiguous bits matching rule. Hence the real malicious antigens were more dissimilar to normal antigens as in the CBR experiment.

5 Related Work

Hofmeyr and Forrest presented in 1999 a paper about the usage of an immune system inspired misbehaviour detection system. Their artificial immune system was designed to work on a wired network observing TCP/IP connections. The pattern matching was based on a r-contiguous bits matching with $r = 12$ bits. The antigen and detector length was 49 bits [9].

Sarafijanovic and Boudec presented in [14] and [15] an artificial immune system based misbehaviour detection system designed for wireless ad-hoc networks. The approach was tested using Glomosim to simulate a network of 40 mobile nodes (1 m/s) of which 5 to 20 nodes were misbehaving. They defined four genes to be used to capture local behaviour at the OSI network layer. The detection rate of the presented system was about 55%. They also used a *danger signal* which allows nodes to inform neighbours on the routing path about misbehaviour. This interaction was adopted from the results by Aickelin et al.[3].

Aickelin et al. have been working on artificial immune systems since 2003 in an interdisciplinary project called *danger theory*. In [2] and [3] they introduced work showing links between intrusion detection systems and artificial immune systems. They also introduced a danger signal approach allowing nodes to judge the misbehaviour information and presented work on adaptive learning mechanisms.

Drozda et al. [6] showed that artificial immune systems can be applied to sensor networks having only low computational costs. The detection rate of the introduced AIS is higher than the one presented by Sarafijanovic and Boudec, as the authors use a static ad-hoc network of 1718 nodes with 10 CBR connections instead of a mobile network.

To our knowledge the Poisson traffic model and the impact of injection distributions on the AIS performance have not been studied before.

6 Conclusions

In this paper we examined the influence of two different packet injection models on the misbehaviour detection performance and provided the hypothesis that the two different packet injection models have only a small influence on the detection system. We conclude that the measured differences of the examined traffic models are not statistically significant and therefore conclude that the hypothesis is true. We are glad to observe that AIS are indeed capable of handling different traffic models. For both models the AIS accomplished a detection ratio above 70% for the misbehaviour rates 30% and 50%. The Poisson based false positives rate was unexpectedly up to 45% lower (worst case) than the CBR based rate, which is an interesting result. We are going to investigate the reason for that further.

We are currently working on a simulation experiment using 50 (100) fixed but randomly chosen connections with a different variety of misbehaving nodes. We are going to investigate more packet injection models and parameters in order to test the AIS using different, more complex attack patterns. The intention of these experiments is to verify which *genes* should be part of an AIS in general and which are independent from the tested packet injection models and are therefore providing good detection results.

Additionally we are working on a sensor node implementation based on Crossbows [16] mica2 mote not only to prove that an AIS is indeed preferable for sensor networks but also to test and verify the simulation results using realistic data collection scenarios.

In our experiments we did not use any advanced clustering methods for data evaluation, therefore these results could still be significantly improved.

Acknowledgements

This work was supported by the German Research Foundation (DFG) under the grant no. SZ 51/24-1 (Survivable Ad Hoc Networks – SANE).

References

[1] S. Forrest, S. A. Hofmeyr and A. Somayaji. A sense of self for unix processes. *In Proceedings of the IEEE Symposium on Research in Security and Privacy*, Los Alamitos, 1996.

[2] U. Aickelin, J. Greensmith, J. Twycross. Immune System Approaches to Intrusion Detection - A Review. *Proc. the 3rd International Conference on Artificial Immune Systems (ICARIS)*, 2004.

[3] U. Aickelin, P. Bentley, S. Cayzer, J. Kim, J. McLeod. Danger theory: The link between ais and ids. *Proc. International Conference on Artificial Immune Systems (ICARIS)*, 2003.

[4] L. Bajaj, M. Takai, R. Ahuja, K. Tang, R. Bagrodia, M. Gerla. GloMoSim: A Scalable Network Simulation Environment. UCLA Computer Science Department Technical Report 990027, May 1999.

[5] D. Dasgupta, F. Gonzalez. An immunity-based technique to characterize intrusions in computer networks. *IEEE Trans. Evolutionary Computation*, vol. 6, no. 3, pp. 281–291, 2002.

[6] M. Drozda, S. Schaust, H. Szczerbicka. Is AIS Based Misbehavior Detection Suitable for Wireless Sensor Networks? *Proc. IEEE Wireless Communications and Networking Conference (WCNC)*, 2007.

[7] M. Drozda, H. Szczerbicka. Artificial Immune Systems: Survey and Applications in Ad Hoc Wireless Networks. *Proc. 2006 International Symposium on Performance Evaluation of Computer and Telecommunication Systems (SPECTS)*, 2006.

[8] P. Helman and S. Forrest. An efficient algorithm for generating random antibody strings. *Technical Report CS-94-07*, University of New Mexico, Albuquerque, NM, 1994.

[9] S. Hofmeyr, S. Forrest. Immunity by Design: An Artificial Immune System. *Proc. Genetic and Evolutionary Computation Conference (GECCO)*, 1999.

[10] Charles A. Janeway Jr. How the immune system works to protect the host from infection: a personal view. *Proc. Natl. Acad. Sci. U S A.*, 2001 Jun 19;98(13):7461-8.

[11] D. Johnson, D. Maltz. Dynamic Source Routing in Ad Hoc Wireless Networks. *Mobile Computing*, Tomasz Imielinski and Hank Korth, Eds. Chapter 5, pp. 153-181, Kluwer Academic Publishers, 1996.

[12] J. Kim, P.J. Bentley. Evaluating Negative Selection in an Artificial Immune System for Network Intrusion Detection, *Proc. Genetic and Evolutionary Computation Conference (GECCO)*, 2001.

[13] S. Marti, T. J. Giuli, K. Lai, M. Baker. Mitigating routing misbehavior in mobile ad hoc networks. *Proc. the 6th annual international conference on Mobile Computing and Networking (MobiCom)*, 2000.

[14] S. Sarafijanović, J.-Y. Le Boudec. An Artificial Immune System for Misbehavior Detection in Mobile Ad-Hoc Networks with Virtual Thymus, Clustering, Danger Signal and Memory Detectors. *Proc. the 3rd International Conference on Artificial Immune Systems (ICARIS)*, 2004.

[15] J.-Y. Le Boudec, S. Sarafijanović. An Artificial Immune System Approach to Misbehavior Detection in Mobile Ad-Hoc Networks. *Proc. Bio-ADIT'04*, 2004.

[16] Crossbow Technology Inc. `www.xbow.com`

[17] ZigBee Alliance® `www.zigbee.org`