Dynamic Machine Learning Algorithm for AODV Routing Attacks Detection

Md Raqibull Hasan[1], Yanxiao Zhao[2], Guodong Wang[3], Yu Luo[4], and Lina Pu[5]

[1] Contract Professionals Incorporation, Waterford Twp, MI 48329.
[2] ECE Department, Virginia Commonwealth University, Richmond, VA 23284.
[3] CS Department, Massachusetts College of Liberal Arts, North Adams, MA 01247.
[4] ECE Department, Mississippi State University, Mississippi State, MS 39762.
[5] School of CSCE, University of Southern Mississippi, Hattiesburg, MS 39406.
Email:mdraqibull.hasan.sdsmt@gmail.com,yzhao7@vcu.edu,guodong.wang@mcla.edu,yu.luo@ece.msstate.edu,lina.pu@ua.edu

Abstract. Ad hoc On-Demand Distance Vector (AODV) routing protocol is vulnerable to some routing attacks including blackhole attack and flooding attack. Typically, these two types of routing attacks are linked with two major malicious behaviors: fake Route Replies (RREPs) and fake Route Request (RREQ) flooding. In this paper, we develop a novel dynamic machine learning approach to detect blackhole and flooding attacks in AODV. The proposed solution primarily determines three distinct features by analyzing Hello, RREQ, and RREP packets in the AODV routing protocol. Then, these features are used to develop a mathematical model for the dynamic learning algorithm. Afterward, we generate the training set of data and assign a threshold for our machine learning model using these features. This training set of data is only valid for N time slots, which is regarded as one iteration. In the following iterations, it will update the latest valid outcomes from the dynamic learning model and determine an updated threshold for the model, which significantly increases the detection accuracy. Extensive simulations have been conducted to evaluate the accuracy and the time overhead of three classifiers, e.g., support vector machine, k-nearest neighbor, and decision tree. The simulation results show that the proposed algorithm can achieve very high accuracy with minimum time overhead to detect malicious behavior in the AODV routing protocol.

Keywords: AODV routing protocol, Routing attack, Smart meter network, Dynamic machine learning

1 Introduction

Smart meter networks are the major components of the smart grid, which are composed of smart meters and Data Aggregation Points (DAPs) [1]. A smart meter is an electronic device that is installed at houses or commercial sites to record consumption of electric energy and communicates that information

EAI MOBIMEDIA 2020, August 27-28, Harbin, People's Republic of China

DOI 10.4108/eai.27-8-2020.2297866

back to the utility for monitoring and billing. Each smart meter equipped with network radio can transmit meter reading periodically or on request by utilities. DAPs are responsible for communicating information or data between smart meters and the utility company.

Since a smart meter network is essentially a multi-hop network, this means some smart meters serve as relay nodes to deliver information to the DAP through multiple hops. Similar to a typical multi-hop network, multiple routes exist in a smart meter network. In other words, there is more than one path between a smart meter and its associated DAP. Therefore, a routing protocol is required to find the appropriate route to DAPs for each smart meter.

In recent years, Ad hoc On-Demand Distance Vector (AODV) routing protocol is recommended for smart meter networks because it meets both the requirements of on-demand and periodic operations. However, the AODV routing protocol is vulnerable to different types of routing attacks. One typical routing attack is the denial of service including blackhole attack and flooding attack. In a blackhole attack, attackers broadcast numerous fake Route Replies (RREPs) with a tampered highest sequence number and minimum hop count. If these fake replies are accepted by other smart meters, blackholes will be created in a smart meter network. Consequently, legitimate data cannot be sent to DAPs. In a flooding attack, attackers broadcast Route Requests (RREQ) packages throughout the network. The entire network keeps busy to forward those fake RREQ packets. Under this attack, the data transmission can be delayed, corrupted, or even blocked. Therefore, how to effectively conduct routing attack detection and detect malicious meters in a smart meter network remains a critical challenge.

There are two problems in the existing work for AODV routing attack detections. The first problem is that the majority of work usually focuses on specific and known routing attacks, e.g., blackhole and flooding detection. It is known that different types of attacks exhibit distinctly. Most approaches detect malicious behaviors by targeting specific anomalous behaviors by adding new control packets to routing protocols. These approaches are not able to detect other types of attacks if present. The second problem is that a constant threshold is commonly used for detection. If an attacker is able to estimate the threshold, it can easily break the intrusion detection system and utilize the routing information to access data packets [2].

In this paper, we introduce a novel dynamic machine learning approach. First, we analyze the complete AODV routing packets including Hello, RREQ, and RREP control packets. Based on this analysis, we obtain three distinct features in the AODV routing protocol. Afterward, these features are utilized to develop a mathematical model of dynamic learning algorithm. Concurrently, these features are also considered in our machine learning model to correlate with the testing data as well as to generate an initial training set of data and a threshold. Every training set of data is only valid for N time slots, which is considered as one iteration. During this N time slots, our dynamic learning model will create new data points, which will be updated in the training set of data and be utilized to determine a new threshold for the next iteration. Therefore, during every N slot

intervals, the training data set will be updated based on the current state of the smart meter network and create a new threshold to detect malicious behaviors. The most promising part is that our proposed machine learning model updates parameters according to the changes of malicious behaviors, which achieves a high detection accuracy. The main contributions of this research are summarized below:

First, we develop formulas by analyzing AODV routing packets including Hello, RREQ, and RREP control packets. Leveraged by our proposed formulas, we obtain three distinct features including average one-hop neighbor distance, the dynamic range of sequence number, and the minimum hop count to analyze entire AODV routing packets.

Second, we derive a mathematical model for dynamic learning algorithm using the three identified features. Initially, these features are placed in a three-dimensional vector space. Therefore, it determines the N number of vectors for N time slots where all of them are considered in the first iteration. Then, we calculate the mean vector from N time slots. Finally, we find out the variance between input sample data and mean vector. If this variance is lower or equal to the current threshold, the input sample data will be considered as normal traffic. This data will be further forwarded to the training model in order to estimate a new threshold for the next iteration. In contrast, if the variance of sample data exceeds the current threshold limit, it will be identified as malicious traffic.

Third, we develop an adaptive machine learning model, where the training set of data is updated and a new threshold is calculated after N time slots interval. In the following iterations, the training model will only update the recent valid inputs of the previous iteration from the dynamic learning model. During the N time slots, the dynamic learning model will calculate the variance of each incoming data and pull the threshold from the memory block simultaneously. Afterward, these two data sets along with input features are evaluated using three classifiers including Support Vector Machine (SVM), k-Nearest Neighbors (k-NN), and Decision Trees (DTs) algorithm.

The rest of paper is organized as follows. Section 2 introduces related work. Section 3 introduces feature identification from AODV routing. Section 4 designs the mathematical model for the proposed dynamic machine learning method. Section 5 shows simulations results and Section 6 draws conclusions.

2 Related Work

In this section, we summarize the state-of-the-art for routing attack detection in AODV under smart meter network, wireless sensor network, or mobile ad hoc network. The followings are some recent work to tackle routing attacks in the AODV routing protocol using machine learning approaches or modified AODV routing mechanisms.

A very recent work [3] conducted a comprehensive survey on various attacks in AODV routing under MANETs. It also introduced some prospective techniques for detecting and predicting routing attacks, which are data mining,

SVM, genetic algorithms (GA), and some other machine-learning approaches. Machine learning and data mining methods for cyber analytic were investigated in [4] for intrusion detection. This work also analyzed the complexity of machine learning algorithms and discussed challenges for cyber-attack defense. In [5], a machine learning-based intrusion detection system was developed to protect critical infrastructures. Among various supervised machine learning classification techniques [6], the k-NN classification algorithm was utilized in wireless sensor networks to separate the anomalous node based on abnormal behaviors [7]. The authors also analyzed the relevant parameter selection and error rate of the intrusion detection system for AODV routing. In addition, an enhanced SVM for packet classification was proposed in [8] to provide unsupervised learning with low false alarm capability.

In [9], the authors proposed a new architecture for intrusion detection in mobile ad hoc networks using the machine learning approach to maximize detection accuracy. In this work, rough set and SVMs have been used for data reduction and classification respectively. The rough set reduces the size of features to simplify the complexity of SVM. In the following work, [10], a novel supervised learning framework was proposed by using a generative adversarial network for improving the performance of the classifier. This framework was utilized to continuously generate other complementary labeled samples for adversarial training and assisting the classification. In addition, several empirical training strategies were proposed to improve the stabilization of the supervised learning framework.

In the cluster wireless sensor network, a beta distribution based dynamic trust management has been proposed in [11]. The proposed method dynamically calculated the reputation and adopted a dynamic threshold to resist the on-off attack, bad-mouth attack, selective forwarding attack, and a mixed attack. In [12–14], some of the recent works have been emphasized on detection and prevention algorithms either for blackhole or flooding attacks under AODV routing. Specifically, in [14], a secure and lightweight routing protocol was proposed to prevent blackhole attacks in constraint-oriented networks. The proposed protocol is a hybrid of medium access control and AODV protocols. In this work, every node was registered with the nearest gateway/cluster head module through the MAC addressing scheme, and only registered nodes were allowed to communicate.

In [15], a logical scheme was proposed to tackle some common cyber-attacks in the smart grid. The hierarchy of the proposed system consisted of three remote terminal units (RTUs), a substation, and a control center, which communicated in two-way data flow in real time scenario. All the critical information of the smart grid, like the voltage, frequency, and voltage angle was encrypted through MD5 hash algorithm and later decrypted at the substation and control center using a key authentication method. In [16], the authors introduced the background of Advanced Metering Infrastructure (AMI) and identified major security requirements in AMI. Specifically, this work illustrated the energy-theft behaviors in AMI using an attack tree-based threat model.

3 Features Identification of AODV routing

As we mentioned previously, RREQ flooding and fake RREPs from blackhole attack are the major problems in AODV. To detect and prevent these attacks, we present a dynamic learning method in this section. In particular, three distinct features are obtained by analyzing Hello Packet, RREQ, and RREP in the AODV routing protocol. These three features are average one-hop neighbor distance, the dynamic range of sequence number, and minimum hop count. We will introduce how to determine these features as follows.

3.1 Calculate One-hop Neighbor Distance

In the AODV routing protocol, each smart meter sends Hello Packets to all of its one-hop neighbors before initiating control packets (RREQ and RREP) to find out the appropriate routes.

During Hello packets communications, smart meters calculate the distance of its one-hop neighbors based on the power label of received Hello Packets. The following equation is used to calculate the one-hop neighbor distance [17]:

$$P = \frac{4\pi D}{0.12476} \times 10^{-12.5}. \tag{1}$$

After that, each smart meter estimates the average one-hop neighbor distance, using the following equation:

$$D_n = \frac{\sum_{i=1}^{n} D_i}{n}. \tag{2}$$

The average one-hop neighbor distance is regarded as the first distinct feature because Hello Packet communication happens before starting the actual routing packets. Using this method, a malicious attacker is being deprived of this information.

3.2 Ranges of Sequence Numbers

In the AODV routing protocol, every routing search deals with two sequence numbers during control packet communications. Initially, a source node initiates the RREQ control packet and includes a sequence number for the desired destination. This sequence number is also named as RREQ sequence number. After receiving RREQ by a destination, the destination acknowledges the source by sending the RREP control packet. This RREP control packet also includes a sequence number. Note that, at this time, the second sequence number is updated and is completely different from the previous one. Therefore, there is a difference between two sequence numbers for every routing setup. In other words, a range of sequence numbers exists for each valid routing, which is predefined by destination. The dynamic range of sequence numbers is regarded as the second distinct feature, which is denoted by S_d in our machine learning model.

3.3 Calculate the Minimum Possible Hop Count

For a smart meter network, every smart meter knows its own and its destination location. Based on the locations, it can predict the distance between source and destination. Under blackhole attacks, each source meter receives multiple RREPs, where fake replies always contain the minimum hop count. To avoid flows of fake replies, each smart meter calculates the minimum possible hop count, H_{min}, from the predicted distance, P_{dis} and the average one-hop neighbor distance, D_n. The equation (3) is used to determine the minimum hop count.

$$H_{min} = \frac{P_{dis}}{D_n} \tag{3}$$

The minimum possible hop count is regarded as the third distinct feature in our dynamic machine learning model. Using this method, we can stop blackhole attackers from pretending one-hop neighbor in its fake reply.

4 Mathematical Model of Dynamic Machine Learning Method

In this section, we design a mathematical model for the proposed dynamic machine learning method. Each smart meter collects three above-mentioned distinct features, which are considered in a three-dimensional vector space, $x_i = (x_{i1}, x_{i2}, x_{i3})$ for i^{th} time slot. Here, each time slot contains a certain Active Routing Time (ART) and a Hello packet communication with one-hop neighbors. For N time slots, we calculate the mean vector of x using the equation (4).

$$\overline{x} = \frac{1}{N} \sum_{i=1}^{N} x_i. \tag{4}$$

Next, we calculate the distance from input data sample x, to the mean vector from equation (5)

$$d(x) = ||x - \overline{x}||^2. \tag{5}$$

If the distance is larger than the threshold T_h ($d(x) > T_h$), that means it is out of range from normal traffic, so it will be regarded as an attack. Here, the projection distance with its maximum value is extracted from the learning data set:

$$T_h = d(x_I), where\ I = \arg\max_i d(x_i), x_i \in D \tag{6}$$

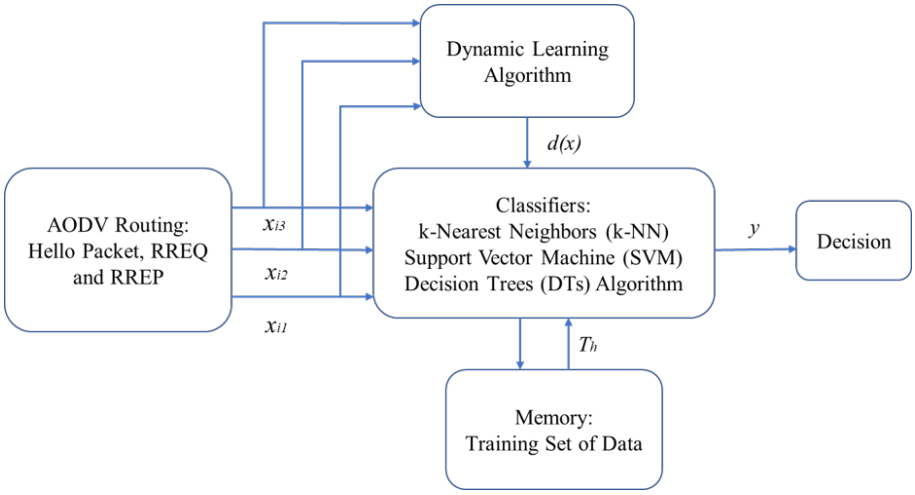

Fig. 1. Intrusion detection in AODV using dynamic machine learning algorithm.

By implementing this mathematical model in AODV routing protocol, we calculate the threshold T_h, and generate training data for our machine learning model. This threshold is only valid for N time slots. After that, it will update the threshold based on current network scenarios. Therefore, each smart meter can operate in a dynamic range of threshold. Now, for every following routing, Hello Packet, RREQ and RREP routing information are analyzed in a dynamic learning model to determine $d(x)$. The calculated $d(x)$ and its corresponding three distinct features are considered as testing data for our machine learning model, where three classifiers are employed including SVM, k-NN and DTs Algorithm. The complete flow chart of our proposed model is depicted in Fig. 1.

5 Simulation Results for Dynamic Learning Method

In our simulation, we consider three distinct features to identify routing attacks in AODV routing protocol, e.g., average one-hop neighbor distance during Hello packet communication, dynamic range of sequence number for each routing, and the minimum hop count. We generate all possible combinations of malicious data and normal data. Using training and cross-validation data, we test three default classifiers (SVM, k-NN and DTs Algorithm) under Python 3.6 Skit-learn module. Since we are dealing with supervised machine learning, accuracy and time overhead are two main performance metrics to evaluate the classifier.

The accuracy of each classifier represents the percentage of detection with our predefined and future randomized data. As shown in Fig. 2, SVM fluctuates its accuracy with the increase of independent variables C as regularization parameter. In contrast, k-NN stabilizes its accuracy around 87%. The time overhead is depicted in Fig. 3, where we find that k-NN produces almost double delay (above

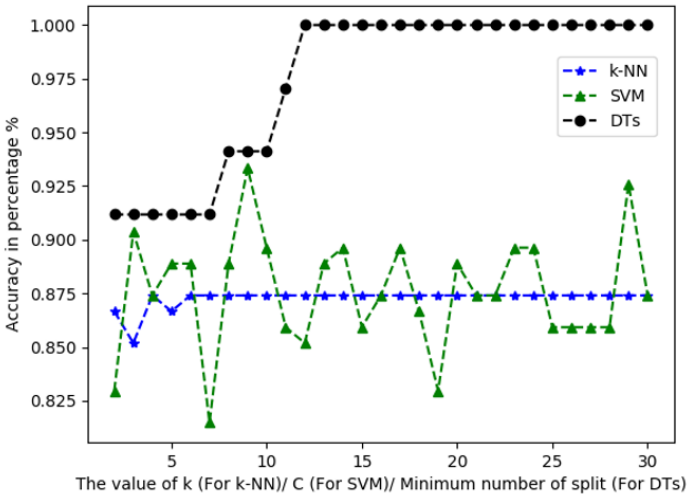

Fig. 2. Accuracy comparison for three classifiers.

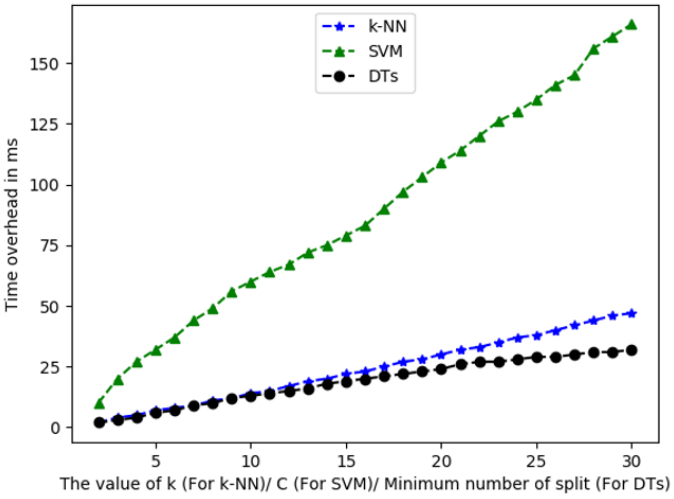

Fig. 3. Time overhead comparison for three classifiers.

135ms) compared to SVM (around 70ms). To address the trade-off between time overhead and accuracy, we introduce a DTs algorithm. As shown in Fig. 2, the DTs algorithm obtains the maximum accuracy (100%) in a 12-minimum splits. In addition, the time overhead as shown in Fig. 3 achieves almost half (around 35ms) of the k-NN. Therefore, the DTs algorithm shows the maximum fitness to identify the routing attacks in AODV routing protocol.

6 Conclusions

In this paper, we proposed a novel dynamic machine learning approach to detect malicious behavior in the AODV routing protocol. To the best of our knowledge, the proposed solution is the first work that combines both the dynamic learning algorithm and the machine learning approach. We mainly focused on two major malicious behavior detection, e.g., RREQ flooding attack and blackhole attack. To detect those malicious behaviors, we developed a dynamic learning algorithm along with the machine learning model. The dynamic machine learning approach is implemented in Python 3.6 using the Scikit-Learn module. Three classifiers were used to evaluate the accuracy and time overhead of the proposed solution. Among those three classifiers, the DTs algorithm achieves 100% accuracy with minimum overhead (around 35ms) to detect the malicious behaviors in the AODV routing protocol.

References

1. M. Lee, O. Aslam, B. Foster, D. Kathan, J. Kwok, L. Medearis, R. Palmer, P. Sporborg, and M. Tita, "Assessment of demand response and advanced metering," *Federal Energy Regulatory Commission, Tech. Rep*, 2013.
2. M. R. Hasan, Y. Zhao, G. Wang, Y. Luo, L. Pu, and R. Wang, "Supervised machine learning based routing detection for smart meter network," in *12th EAI International Conference on Mobile Multimedia Communications, Mobimedia 2019*. European Alliance for Innovation (EAI), 2019.
3. M. S. Hussain and K. U. R. Khan, "A survey of ids techniques in manets using machine-learning," in *Third International Conference on Computational Intelligence and Informatics*. Springer, 2020, pp. 743–751.
4. A. L. Buczak and E. Guven, "A survey of data mining and machine learning methods for cyber security intrusion detection," *IEEE Communications surveys & tutorials*, vol. 18, no. 2, pp. 1153–1176, 2015.
5. L. Cazorla, C. Alcaraz, and J. Lopez, "Towards automatic critical infrastructure protection through machine learning," in *International Workshop on Critical Information Infrastructures Security*. Springer, 2013, pp. 197–203.
6. S. B. Kotsiantis, I. Zaharakis, and P. Pintelas, "Supervised machine learning: A review of classification techniques," *Emerging artificial intelligence applications in computer engineering*, vol. 160, pp. 3–24, 2007.
7. W. Li, P. Yi, Y. Wu, L. Pan, and J. Li, "A new intrusion detection system based on knn classification algorithm in wireless sensor network," *Journal of Electrical and Computer Engineering*, vol. 2014, 2014.

8. T. Shon and J. Moon, "A hybrid machine learning approach to network anomaly detection," *Information Sciences*, vol. 177, no. 18, pp. 3799–3821, 2007.
9. T. Poongothai and K. Duraiswamy, "Intrusion detection in mobile adhoc networks using machine learning approach," in *International Conference on Information Communication and Embedded Systems (ICICES2014)*. IEEE, 2014, pp. 1–5.
10. Y. Z. S. L. J. F. Yin, Chuanlong and H. Zhang, "Enhancing network intrusion detection classifiers using supervised adversarial training," *The Journal of Supercomputing*, pp. 1–30, 2019.
11. Y. Sun and Y. Zhao, "Dynamic adaptive trust management system in wireless sensor networks," in *5th International Conference on Computer and Communications (ICCC)*. IEEE, 2019, pp. 629–633.
12. M. R. Hasan, Y. Zhao, G. Wang, Y. Luo, and R. M. Winter, "Enhanced aodv: Detection and avoidance of black hole attack in smart meter network," in *2017 26th International Conference on Computer Communication and Networks (ICCCN)*. IEEE, 2017, pp. 1–6.
13. M. R. Hasan, Y. Zhao, Y. Luo, G. Wang, and R. M. Winter, "An effective aodv-based flooding detection and prevention for smart meter network," *Procedia Computer Science*, vol. 129, pp. 454–460, 2018.
14. R. K. M. A. A. M. A.-Z. M. Z. M. S. A. Adil, Muhammad and R. Ahmed, "Mac-aodv based mutual authentication scheme for constraint oriented networks," *IEEE Access*, vol. 8, pp. 44 459–44 469, 2020.
15. V. M. A. A. S. P.-R. P. U. M. Singh, Neeraj Kumar and M. Bhatt, "Identification and prevention of cyber attack in smart grid communication network," in *International Conference on Information and Communications Technology (ICOIACT)*. IEEE, 2019, pp. 5–10.
16. R. Jiang, R. Lu, Y. Wang, J. Luo, C. Shen, and X. S. Shen, "Energy-theft detection issues for advanced metering infrastructure in smart grid," *Tsinghua Science and Technology*, vol. 19, no. 2, pp. 105–120, 2014.
17. E. P. GRID, "Utility-scale smart meter deployments," *IEI Report*, 2014.