ML-IDS: A Machine Learning Approach to Detect Wormhole Attacks in Internet of Things

Prachi Shukla
Dept. of Electrical Engineering
Columbia University
New York, NY 10027
ps2829@columbia.edu

***Abstract*—Internet of Things (IoT) is an emerging business model aimed to connect many low-power embedded devices with internet. IoT has many applications in building smart cities, smart environment, e-health care, etc. Due to the presence of unreliable internet and new routing protocols for low-power devices, IoT requires innovative security solutions. In this paper, we present three new Intrusion Detection Systems (IDSs) for IoT: 1) K-means clustering unsupervised learning based IDS; 2) decision tree based supervised IDS; and 3) a hybrid two stage IDS that combines K-means and decision tree learning approaches. To the best of our knowledge, these are the first machine learning based IDSs (together called as ML-IDS) for IoT. All the three IDS are centralized and scalable approaches. The K-means approach achieves 70-93% detection rate for varying sizes of random IoT networks. Decision tree based IDS achieves 71-80% detection rate and the hybrid approach attains 71-75% detection rate for the same network sizes. Although the hybrid IDS obtains lower detection rate, it is more accurate than the other two approaches. The hybrid approach eliminates the false positives significantly, while the other two IDS suffer from a higher number of false positives. Similar results are also obtained for regular mesh, star and ring topologies of IoT networks, each comprising 16 nodes.**

***Keywords*—Internet of Things (IoT); security; machine learning; intrusion detection systems**

I. Introduction

Internet of Things (IoT) is an innovative concept of connecting thousands of low-power embedded devices with each other and with the global internet. Some of the major applications of IoT include: 1) *smart cities* - monitoring parking spaces, traffic congestion, managing street lighting; 2) *smart environment* - forest fire detection, control of CO2 emissions; and 3) *smart industrial control* - monitoring air quality, temperature; and 4) *eHealth* - monitoring patient conditions in hospitals and elderly homes, UV radiation detection, etc. With the above applications, the IoT exchanges vast amounts of critical and private data, whose security if compromised can lead to severe economic consequences as well as threaten human lives [1].

IoT is different from traditional internet in the sense that it uses a compressed version of the *IPv6* protocol called *6LoWPAN* (Low-power Wireless Personal Area Network) in the network layer. 6LoWPAN uses a small packet size and has a low bandidth of 20-250 kbps. The 6LoWPAN networks also use a new, recently standardized routing protocol for low-power and lossy networks (RPL) [2]. In the application layer, 6LoWPAN network uses a Constrained Application Protocol (CoAP) instead of widely-used connection-oriented web protocol like HTTP, which is infeasible for IoT [3]. Finally, at the link layer, IoT uses IEEE 802.15.4 protocol, which is a low-cost, low-speed and low-power wireless personal area network, compared to 802.3/802.11 standards for traditional internet.

There are major security concerns associated with 6LoWPAN networks. These networks are directly connected to the unreliable and untrusted internet. Therefore, an attacker can get access to the IoT devices through the internet. Moreover, attacks can also originate within the 6LoWPAN networks due to one or more compromised devices. A recent example of a security breach of an IoT system is the compromise of Google Nest Thermostat, where the attackers used a modified boot sequence to insert arbitrary payloads into the device [4].

Wormhole attacks are one of the most difficult attacks to prevent or detect. There has been a significant amount of research in detecting these attacks in the traditional internet [5], [6]. During this attack, a high-speed wormhole tunnel is created between two distant compromised routers. This tunnel is then used to modify the routing behavior of the network by sending most of the traffic through the tunnel. This change in routing behavior is because the other routers connected to the victims can find the shortest path through this tunnel. The net effect is that the routers, which are very far apart also appear to be neighbors. The high-speed tunnel can be created using low-latency wired links or high-power wireless links. RPL based 6LoWPAN networks have also been shown to be affected by these wormhole attacks [7].

In this paper, we introduce novel lightweight Intrusion Detection Systems (IDS) for RPL based 6LoWPAN networks to detect wormhole attacks in IoT. We introduce three new machine learning based IDS for IoT: 1) K-means clustering unsupervised learning based IDS (KM-IDS); 2) a supervised decision tree based IDS (DT-IDS); and 3) a two-stage hybrid IDS that combines both K-means clustering and decision tree based approach to detect wormhole attacks in IoT. *To the best of our knowledge, these are the first machine learning based approaches to detect wormhole attacks in IoTs*. Together, these IDS are called ML-IDS.

The KM-IDS is a centralized approach, where K-means clustering is used to divide the router nodes of a 6LoWPAN network into several clusters. Each cluster is called a *safe zone*, which is a new metric that we introduce to measure how far a router can communicate without involving any malicious

wormhole tunnel. A wormhole attack is detected if a router tries to add new neighbors outside its safe zone.

The DT-IDS is also a centralized approach, where the IDS uses training data to learn the *safe distance* between any two neighboring routers. After learning, the IDS can detect a wormhole attack if the routers requesting to become neighbors are more than *safe distance* apart. This IDS creates a decision tree during the learning phase, which is then used to decide if any two routers can be made neighbors or are victims of attack.

Finally, we also introduce a two-stage hybrid IDS, which first applies K-means clustering to identify the safe zones and detect wormhole attacks. K-means is followed by the use of decision tree in the second stage that filters out some of the false positives, providing more accurate detections. KM-IDS is a very conservative approach, where any connection request between the safe zones is seen as a potential attack. In contrast, the use of decision tree as a second stage in hybrid IDS, provides more flexibility and allows some legal connection requests to be granted, reducing the number of false positives.

We also performed an exhaustive experimental analysis on varying sizes of random IoT networks as well as fixed size regular IoT topologies. The K-means approach achieves 70-93% detection rate. Decision tree based IDS achieves 71-80% detection rate and the hybrid approach attains 71-75% detection rate for the same network sizes. Although the hybrid IDS obtains lower detection rate, it is more accurate than the other two approaches. The hybrid approach eliminates the false positives significantly, while the other two IDS suffer from a higher number of false positives. Similar results are also obtained for regular mesh, star and ring toplogies of IoT networks, each comprising 16 nodes.

Rest of the paper is organized as follows: Section II covers the background on RPL protocol and network architecture and also presents related work on IDS for IoTs. Section III demonstrates an example of a wormhole attack in RPL based network and also gives an overview of the new K-means based IDS approach. Section IV presents the new decision tree based IDS. A two stage hybrid of K-means and decision tree based IDS is presented in Section V. Experimental results are shown in Section VI. Finally, we conclude in Section VII.

II. Background and Related Work

The 6LoWPAN networks in IoT use the RPL routing protocol. This section presents the background on RPL: its architecture, routing policies and related work on IDS.

A. Architecture

RPL creates a Destination-Oriented Directed Acyclic Graph (DODAG) as shown in the Fig. 1(a). This graph starts with a root node, which is connected to the global internet as its parent and some low-power devices as its children. This root is called the 6BR node. Each node of the DODAG contains the following information: node ID (an IPv6 address), one or more parents, a list of neighbors, and a rank that determines the node's position with respect to the root and other nodes.

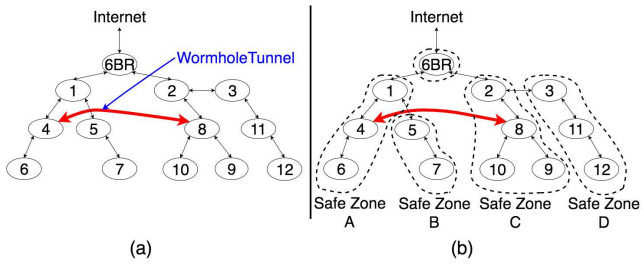

Fig. 1. RPL network: (a) Wormhole, (b) Safe zones.

B. Routing Policies

Two kinds of routing policies are supported: source routing and table-based routing. In source routing, the packet carries the entire path to the destination(s). In table-based routing, all nodes maintain a routing table to know where to send the packets (upwards or downwards).

In this work, we use table-based routing. These tables are continuously updated when new shorter paths are discovered by the nodes: adding new neighbors to reach the destination(s) faster.

C. Related Work

There has not been much research focusing on IDS for IoT. Current approaches are complex solutions, such as using topology reconstruction to detect network graph inconsistencies during sinkhole attacks [3]. This approach deploys IDS at all nodes, which can lead to significant area and energy overhead.

Another work targets rank attack and local repair attack [8]. These attacks are detected using a finite state machine based monitoring network that checks for malicious routing behaviors in each router node. This approach is also distributed and can have significant overhead due to extra monitoring network. In contrast, our machine learning based proposed approaches i.e. KM-IDS, DT-IDS and Hybrid-IDS are centralized approaches. A centralized lightweight attack detection IDS is only deployed at the 6BR.

IDS for wormhole attacks. There is considerable research on detecting wormhole attacks in traditional ad hoc networks. However, there is not much research in designing IDS for wormhole attacks in IoT. In this section, we will present wormhole detection schemes in ad hoc networks followed by a wormhole IDS for IoT, which also forms a baseline for the new work.

Wormhole IDS for ad hoc networks. Raju et al. use an average one hop round trip time (RTT) metric for detection [9]. If the time taken for communication on long paths is more than the round trip time of that path (average one hop RTT * number of hops) then this path is considered to be suspicious. This approach may not be efficient in case of use of high-speed links, where it will not be able to detect attacks. Also the approach will raise more false positives in case of increased congestion. Yifeng Zhou et al. use a distance verification process to detect wormhole attacks in wireless ad hoc networks [10]. Distance verification is performed by comparing the received signal strength (RSS) of received packets against the distance to the sender, computed from the address information in the packets.

Baseline wormhole IDS for IoT. A hybrid centralized-distributed IDS is proposed to detect wormhole attacks in IoT [11]. In this work, the 6BR contains centralized modules that gather requests from other nodes to become neighbors, as well as modules to detect attacks based on the distance between the requesting nodes. The nodes contain distributed modules that send requests to 6BR to become neighbors. The *Attacker detecion* module in 6BR calculates the distance between the requesting nodes and if the distance is greater than a fixed *range*, then an attack is detected otherwise the request is granted. The range used in this work for simulation is 100m and constant.

One of the main limitations of this baseline work is that it uses a fixed range, whose value does not depend on the IoT network properties, such as topology and the distances between nodes. Range is an important parameter and directly determines the attack detection rate. A very small range compared to the average distances between the nodes may lead to a lot of false detections. While, a high range can lead to small detection rate. Therefore, the value of the range parameter must be intelligently selected while considering topology and node distances.

In the new work, we use machine learning approaches like K-means and decision trees to select the optimal range value. Based on this range, we then determine the detection rate. This is the first time machine learning has been used in IDS to detect attacks in IoT networks.

III. Proposed KM-IDS Approach: An Overview

In this section, we consider an example of a wormhole attack in an RPL network. A new K-means clustering algorithm (KM-IDS) is proposed for detection of such attacks.

A. Wormhole Attack Example

Fig. 1(a) shows an RPL network under attack. This network contains the 6BR as the root and low-power devices as other nodes of the DODAG. A high-speed wormhole tunnel has been formed between the nodes 4 and 8. These nodes and the neighbors (1, 6, 2, 9, 10) will start relaying packets through this tunnel. Due to this high-speed traversal, nodes 9 and 6 will appear to be neighbors and will request the 6BR to update the routing tables. However, these nodes are very far apart and should not be neighbors, hence may become victims of this attack.

A simple centralized approach to detect this attack would be if the IDS in 6BR receives requests from two very far apart nodes to become neighbors. An unsupervised machine learning approach is employed to detect this malicious scenario.

B. K-means clustering based IDS (KM-IDS).

K-means is a well-known and highly-efficient clustering algorithm. The basic idea is to divide the nodes into various clusters based on their locations from the 6BR as shown in the Fig. 1(b). These clusters are called safe zones. If the 6BR receives request to update neighbors from two nodes of the same safe zone (e.g. 2 and 9 of safe zone C) then the request is granted. But if a request is received from two nodes belonging to separate safe zones (e.g. nodes 9 and 6) then the IDS of 6BR detects a wormhole attack and denies the request.

Algorithm 1 shows the steps involved in KM-IDS. In Steps 1-2, the network is first initialized and each nodes x,y coordinates are entered into the 6BR. Next step for 6BR is to divide the network in K safe zones. 6BR determines K by first plotting a graph of distortion cost function for various potential K values and then using the Elbow-method to choose the optimal value of K. In Step 3, after optimal K is selected, 6BR randomly initializes K safe zone centroids ($C_1, C_2, , C_k$). Steps 5-7 are safe zone assignment steps where each node is assigned to the safe zone with the closest centroid to the node. In Steps 8-10, 6BR moves the centroids: the coordinates of centroids of each safe zone are reassigned to the mean of the coordinates of the its assigned nodes. 6BR repeats the safe zone assignment and moving the centroid steps 30 times (or, p iterations) to get the optimal centroids for each safe zone. Steps 14-18 show that in KM-IDS, any request from a node to become a neighbor of another node located in a different safe zone is considered as a wormhole attack. In this case, 6BR denies the request and flags an attack.

Algorithm 1 K-means clustering based KM-IDS algorithm

1: Input: test_data of n nodes $Y_1, Y_2, ..., Y_n$
2: Determine optimal K (number of safe zones) by minimizing the distortion function, J
3: Randomly initialize K safe zone centroids: $C_1, C_2, ..., C_k$
4: **repeat**
5: **for** i = 1 to n **do**
6: S(i) = index (from 1 to K) of safe zone centroid closest to Y_i
7: **end for**
8: **for** k = 1 to K **do**
9: C(k) = average of points assigned to safe zone k
10: **end for**
11: **until** p iterations
12: Let N_i = Number of nodes in S(i)
13: $attacksDetected = 0$
14: **for** i = 1 to K **do**
15: **for** j = i+1 to K **do**
16: $attacksDetected+ = N_i.N_j$
17: **end for**
18: **end for**

It is important to divide the network into optimal number (K) of safe zones. Distortion function, J, gives the cost w.r.t nodes (X) as well as cluster centroids (C) [12]. J is the mean of the sum of squared distance between a node and the cluster centroid that is it assigned to. J is mathematically represented as below:

$$J = \int \frac{1}{n} \sum_{i=1}^{n} ||Y_i - C(S(i))||^2$$

It should also be noted that very small number of safe zones will lead to lower detection rate. While large number of safe zones can cause too many false positive detections. We also understand that a very high detection rate will include more false positive attacks because it is unlikely for so many nodes to not be allowed to interact directly in a network.

IV. Proposed DT-IDS Approach: An Overview

In this section, we present the decision tree based IDS (DT-IDS) for detection of wormhole attacks in IoT.

Decision Tree based IDS (DT-IDS). We first use the decision tree algorithm to train the IDS on a given network whose topology is known in advance. The main idea is to

determine a threshold (or a *safe distance*), which is a measure of the permissible distance between any two nodes to be directly connected. We calculate the threshold by taking the mean of the sum of distances between two directly connected nodes. This threshold can now be used to determine wormhole attacks in a new similarly distributed network whose topology is unknown. In the new network, if the 6BR receives request to update neighbors from two nodes that are at a distance $\leq$ threshold apart, request will be granted. But if the distance between two nodes is greater than the threshold, then the IDS of 6BR detects a wormhole attack and denies the request. It is important to note that both the trained and the new networks must be similar in distribution and size to get more accurate detection rate with lesser false positives.

Algorithm 2 shows the steps involved in DT-IDS. In Step 1, 6BR is first trained in DT-IDS where the training network is entered in the 6BR. First, coordinates of each node are entered into the 6BR and an adjacency matrix is initialized with the nodes and the distances between each node, as shown in Steps 2-10. In Steps 12-16, the adjacency matrix is updated with the training network connections. A threshold distance is determined in Step 17 by taking the mean of all the connections in the training network. In Steps 19-20, a similarly distributed network is entered into 6BR and the coordinates of each node are stored in an adjacency matrix. In Steps 22-27, the new adjacency matrix is scanned and wormhole attack detections are made. In DT-IDS, any request from a node to become a neighbor of another node located at a distance greater than threshold is considered as a wormhole attack. In this case, 6BR denies the request and flags an attack.

Algorithm 2 Decision Tree based DT-IDS algorithm

```
    Input: input_data of n nodes X_1, X_2, ....., X_n and corresponding training_data
 2: Create adjacency matrix, A of size nxn
    for i = 1 to n do
 4:     for j = i to n do
            A[i][j].dist = dist(X_i, X_j)
 6:         if(i=j) A[i][j].connected=true
            else A[i][j].connected=false
 8:         A[j][i] = A[i][j]
        end for
10: end for
    distance = 0
12: for each unique tuple (Xi, Xj, true) in training_data do
        A[i][j]. connected = true
14:     A[j][i]. connected = true
        distance += A[i][j].dist
16: end for
    threshold = mean(distance)
18: Input: test data of n nodes Y1,Y2,.Yn
    Create adjacency matrix, B of size nxn
20: Initialize B following Steps 3 to 10
    attacksDetected = 0
22: for i = 1 to n do
        for j= i+1 to n do
24:         if(B[i][j] greater than threshold)
            attacksDetected++
26:     end for
    end for
```

V. Proposed Hybrid-IDS Approach: An Overview

In this section, we propose a two stage hybrid-IDS that uses both K-means clustering and decision tree approaches to accurately detect wormhole attacks in IoT.

Hybrid-IDS. In this approach, every request for a direct connection made to 6BR is approved or rejected using a two-stage IDS system. We first train the IDS with a known network topology to determine a threshold for direct connections using DT-IDS. In a new network, which is similar to the trained network, the IDS first divides the network into an optimal number of safe zones using KM-IDS as shown in the Fig. 1(b). In the native KM-IDS implementation, if the 6BR receives request to update neighbors from two nodes belonging to separate safe zones (e.g. nodes 9 and 6) then the IDS of 6BR would detect a wormhole attack and deny the request. However, in Hybrid-IDS, if the distance between the nodes 9 and 6 is within the threshold that was determined using DT-IDS, the request will be granted. By using KM-IDS in Step 1 and DT-IDS in the Step 2, we are introducing a filter to eliminate false positives that were generated by KM-IDS.

Algorithm 3 shows the steps involved in Hybrid-IDS. In Steps 1-2, 6BR is first trained with a training network to determine a threshold using DT-IDS. In Step 3, a similarly distributed network is initialized and each nodes coordinates are entered into the 6BR. In Step 4, 6BR divides the new network into K safe zones using KM-IDS. Steps 6-15 is the 2-step process to determine wormhole attacks in Hybrid IDS. In the inner for loops in Steps 8-13, a request from a node to become a neighbor of another node located in a different cluster is considered as a candidate for wormhole attack. Final decision is taken in the outer for loops (Steps 6, 7, 14, 15) in which if the distance between the two nodes is greater than the threshold, 6BR denies the request and flags an attack.

Algorithm 3 Hybrid-IDS algorithm

```
    Input: input_data of n nodes X_1, X_2, ....., X_n and corresponding training_data
    Determine threshold using DT-IDS
 3: Input: test_data of n nodes Y_1, Y_2, ....., Y_n
    Determine K safe zones using KM-IDS
    attacksDetected = 0
 6: for i = 1 to K do
        for t in S(i) do
            for j = i+1 to K do
 9:             for u in S(j) do
                    if dist(t,u) greater than threshold
                    attacksDetected++
12:             end for
            end for
        end for
15: end for
```

VI. Experimental Results

We show the effectiveness of the proposed techniques by measuring the attack detection rates for varying network sizes and topologies. Two types of IoT networks are considered in our experiments: 1) Random networks where nodes are located at random coordinates. Four sizes are considered for these random networks: 10, 50, 100, 200. 2) topologies where nodes are located in regular mesh, ring and star topologies, which are commonly used for IoT networks. Each of these topologies consist of 16 nodes. All IDS approaches were implemented in C++.

A. Random Networks

In this experiment, five different schemes are compared with two baseline and three proposed approaches. The baseline schemes are taken from [11], which uses a fixed range to determine attack detection. In baseline, if the distance between the requesting nodes is greater than the fixed range then attack is detected, otherwise the request is granted. The goal is to show that the new schemes lead to a more accurate detection

than the baseline that uses a fixed ad hoc range. The safe-zones and the safe distances proposed in the new schemes are determined based on learning, considering the various factors of topology and distances between nodes.

The five schemes are discussed below:

1) **Base-high**: This is the baseline wormhole IDS, which uses a fixed range to determine attack detection [11]. In this scheme, a high value is used for the ad hoc range, compared to the average distance between the nodes.
2) **Base-low**: This scheme is also based on the baseline wormhole IDS [11]. The range in this scheme is selected to be a low value compared to the average distance between the nodes.
3) **KM-IDS**: This scheme uses the unsupervised K-means clustering based IDS.
4) **DT-IDS**: This scheme uses the supervised decision tree based IDS.
5) **Hybrid-IDS**: Hybrid is 2-stage IDS that uses K-means clustering in the first step and decision trees in the second step.

Fig. 2 shows the detection rates for four network sizes: 10, 50, 100 and 200 nodes, respectively. Base-high scheme achieves the lowest detection rate due to a very high value of the selected range. Due to this range, the IDS grants most of the requests by the nodes to become neighbors as their distances are smaller than the range. On the other hand, the Base-low scheme achieves very high detection rates because of a very small value of selected range. In this case, most of the requests will be rejected as the distance between the requesting nodes is greater than the small range, leading to very high false positive detections.

Results of these two baseline schemes motivate for a more intelligent analysis where the range is picked based on a learning model. This model must consider the average distances between nodes and their locations in the network topology. All three new IDS (KM-IDS, DT-IDS, hybrid-IDS) use learning based approaches to create safe-zones and determine attacks if the requesting nodes lie in different zones.

The optimal number of safe zones needed to achieve these detection rates in KM-IDS are determined by plotting the J-K graphs for each of the four networks. By using the *Elbow method* and considering the detection rates with lesser false positives, we were able to decide an optimal value for K. The J-K plots for each network size is shown in the Fig. 3, 4, 5, and 6, respectively. The optimal K for each network is determined to be: 3, 8, 11, and 14, respectively. These values increase with the size of the network.

The results show the significant amount of wormhole attack detection (70-93%). DT-IDS achieves a detection rate of 71-80% by smartly selecting the threshold range using a decision tree based learning model. Finally, the Hybrid-IDS that uses K-means in the first step, followed by decision trees, achieves a detection rate of 71-75%. All three approaches are also shown to be highly scalable.

We believe that high rate of attack detection in KM-IDS includes false positives which are significantly eliminated by introducing the two-stage Hybrid-IDS. In Hybrid-IDS, the false positive elimination is marked by a reduction in attack detection rate (2-24%) over KM-IDS. Hybrid-IDS detection rate is 0.5-13% lower than that of DT-IDS. This result may be the result of false positive elimination from the attacks detected by DT-IDS. In DT-IDS, there may be some nodes that are close enough to each other but outside the threshold, due to which DT-IDS detects it as an attack.

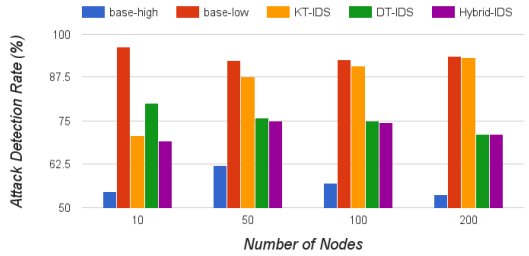

Fig. 2. Detection rates: comparing KM-IDS, DT-IDS and Hybrid-IDS.

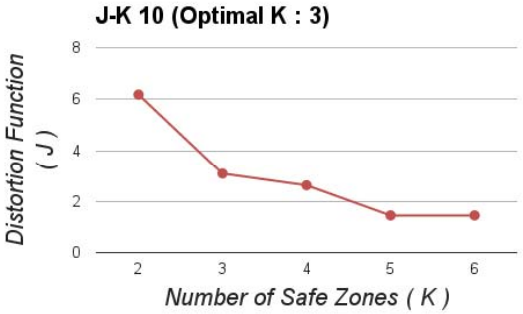

Fig. 3. J-K plot for IoT network with 10 nodes.

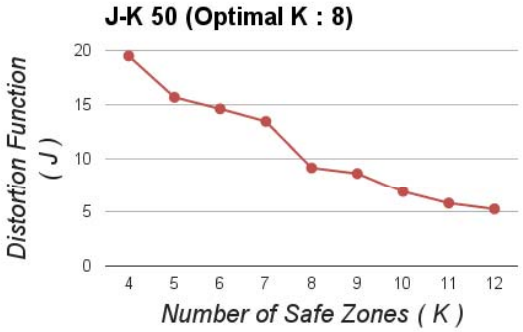

Fig. 4. J-K plot for IoT network with 50 nodes.

Hybrid-IDS is seen to eliminate false positives from both DT-IDS as well as KM-IDS. The networks that we used for experiments were randomly-generated. The advantages of using Hybrid-IDS and KM-IDS can be easily seen in our experiments. We believe that given a real network, the advantages of using Hybrid-IDS over DT-IDS will be more visible since the determined threshold will be more accurate.

B. Topologies

Three regular topologies are considered: mesh, ring and star, which are commonly used for IoT networks. These topologies are shown in Fig. 7, 8, and 9. Each of these topologies consist of 16 nodes.

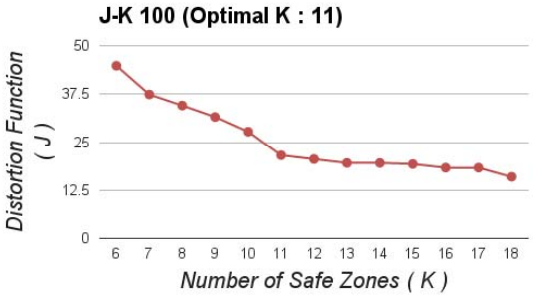

Fig. 5. J-K plot for IoT network with 100 nodes.

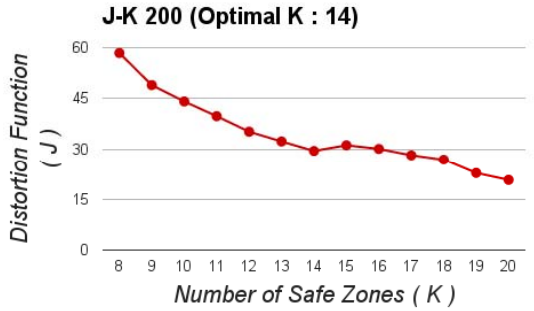

Fig. 6. J-K plot for IoT network with 200 nodes.

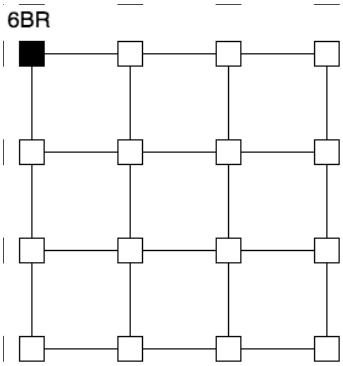

Fig. 7. Mesh topology.

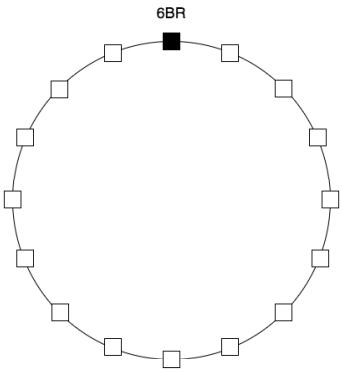

Fig. 8. Ring topology.

Fig. 10 shows the detection rates for mesh topology of KM-IDS, DT-IDS and Hybrid-IDS for varying number of safe zones. There are two important observations:

- Detection rates of KM-IDS and Hybrid-IDS increase as the number of safe zones increase. This result is because as the number of safe zones increase, the number of nodes in each zone decreases and there will be more inter-zone communication leading to more detection of wormholes. On the other hand, detection rate of DT-IDS remains constant at 65%.
- We pick 4 safe zones as the optimal number of clusters for KM-IDS and Hybrid-IDS. For this case, Hybrid-IDS achieves 55% detection rate, which is 11.7% lower detection rate than KM-IDS and 10% lower than DT-IDS due to elimination of some false positives.

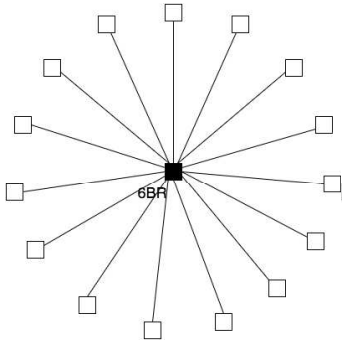

Fig. 9. Star topology.

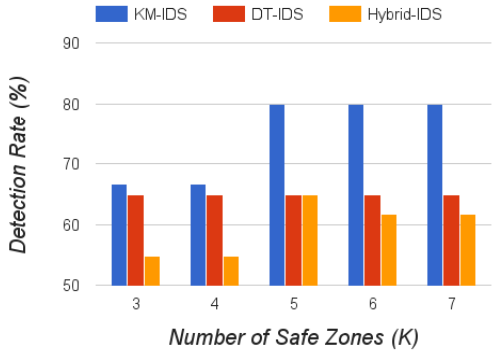

Fig. 10. Detection rates for mesh: comparing KM-IDS, DT-IDS and Hybrid-IDS for varying number of safe zones.

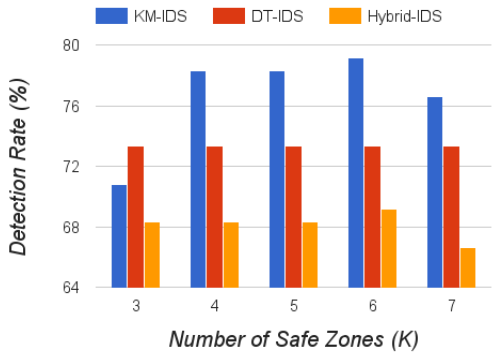

Fig. 11. Detection rates for ring: comparing KM-IDS, DT-IDS and Hybrid-IDS for varying number of safe zones.

Fig. 11 and 12 show the detection rates for ring and star topologies for varying number of safe zones. There are two important observations:

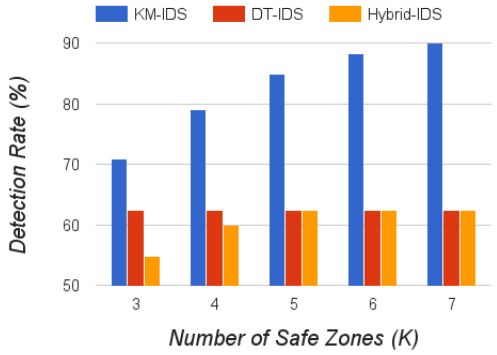

Fig. 12. Detection rates for star: comparing KM-IDS, DT-IDS and Hybrid-IDS for varying number of safe zones.

- Like in mesh, for ring and star also, detection rates of KM-IDS and Hybrid-IDS increase as the number of safe zones increase. On the other hand, the detection rate of DT-IDS is constant: 73.3% in case of ring and 62.5% in case of star.
- Similar to mesh, we pick 4 safe zones as the optimal number of clusters for KM-IDS and Hybrid-IDS, in case of both ring and star topologies. For ring, Hybrid-IDS achieves 68.3% detection rate, which is 10% lower than KM-IDS and 5% lower than DT-IDS. For star, Hybrid-IDS achieves 60% detection rate, which is 19% lower than KM-IDS and 2.5% lower than DT-IDS due to elimination of some false positives.

VII. Conclusion

In this paper, three machine learning based centralized IDS are proposed for RPL networks in IoT: unsupervised K-means based IDS (KM-IDS), supervised decision tree based IDS (DT-IDS) and a two-stage Hybrid-IDS that combines K-means and decision tree approaches. To the best of our knowledge, this is the first time machine learning has been used to develop IDS for IoT. The K-means approach achieves 70-93% detection rate for varying sizes of random IoT networks. Decision tree based IDS achieves 71-80% detection rate and the hybrid approach attains 71-75% detection rate for the same network sizes. Although the hybrid IDS obtains lower detection rate, it is more accurate than the other two approaches. The hybrid approach eliminates the false positives significantly, while the other two IDS suffer from a higher number of false positives. Similar results are also obtained for regular mesh, star and ring topologies of IoT networks, each comprising 16 nodes.

References

[1] N. Cam-Winget, A.-R. Sadeghi, and Y. Jin, "Can iot be secured: Emerging challenges in connecting the unconnected," in *DAC*, 2016, pp. 71.3:1–6.

[2] T. Winter *et al.*, "RPL: IPv6 routing protocol for low-power and lossy networks," *RFC 6550*, 2012.

[3] S. Raza, L. Wallgren, and T. Voigt, "SVELTE: real-time intrusion detection in the internet of things," *Journal of Ad Hoc Networks*, vol. 11, no. 8, pp. 2661–2674, 2013.

[4] G. Hernandez *et al.*, "Smart nest thermostat: A smart spy in your home," *Black Hat USA*, 2014.

[5] Y. Hu, A. Perrig, and D. B. Johnson, "Packet leashes: A defense against wormhole attacks in wireless networks," in *IEEE INFOCOM*, 2003, pp. 1976–1986.

[6] N. Song, L. Qian, and X. Li, "Wormhole attacks detection in wireless ad hoc networks: A statistical analysis approach," in *IPDPS*, 2005.

[7] L. Wallgren, S. Raza, and T. Voigt, "Routing attacks and countermeasures in the rpl-based internet of things," *IJDSN, Article ID: 794326, 11 pages*, 2013.

[8] A. Le, J. Loo, K. K. Chai, and M. Aiash, "A specification-based IDS for detecting attacks on rpl-based network topology," *Information 2016, 7(2), 25*.

[9] R. V, K. V, and K. K, "A simple and efficient mechanism to detect and avoid wormhole attacks in mobile ad hoc networks," in *ICCS*, 2012.

[10] Z. Yifeng, L. L, and L. Li, "Wormhole attack detection based on distance verification and the use of hypothesis testing for wireless ad hoc networks," in *ICPC*, 2009.

[11] P. P and C. G, "Real time intrusion and wormhole attack detection in internet of things," in *IJCA*, 2015.

[12] A. Ng, "Web lecture series on machine learning," 2012.