Technical Disclosure Commons

Defensive Publications Series

February 2020

USING CLUSTERING ALGORITHM AND FOG COMPUTING FOR EN-ROUTE FILTERING IN LOW POWER AND LOSS NETWORKS

Lele Zhang

Akram Sheriff

Chuanwei Li

USING CLUSTERING ALGORITHM AND FOG COMPUTING FOR EN-ROUTE FILTERING IN LOW POWER AND LOSS NETWORKS

AUTHORS:

Lele Zhang
Akram Sheriff
Chuanwei Li

ABSTRACT

Due to trustless link quality in Wireless Mesh Networks (WMNs), Power Outage Notification (PON) and Power Restoration Notification (PRN) messages are often dropped or delayed en-route, which may fail to satisfy customer requirements in practice. Therefore, proposed herein are techniques that use machine learning and Fog computing to efficiently deduce missing PON/PRN messages.

DETAILED DESCRIPTION

Vendors are developing multi-hop wireless mesh networks (WMNs) for smart grid business use and to provide interoperability for with Advantaged Metering Infrastructure (AMI) and Distributed Automation (DA) devices. These WMNs utilize IPv6 Routing Protocol for LLNs (RPL) to establish a tree-based multi-hop topology network based on the RF and PLC medium by using IEEE802.15.4g and P1901.2 protocols. In general, the networks are usually constrained with limited power/energy, bandwidth, and memory resources, are often deployed in hostile environments, and utilize wireless communication.

In WMNs, there are two en-route filtering problems that must be considered. The first problem is that most of the nodes can be easily compromised by an attacker, who can then inject false report and launch path-based DoS (PDoS) attacks. For example, a compromised node will frequently send joining requests to neighbors (e.g., Eapol requests), and then the receivers will forward these requests to the border router (BR) because they could not verify the legitimacy for these requests. This problem is shown below in Figure 1.

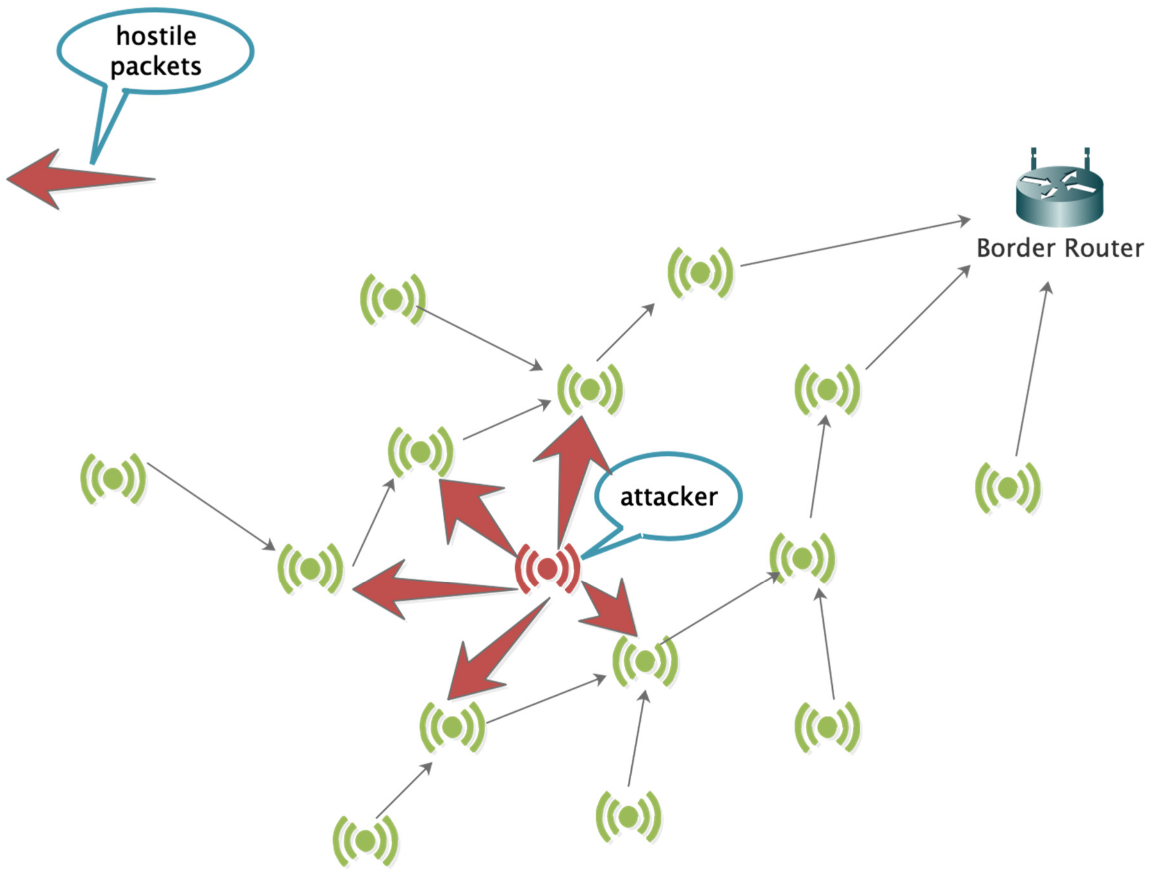

Figure 1

The second problem is that, due to multi-hop topology, a WMN may, in practice, have many hops. As such, the messages from a far node may be invalid when they are transmitted by an intermediate node. For example, an instance of a mesh has up to 24 hops nodes and one node of hop 24 sends a report to the border router and requests a response back within five (5) minutes. Consequently, this report is forwarded hop-by-hop towards the border router. However, due to a poor signal environment, there are many retransmissions on the path. After 5 minutes, the report just achieves the intermediate node of hop 10, so the original node will re-send the same report. This wastes bandwidth resources and the border router will receive multiple duplicated reports and send back multiple responses. This problem is shown below in Figure 2.

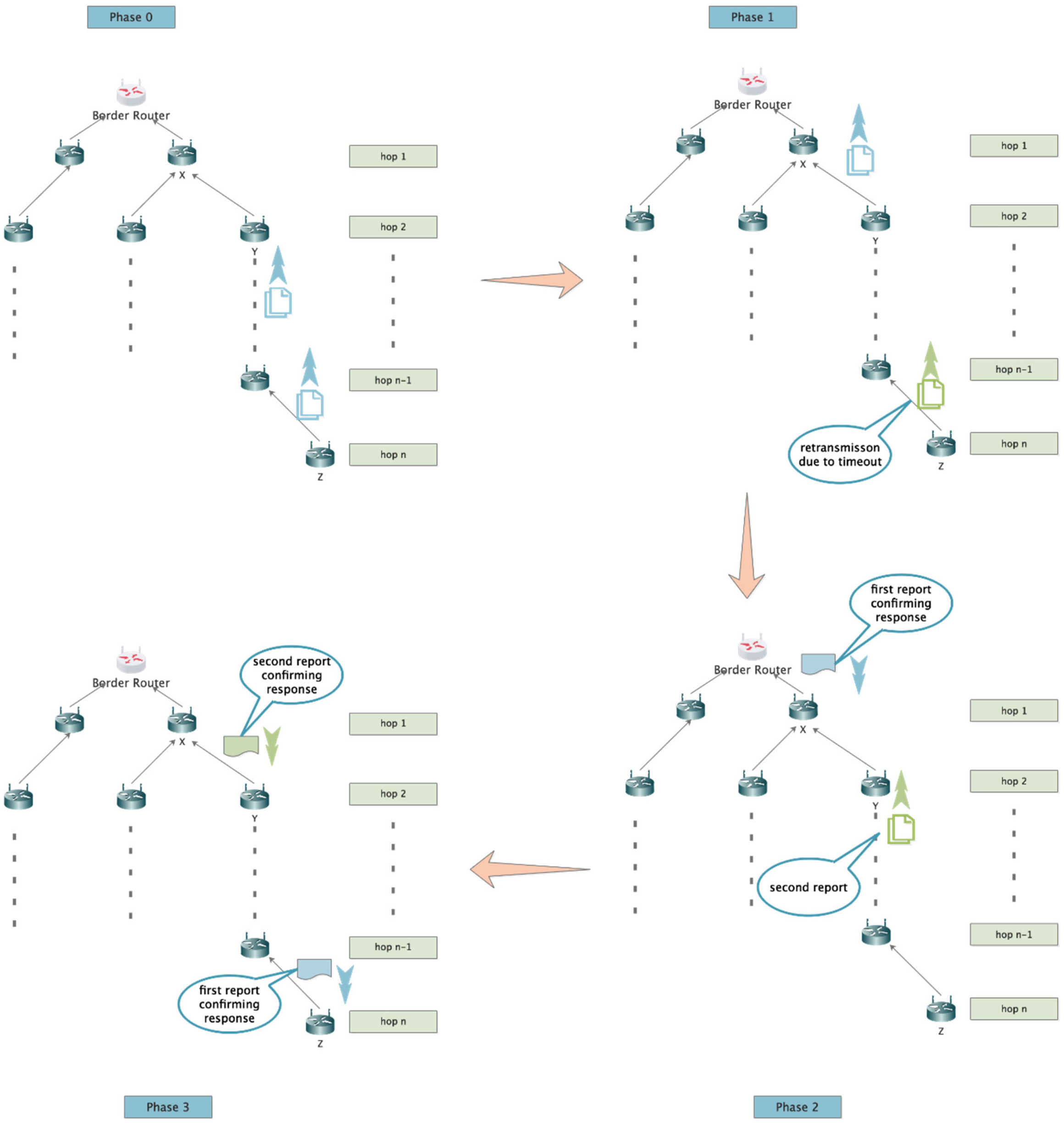

Figure 2

In order to address both of the above problems, the techniques presented herein include a novel of en-route filtering process, based on a clustering algorithm, to drop duplicate packets at the intermediate nodes. It is estimated that all of abnormal messages (e.g., false reports, PDoS messages or timeout messages) could be distinguished from legitimate messages with the following measurements (but not limited to):

- Source address: The abnormal data usually comes from some fixed nodes, such as compromised nodes or timeout nodes, and this can be easily detected by their source addresses.
- RPL RANK: This metric denotes the link distance (i.e., hop depth) between a node and the base station (BS, e.g. border router/root node/sink node/etc.). If a node is too far away from the base station, then its packets will probably be dropped due to timeout.
- Upstream routing path: A compromised node may often forge its source address (i.e., MAC address) to cheat an en-route filtering rule. However, the compromised node cannot change its mediate routing path. For example, an attacker sends multiple false reports to the base station frequently with different source addresses, but its reply node is always node X.
- Packet size: Either false data or timeout packet usually is sent repeatedly, so that the length is same during a period of time.
- The lifetime of the packet: The time interval between the moment that the abnormal packet is generated by the original node and the moment this packet is received by an intermediate node which is responsible for the en-route filtering. For example, as shown in Figure 2, above, node Z will re-send a report if its report could not reach the border router in 5 minutes. With a statistical method, node Y finds that if node Z's packet could not be received in more than 4 minutes, the border router will not receive the packet in time either. Thus, node Y can drop this packet in advance in case of exceeding 4 minutes lifetime.
- etc.

As noted, the techniques presented propose the use of a clustering algorithm (e.g., K-means) to distinguish abnormal data from normal traffic, based on the above metrics. In general, there are two kinds of data in the proposed model, one is normal traffic and the other is abnormal traffic, so that the clustering algorithm is much suitable for this problem. In one example unsupervised learning method, the samples of the clustering algorithm only have feature sets rather than tagged results, such as K-means. The detailed algorithm as described as below in four (4) steps.

Step One

The intermediate node collects all the measurements for every forwarding packet, so each packet could be denoted as one N-dimension vector, N presents the quantity of measurements. At last, all of these N-dimension vectors are placed into a sample set, such as K = {u_1, u_2, u_3, ..., u_k}.

Step Two

Pick up two vectors randomly as cluster centroids, like c_1, c_2 belongs to set K.

Step Three

The other vectors will be divided into two clusters. The vectors closed to c1 will belong to cluster 1, and the other will be members in cluster 2. Then, a new cluster is obtained for a centroid for either cluster, and the clusters are re-divided in accordance with the prior method. This operation will be repeated until both cluster centroids could not move any more, which is illustrated below in Figure 2.

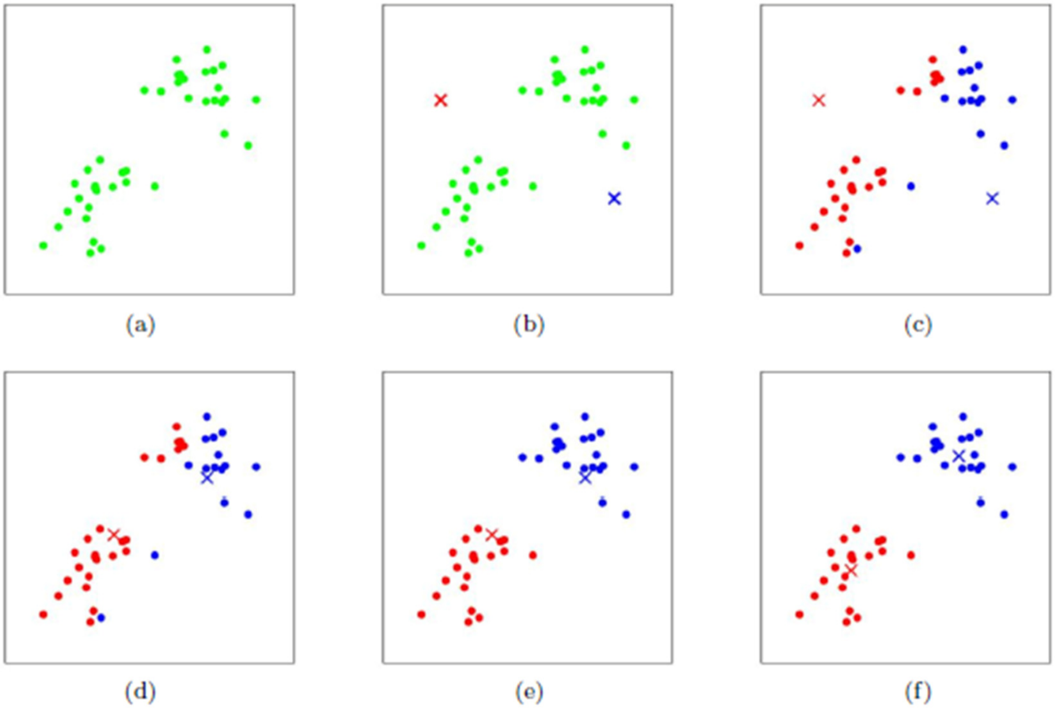

Figure 3

Step Four

In practice, each vector presents a packet, thus one cluster contains illegal packets and the others are normal packets. There are many methods to determine which group is abnormal, such as a smaller cluster (because most of packets are normal in practice), query the border router or cloud with one packet of either cluster, and so on. Once the abnormal cluster is identified, all of the vectors (i.e., illegal packets) will be dropped.

Additionally, in resource-constrained wireless networks, the normal nodes could not execute such a complex machine learning algorithm. Therefore, proposed herein use to distribute some fog/edge computing device among the WMN normal nodes, which control the en-route filtering of illegal packets through use of the clustering algorithm. This is shown below in Figure 4.

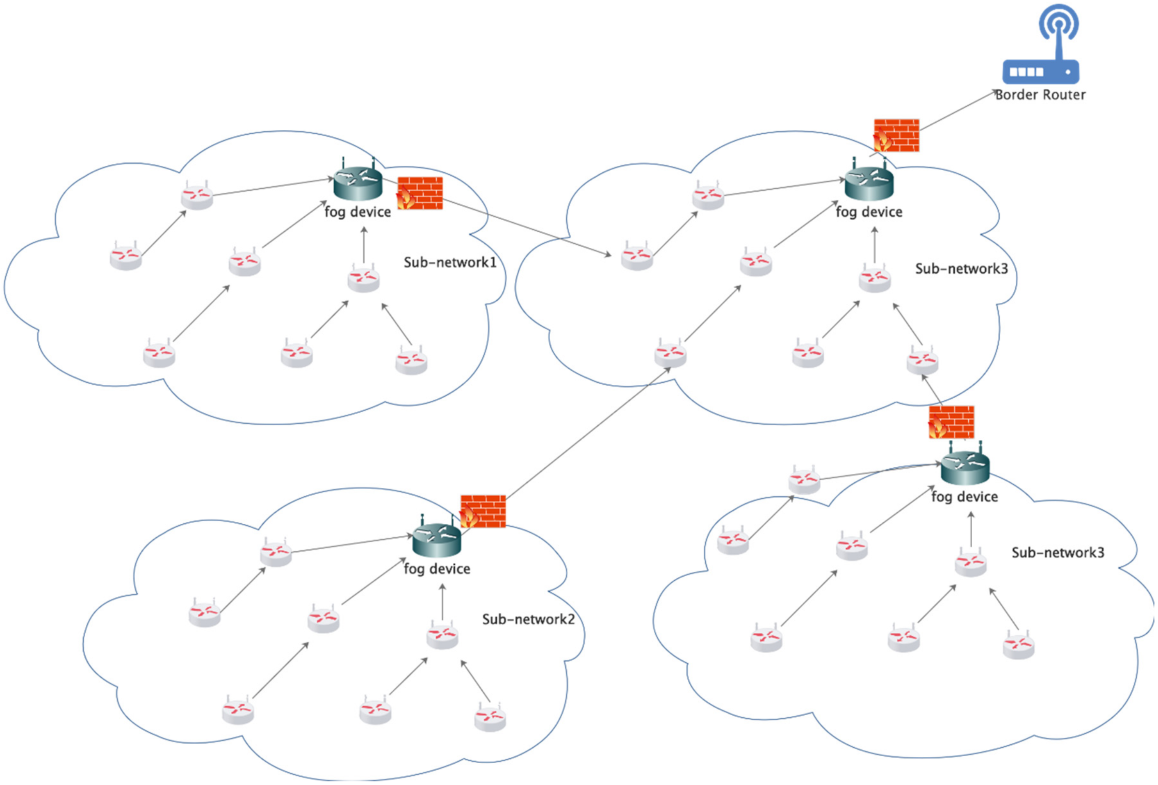

Figure 4

In certain examples, it is proposed that each fog device could cover almost the same quantity of nodes.

In practice, there are many methods could achieve this purpose, such as simulated annealing or Monte Carlo algorithms. However, using clustering algorithm and fog computing technologies, most of the illegal packets will be dropped en-route by fog nodes. With the more data collected by fog nodes, the result will be more accurate.