Detecting Forwarding Misbehavior in Clustered IoT Networks

Nalam Venkata Abhishek
Department of Electrical and Computer Engineering, National University of Singapore
abhishek_nalam@u.nus.edu

Teng Joon Lim
Department of Electrical and Computer Engineering, National University of Singapore
eleltj@nus.edu.sg

Anshoo Tandon
Department of Computer Science, National University of Singapore
dcsansh@nus.edu.sg

Biplab Sikdar
Department of Electrical and Computer Engineering, National University of Singapore
bsikdar@nus.edu.sg

ABSTRACT

Internet of Things (IoT) devices in clustered wireless networks can be compromised by compromising the gateway which they are associated with. In such scenarios, an adversary who has compromised the gateway can affect the network’s performance by deliberately dropping the packets transmitted by the IoT devices. In this way, the adversary can actually mimic a bad radio channel. Hence, the affected IoT device has to retransmit the packet which will drain its battery at a faster rate. To detect such an attack, we propose a centralized detection system in this paper. It uses the uplink packet drop probability of the IoT devices to monitor the behavior of the gateway with which they are associated. The detection rule proposed is given by the generalized likelihood ratio test, where the attack probabilities are estimated using maximum likelihood estimation. Results presented show the effectiveness of the proposed detection mechanism and also demonstrate the impact of the choice of system parameters on the detection algorithm.

CCS CONCEPTS

- Security and privacy → Intrusion detection systems; Mobile and wireless security;

KEYWORDS

Generalized Likelihood Ratio, Maximum Likelihood Estimation, Packet Drop Probability, Malicious Gateway

ACM Reference Format:


1 INTRODUCTION

With the continuing growth of the Internet of Things, the demand for connecting resource-constrained devices to the Internet has been increasing quickly. It is estimated that more than fifty billion devices would be connected to the Internet by 2050 [5]. This has brought in many issues like security, network scalability, etc. [4, 7, 8, 17]. The increasing number of devices in the networks can lead to radio access network congestion. Hence, the need to resolve the network scalability issue becomes apparent. Clustering the devices is one solution to the problem [2]. A set of IoT devices are clustered and are assigned a gateway (or cluster head) which would assist in forwarding the traffic to and from the Base Station (assuming a cellular architecture). The gateways are generally those devices that have superior transmission and processing capabilities. The clustering of the IoT devices can be implemented in many ways, e.g., based on the geographical location, the radio link quality, etc. In terms of security, such a solution creates an opportunity for an adversary to compromise a set of IoT devices merely by compromising their gateway.

Even though message security can be achieved using various approaches [12], the adversary can still compromise the network by misusing the properties of the wireless communication channel. In this paper, the adversary who has compromised the gateway makes the IoT devices retransmit their packets by mimicking a bad radio link between IoT device and gateway. This will affect the network’s performance and at the same time drain the batteries of the IoT devices at a faster pace. Hence, deploying Intrusion Detection systems (IDS) to detect such attacks would be necessary [21]. The major contributions of our paper are listed as follows:

1. The gateways are monitored using a detection algorithm, implemented at the access point, based on the Generalized Likelihood Ratio Test. The statistic required for the same is shared using a side channel from the IoT device to the access point.

2. We present an analytical method to estimate the parameters of the adversary required by the detection system.

3. The impact of the observation interval is demonstrated using extensive simulations and it can be seen that the performance of the IDS improves with the increase in the interval size.

4. The numerical results demonstrate the trade-off between the false alarm and missed detection probabilities as a function of the threshold and that the detection of an adversary becomes easier with increasing attack probability.
1.1 Related Work
Examples of attacks which cryptography cannot defend against are Selective Forwarding, Black Hole and Channel degradation. Researchers in the past have proposed to overcome such attacks using various methods. Machine learning algorithms (such as the ones in [20]), when designed using sufficient and appropriate training data samples, can provide the desired performance. However, in reality, it is problematic to inject packets into the networks to build the training data. Authors in [15] propose a detection technique called SVELTE to detect the presence of a selective forwarding attack. The proposed system detects the adversary when it filters all the packets or sends only the mapping request packets. In [16], the authors have presented an approach based on the channel conditions to detect selective forwarding attacks. A similar approach was proposed in [11] to detect forwarding misbehavior of nodes. However, a sensor monitoring the data packets of the forwarding nodes can be expensive in terms of the energy consumed. Detecting selective forwarding attacks using the traffic eavesdropped by monitor nodes was proposed in [18]. It is however not practical if the system requires a large number of monitor nodes. In [13], the authors have proposed to detect selective forwarding attacks by random selection of a single checkpoint node. To implement it, however, we need to make major changes to the existing protocols. A sequential probability ratio based detection system was presented in [6] for detecting selective forwarding attacks. Their decision is based on the expected transmission count of the nodes. A lightweight heart-beat protocol is proposed in [19]. In this approach, an echo is sent to every node in the network. A selective forwarding attack is detected when there is no reply received from the affected nodes. However, an intelligent attacker might simply refrain from dropping the echo packets and stay undetected.

In our previous work [3], we considered an adversarial gateway which corrupts packets to be forwarded to IoT devices (i.e. the gateway attacks the downlink channel of the IoT devices). The probability of attack was assumed to be constant over all the devices. In contrast to [3], in this paper we consider an adversarial gateway which attacks the uplink channel of the IoT devices. Further, the probability of attack can be different for different devices. In this paper, we also discuss and analyze the impact of the observation interval and attack probabilities on the proposed IDS’s performance.

1.2 Paper Organization
The rest of the paper is organized as follows. In Section 2, the network and adversary models are described. In Section 3, we propose the detection system and design the key parameters of the system. In Section 4, results are presented to show the effectiveness of the system and also to show the impact of the choice of the key parameters on the detection performance. Lastly, in Section 5, we conclude our paper and present a few directions for future research.

2 SYSTEM MODEL
2.1 Network Model
In general, a clustered IoT network will have more than one gateway for assisting the IoT devices. Since the algorithm we present is the same for every gateway, we chose the following simpler but still appropriate network. We consider an IoT network with one access point, AP, one gateway, G, and M IoT devices, $D_j \forall j = \{1, 2, \cdots, M\}$, associated with the gateway. The IoT devices exchange information with the secured access point AP via the gateway. One possible way to implement such a model is to follow the IEEE 802.11ah specification where the gateways in the network operate as “decode and forward” relays [1].

It is assumed that every IoT device has the ability to directly communicate with the access point wirelessly, but at a bit rate lower than it can communicate with its gateway. To extend the coverage of network in order to enable the IoT devices to communicate with the AP and vice-versa, techniques like frequency hopping, power boosting, etc. can be used [14]. Before the IoT device associates itself with the AP, it will disassociate itself from its gateway G. The same side channel will be used by the proposed detection system which will be elaborated upon in Section 3. The network model is illustrated in Fig. 1. The dashed lines representing direct links between IoT devices and the AP are low-rate connections and the solid line represents the link via G. The wireless channel between any two devices in the network is assumed to be memoryless.

For any network in normal operation, there is a non-zero probability of decoding the bits in a packet in error due to various naturally occurring channel and network non-idealities, and/or protocol level behavior. In such a case, the packet will be dropped and a retransmission will be requested by the device. The packet drop probability (PDP) of a packet received by gateway G from a device $D_j$ is assumed to be known and denoted by $a_j$ (i.e. uplink PDP). One of the possible ways to estimate the natural or normal PDP is by measurements when the network is operating normally.

![Network Model Illustration.](image-url)

2.2 Adversary Model
We now describe the strategy employed by an adversary who has compromised the gateway G. The adversary tries to disrupt the communication between the access point and IoT devices connected to G. This is achieved by deliberately requesting the IoT devices to retransmit a successfully received packet. The probability that the gateway requests the device $D_j$ to re-transmit a successfully received packet is $\delta_j$. When a request for re-transmission (sent by the gateway) arrives at the device, the device interprets that the
packet previously sent was received in error by the gateway. Hence, the IoT device re-transmits the packet. By deliberately sending such retransmission requests, the attacker can adversely impact the battery lifetime of the IoT devices and at the same time degrade the performance of the network as measured by other quality of service parameters such as throughput and delay. In the presence of such an attack, a packet can be dropped either due to the network non-idealities or the action of the gateway. Hence, the PDP of a packet received from a device $D_j$ communicating with the AP via malicious gateway $G$ is given by:

$$\beta_j = \delta_j + (1 - \delta_j)\alpha_j$$  \hspace{1cm} (1)

where $\delta_j, j \in \{1, \cdots, M\}$ are unknown random variables. Such attacks are difficult to detect since the attacker is mimicking a bad radio channel.

3 INTRUSION DETECTION SYSTEM

In this section, we present our intrusion detection system (IDS) which is proposed to be implemented at the AP. In the presence of an attack, it is evident from Section 2.2 that PDP of the uplink packets of all IoT devices with $\delta_j > 0$ increases. Hence, we use this parameter to identify whether the gateway is malicious or not. The proposed IDS performs a binary hypothesis test with the following hypotheses:

- Hypothesis $H_1$: Gateway is compromised and is selectively dropping the packets.
- Hypothesis $H_0$: Gateway is not compromised and is in normal operation.

The Intrusion Detection System requires the following abilities to be enabled in the network:

1. All the IoT devices track the total number of packets sent to the gateway, which includes the number of packets retransmitted due to a NACK received from the gateway either explicitly or implicitly.
2. The AP keeps track of the number of uplink packets $(R_j)$ of $D_j$ successfully received from $G$.
3. At regular intervals, say with a time period $T$, each IoT device $D_j$ updates the AP about the number of packets sent $(P_j)$ using the side channel mentioned in Section 2.1. In this paper, for simplicity, we assume $P_j$ varies linearly with $T$, i.e., $P_j = \lambda_jT$. Without loss of generality, we assume the unit of $T$ is ms and the unit of $\lambda_j$ is packets/ms.

3.1 Probability Distributions of the Hypotheses

We denote the number of uplink packets of device $D_j$ dropped by the gateway by $N_j$. Then,

$$N_j = P_j - R_j.$$  \hspace{1cm} (2)

We assume that the packet drops of different devices are independent. When there is no attack, packets are dropped with probability $\alpha_j$, and hence the probability distribution of the variables $N_j, j \in \{1, \cdots, M\}$ under the hypothesis $H_0$ is given as follows:

$$P(N_j = k|H_0) = \frac{(P_j)^k}{k!}(1 - \alpha_j)^{P_j-k}$$  \hspace{1cm} (3)

for $k \in \{1, \cdots, P_j\}$. Similarly, when the gateway is compromised the PDP increases to $\tilde{\beta}_j$, and the probability distribution of the variables $N_j, i \in \{1, \cdots, M\}$ under the hypothesis $H_1$ is given as:

$$P(N_j = k|H_1) = \frac{(P_j)^k}{k!}(\beta_j)^k(1 - \beta_j)^{P_j-k}$$  \hspace{1cm} (4)

for $k \in \{1, \cdots, P_j\}$. We can assume that the wireless channels used by the IoT devices in the network are independent since they will likely be placed more than a few wavelengths apart from each other. Using this assumption, variables $N_j, j \in \{1, \cdots, M\}$ are independent. The joint probability distribution under hypothesis $H_0$ is now defined below, where $N = [N_1, \cdots, N_M]$ and $n = [n_1, \cdots, n_M]$.

$$P(N = n|H_0) = \prod_{j=1}^{M} P(N_j = n_j|H_0).$$  \hspace{1cm} (5)

Similarly, the joint probability distribution under hypothesis $H_1$ is

$$P(N = n|H_1) = \prod_{j=1}^{M} P(N_j = n_j|H_1).$$  \hspace{1cm} (6)

3.2 Detection Algorithm

The likelihood ratio test (LRT) [10], which is known to maximize the probability of detection for any given probability of false alarm, is the optimum detection rule. The LRT decides in favor of $H_1$ if and only if the following holds:

$$\frac{P(N = n|H_1)}{P(N = n|H_0)} > \gamma.$$  \hspace{1cm} (7)

Since the equation in (7) involves parameters $\delta_j, j \in \{1, \cdots, M\}$ which are assumed unknown at the detector, we use the Generalized LRT (GLRT) [10] where the unknown parameters are replaced with their maximum likelihood estimates (MLE) [9]. This will be further elaborated on in Section 3.3. Assuming that $\hat{\delta}_j, j \in \{1, \cdots, M\}$ are the MLEs of $\delta_j, j \in \{1, \cdots, M\}$, we now proceed to derive the detection algorithm as follows where $\hat{\beta}_j$ is replaced by $\hat{\beta}_j = \delta_j + (1 - \delta_j)\alpha_j$. The detection algorithm decides in favor of hypothesis $H_1$ when

$$\sum_{j=1}^{M} \frac{(\hat{\beta}_j)^{n_j}(1 - \hat{\beta}_j)^{P_j-n_j}}{n_j!} > \gamma$$  \hspace{1cm} (8)

$$\Rightarrow \prod_{j=1}^{M} a_j^{n_j}(1 - \hat{\beta}_j)^{P_j-n_j} > \gamma$$  \hspace{1cm} (9)

$$\Rightarrow S = \sum_{j=1}^{M} S_j > \log(\gamma) = \Gamma$$  \hspace{1cm} (10)

where $S_j = n_j \log(a_j) + P_j \log(1 - \hat{\beta}_j)$ and $a_j = \frac{\hat{\beta}_j}{\alpha_j(1 - \hat{\beta}_j)}$.

3.3 Probability Estimation

In this section, we derive the MLEs of the probabilities $\delta_j, j \in \{1, \cdots, M\}$. This is obtained by maximizing (6) over $\delta_j, j = \{1, \cdots, M\}$. It can be observed that the values of the probabilities which maximize (6) are the same values which maximize their individual probability distributions. Hence, the MLE of $\delta_j$ is obtained by setting the derivative of $P(N_j = n_j|H_1)$ with respect to $\delta_j$ to zero,
under the constraint that $\delta_j \geq 0$, i.e.,

$$\hat{\delta}_j = \max \left(0, \frac{n_j}{P_j - \alpha_j} \right).$$

(11)

We now provide an upper bound on the variance of the estimate $\hat{\delta}_j$. Since estimating the mean ($\hat{\mu}_j$) and the variance ($\sigma_j^2$) of the estimate of $\delta_j$ ($\neq 0$) is difficult, we calculate bounds on both the mean and variance. Say, $\hat{\delta}_j = \frac{n_j}{P_j - \alpha_j}$ which implies that $\hat{\delta}_j = \max(0, \hat{\delta}_j)$. It can be seen that $\hat{\delta}_j \leq \delta_j$ which implies that $E[\hat{\delta}_j] \leq E[\delta_j]$. Also, $\delta_j^2 \geq \hat{\delta}_j^2$ which implies that $E[\delta_j^2] \geq E[\hat{\delta}_j^2]$. Using these observations, the following can be inferred:

$${E[\hat{\delta}_j^2] - (E[\hat{\delta}_j])^2} \leq {E[\delta_j^2] - (E[\delta_j])^2}.$$  

(12)

Hence, the bounds on the mean and the variance of $\hat{\delta}_j$ are as follows:

$$\hat{\mu}_j \geq \mu_j \text{ and } \sigma_j^2 \leq \sigma_j^2$$

(13)

where $\mu_j \pm \delta_j$ is the mean of $\hat{\delta}_j$ and $\sigma_j^2 \pm \beta_j(1-\delta_j)$ is variance of $\hat{\delta}_j$. Using (13) and the fact that $P_j = \lambda_j T$, we get

$$\sigma_j^2 \leq \frac{\beta_j(1 - \delta_j)}{\lambda_j T (1 - \alpha_j)}.$$

(14)

It can be seen from (14) that the variance decreases as we increase $T$. Hence, for a higher $T$, a more accurate estimate can be expected with higher probability.

### 3.4 Performance Characteristics

To evaluate the performance of the algorithm in (10), we use the false alarm and missed detection probabilities. The probability that the detection system decides on $H_1$ in the absence of an attack is defined as the false alarm probability ($P_{FA}$). The probability that the detection system decides on $H_0$ in the presence of an attack is defined as the missed detection probability ($P_{MD}$). We have

$$P_{FA} = P(S > \Gamma | H_0)$$

(15)

$$P_{MD} = P(S \leq \Gamma | H_0).$$

(16)

Using the expressions obtained for the estimated attack probabilities in (11), the expression for $S_j$, for $N_j = n_j$, can be written as follows:

$$S_j = \begin{cases} n_j \log \left( \frac{n_j(1-\alpha_j)}{\alpha_j P_j T (1 - \alpha_j)} \right) + P_j \log \left( \frac{\lambda_j T - n_j}{\lambda_j T (1 - \alpha_j)} \right), & \text{if } \frac{n_j}{P_j} > \alpha_j \\ 0, & \text{if } \frac{n_j}{P_j} \leq \alpha_j \end{cases}$$

We know that the probability distribution of $N_j$ is binomial but finding the distribution of $S_j$ is not trivial. Hence, it is difficult to find analytical expressions for $P_{FA}$ and $P_{MD}$. We therefore use numerical techniques to obtain the threshold $\Gamma$ for which the false alarm probability is equal to a desired value $\epsilon$. Note that the attack probabilities are zero under hypothesis $H_0$, and so $P_{FA}$ is independent of the attack probabilities chosen by the adversary. However, for the threshold chosen to meet the desired $P_{FA}$ value, the corresponding $P_{MD}$ is a function of the attack probabilities. The proposed IDS’s performance in terms of $P_{MD}$ will still be acceptable as long as attack probabilities are not too small, which we can assume to be true since otherwise, the attack would not be effective.

### 4 RESULTS

In this section we present simulation results to show the effectiveness of the system and variation of the performance characteristics with $T$. Secondly, we demonstrate the impact of the threshold $T$ on the performance of the IDS. Lastly, the impact of $T$ on the variance of the estimated attack probabilities is demonstrated. For this, we use a network setup with one access point, one gateway and eight IoT devices associated with the gateway.

#### 4.1 Variance of the MLE Estimates

We now demonstrate the impact of $T$ on $\hat{\delta}_j$, the estimates of the attack probabilities. We chose the device $D_j$ whose parameters are available in Table 1 and ran the following steps:

1. For a given value of $T$, we determine the number of packets dropped for the IoT Device $D_j$.
2. We then calculate $\hat{\delta}_j$ using (11).

The sample variance of $\hat{\delta}_j$, for a given $T$, is calculated using the estimates obtained from $10^5$ Monte Carlo simulations. The results obtained are plotted in Fig. 2. It can be seen from the figure that the variance decreases as the value of $T$ increases. By increasing $T$, the number of packets observed will increase. With such an increase, more information is available which implies better accuracy. It can also be seen that the upper bound calculated is very close to the real estimated value.

#### 4.2 Performance Characteristics

The parameters used for demonstrating the performance of the detection algorithm in (10) are tabulated in Table 1. We choose two different time periods ($T$) whose values are 400ms and 800ms. For calculating $P_{FA}$, the following steps were followed for a given $T$:

1. We setup the network using the Hypothesis $H_0$, i.e., all the values of $\delta_j, j \in \{1, \cdots, M\}$ are equal to zero.
<table>
<thead>
<tr>
<th>IoT Device</th>
<th>$\alpha_j$</th>
<th>$\delta_j$ (under $H_1$)</th>
<th>$\lambda_j$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D_1$</td>
<td>0.0789</td>
<td>0.2</td>
<td>0.125</td>
</tr>
<tr>
<td>$D_2$</td>
<td>0.2028</td>
<td>0</td>
<td>0.125</td>
</tr>
<tr>
<td>$D_3$</td>
<td>0.0891</td>
<td>0</td>
<td>0.125</td>
</tr>
<tr>
<td>$D_4$</td>
<td>0.2343</td>
<td>0.2</td>
<td>0.125</td>
</tr>
<tr>
<td>$D_5$</td>
<td>0.2136</td>
<td>0</td>
<td>0.125</td>
</tr>
<tr>
<td>$D_6$</td>
<td>0.0612</td>
<td>0</td>
<td>0.125</td>
</tr>
<tr>
<td>$D_7$</td>
<td>0.0478</td>
<td>0</td>
<td>0.125</td>
</tr>
<tr>
<td>$D_8$</td>
<td>0.0605</td>
<td>0.2</td>
<td>0.125</td>
</tr>
</tbody>
</table>

Table 1: Parameters of the Devices

2. In every iteration, using simulations, we determine the number of packets dropped for every IoT Device.

3. The values of $\delta_j, j \in \{1, \cdots, M\}$ are calculated using (11).

4. We then plug in the values in (10) and compare with a predefined threshold ($\Gamma$) to decide which hypothesis is true.

$P_{FA}$ is obtained by averaging over $10^6$ such Monte Carlo simulations. The same is repeated for calculating $P_{MD}$ with the only difference being that the network is setup using Hypothesis $H_1$. The attack probabilities used in this case are mentioned in Table 1. The variation of $P_{FA}$ and $P_{MD}$ as a function of the normalized threshold ($\Gamma' = \frac{\Gamma}{\lambda_j}$) for the setup is shown in Fig. 3. It can be seen that as $\Gamma'$ increases, the value of $P_{FA}$ decreases and the value of $P_{MD}$ increases. The performance of the detection algorithm with increasing $T$ is shown in Fig. 4, using plots of $P_{FA}$ and $P_{MD}$ against $T$. As can be seen from Fig. 4, both $P_{FA}$ and $P_{MD}$ decrease with increasing $T$ for the same $\Gamma'$. The same can be inferred from Fig. 5 where we are able to achieve a better performance for a higher value of $T$.

4.3 Impact of Attack Probabilities

To demonstrate the effect of the attack probability on $P_{MD}$, we varied $\delta_1$ from 0.1 to 0.5. The results obtained, shown in Fig. 6, depict that $P_{MD}$ decreases with increasing attack probability. To demonstrate the effect of increasing devices under attack, we used the following scenarios:

- Scenario 1: The parameters mentioned in Table 1 are used except for the values of $\delta_4$ and $\delta_8$ which are zero in this case.
- Scenario 2: The parameters mentioned in Table 1 are used except for the value $\delta_4$ which is zero in this case.
- Scenario 3: The parameters mentioned in Table 1 are used.

The results obtained, shown in Fig. 7, depict that as the number of devices effected increase, the value of $P_{MD}$ decreases. Hence, there is a trade-off between the adversary’s choice of attack probabilities and the probability of the attack being discovered.
5 CONCLUSION AND FUTURE WORK

A novel approach for detecting an adversary who is corrupting the communication between an IoT device and the access point by compromising the gateway is presented. The condition for detection is derived using GRLT and is based on the number of packets transmitted by the IoT devices and dropped at the gateway. The estimates for the probabilities $\delta_j, j \in \{1, \ldots, M\}$ are obtained using MLE. Results presented demonstrate the impact of the choice of $T$ on performance characteristics, the effectiveness of the detector with varying $\Gamma$ and the tightness of the upper bound to the real value. The impact of the choice of $\Gamma$ on detector performance was demonstrated using numerical results.

In our future work, we would be working on obtaining analytical results for $P_{FA}$ and $P_{MD}$. Another interesting future work is the design of an optimum attack where the attacker affects the network’s performance while staying undetected.

ACKNOWLEDGMENTS

This research is supported by the National Research Foundation, Prime Minister’s Office, Singapore under its Corporate Laboratory@University Scheme, National University of Singapore, and Singapore Telecommunications Ltd.

REFERENCES