Deep Reinforcement Learning based QoS-aware Secure Routing for SDN-IoT

Aluri Jagan Mohini

Summer term 2020

In IoT (Internet of Things) systems, to handle all the devices efficiently and to overcome the security issues, SDN (Software Defined network) is incorporated into IoT. However, default routing protocols of SDN such as OSPF (Open Shortest Path First) is vulnerable to the flow changes when the network is under attack. To overcome the above mentioned issue, Deep reinforcement learning based QoS-aware Secure routing Protocol (DQSP) is proposed. DQSP guarantees the QoS and is proved to be more efficient than the OSPF.

1 Introduction

IoT is a dominating force in the field of IT for more than a decade due to its versatility in connecting vast number of diverse devices and its usage has been predominant in many sectors like agriculture, marketing, smart services etc.,. However, with increasing heterogeneous devices, the complexity also increases in the system thus affecting the QoS (Quality of service) and leading to security issues. In order to overcome the above issues, SDN has been integrated into IoT which offers a centralised control over these distributed devices [2]. The SDN-IoT architecture consists of sensing layer, data layer and controller layer [1].

Even though SDN-IoT architecture has good programmable and controllable features, it has some disadvantages as well. Due to increased services and the network traffic in SDN-IoT, traditional routing protocols of SDN-IoT namely OSPF (Open Shortest Path First) and RIP (Routing Information protocol) fail to provide solutions to large number of heterogeneous networks according to their demands as they are capable of handling only shortest routes. Moreover these protocols cause packet loss and congestion especially when the rate of the message requests is high or when the network is under attack.

To concentrate on routing optimization, Deep reinforcement learning based QoS aware Secure routing Protocol (DQSP) in SDN-IoT system was proposed [1]. The objective of this proposal is to enhance security and Quality of Service (QoS) in SDN-IoT which traditional protocols lacked. The DQSP is model free and the maximum cumulative rewards obtained in training process are taken as the learning goal achieved. Finally, the experiments were conducted to prove the dominance of DQSP over traditional OSPF.

2 PROBLEM DEFINITION AND ATTACK MODELS

2.1 network Model and Problem Definition

In SDN-IoT, Sensing layer consists of multiple IoT sensor devices deployed in it and these devices are the sources for generated data. The network of the data layer is an un directed graph G(V,E), where vertices are switch nodes and edges are connections between them. Each switch maintains a flow table and the controller gives the flow entry based on requirements of environment [1].

In a traditional routing protocol, once the request message gets delivered from sensing layer, switches in data layer checks whether there is a flow entry for this request or not and if there is a flow entry, then the packet is forwarded else the switch directs the

packet-in-message to the SDN controller. Once it is received by the controller, it decides to send the forwarding policy to Open Flow enabled switch for updating the flow changes. It is noted that if an attack occurs during the forwarding process, the traditional routing protocols are more prone to congestion in the network and that causes delay during data transmission which results in reduced throughput.

On the other hand the proposed DQSP acts as a workaround to above mentioned problem as the additional agent layer in DQSP receives environmental readings from the control layer and it continuously adjusts the network behaviour using cumulative rewards thus improving the overall routing as shown in Figure 1.

2.2 Attack Model

The SDN-IoT system is susceptible to the following attacks [1] :

1) Gray hole attack: Gray hole attack is an attack caused internally by the nodes. It occurs when the legitimate internal nodes act maliciously for a certain duration by dropping packets. The nodes may return to their original states some time later or in the worst case they may even go undetected. In SDN-IoT, the gray hole attack is because of a malicious switch intentionally altering the flow table which results in huge amount of packet loss.

2) Distributed denial of service (DDoS) attack: DDoS is the second type of attack which is caused by an external attacker. Here the switch receives unwanted extra requests by the attacker that are not present in its flow table which may cause flooding of data. It may also result in packet loss and communication blockage due to congestion thus degrading the overall network performance.

3 THE PROPOSED DQSP SCHEME

3.1 Architecture of DQSP

In addition to the above three layers of SDN-IoT, the Deep reinforcement learning based QoS-aware Secure routing Protocol (DQSP) Architecture consists of an additional Agent layer on top of the controller as shown in Figure 1. Once DQSP gets to know its

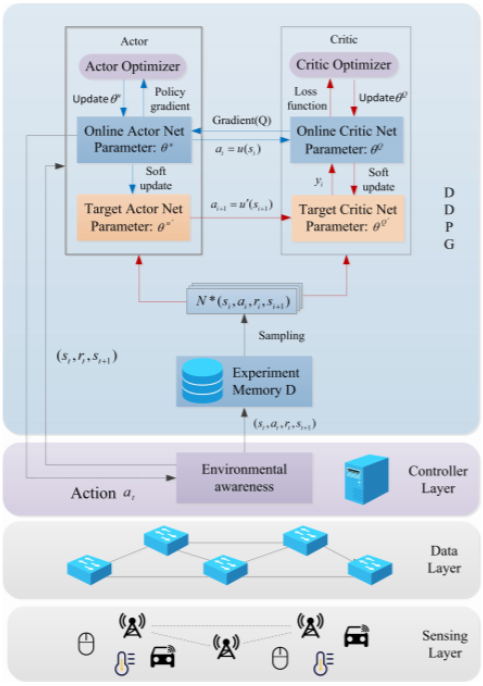

Figure 1: The DQSP architecture [1].

underlying environment, the appropriate routing policy would be generated and Agent layer takes complete responsibility for evaluating the routing policy based on the rewards obtained and it improves the overall network performance by reinforcement learning.

3.2 Related definitions of DQSP

For the experiments carried out, understanding of these basic definitions is necessary.

1) State : A state is a vector [1] which stores the values for frequency of packet-in message, the occupancy rate of the flow table and the channel occupancy rate between the switch node and controller at each time slot t.

2) Action : Action is the decision made by an agent to assign the available switch as a next hop [1].

3) Reward function : Rewards are the performance evaluation metrics considered by an agent to evaluate the efficiency of an action. It can be used for improving the next action. In [1] Reward function

R(t) at time slot t is represented as,

$$R(t) = \frac{1}{|M|} \sum_{i \in M} (\alpha R_{v_i}^{attack}(t) + \beta R_{v_i}{}^{qos}(t)) \quad (1)$$

where $R_{v_i}^{attack}(t)$ is the attack reward and $R_{v_i}^{qos}(t)$ is the QoS reward. |M| is the number of transmissions at time t. α and β are tuning weights and they vary in such a way that their combined sum is always 1.

3.3 The detailed process of DQSP

The DQSP process uses DDPG (Deep Deterministic Policy Gradient) algorithm and the applicability of DQSP to achieve secure routing is discussed below:

1) DDPG : Figure 1 shows the architecture of DDPG in it. DDPG is a deep reinforcement learning method and it follows the actor-critic model from reinforcement learning. The actor in this DDPG learning method contains main actor network denoted by parameter μ and the target actor network denoted by parameter $\mu^{'}$. The critic also contains main critic denoted by parameter Q and target critic denoted by parameter $Q^{'}$. The main actor network updates the policy for every mapping from state to an action and the output a(t) obtained in main actor network is inputted to the main critic network to obtain the value of Q [3].

2) Sampling: Sampling algorithm is the process of generating routing samples. The exploration in an environment by an agent leads to the generation of routing samples. The samples are in the form of (s(t), a(t), r(t), s(t+1)). where s(t) is the initially observed state, a(t) is the action processed in state s(t) which is the output of actor network and r(t), s(t+1) are the reward obtained and new state observed respectively [1].

3) Training: Training is the process where the main network and target network of both critic and actor gets optimized. In this process, the value of Q is obtained when action a(t) is given as an input to main critic network, which is called as the Deep-Q learning approach. The target value of main critic is obtained by adding the current reward r(t) and the value of Q obtain in the state s(t+1). a(t+1) is the action obtained when s(t+1) is given as an input to the target actor network μ', when a(t+1) is given as input to the target critic, the value of Q' is obtained. Likewise, there are different Q values for different actions obtain in same state s(t). Finally, the target network is updated with target actor μ' and target critic Q' [1].

4 RESULTS

4.1 Experiment Setup

DQSP was implemented using Tensor flow at the back end and for the experiment a network consisting of 10 switch nodes with one of them being source and one node as destination was considered. While forwarding more than one node can be selected for next hop. It is assumed that if some particular number of internal switches are attacked then it may reduce network performance. The training of DDPG agent is done under the randomly chosen attacked nodes in a DQSP training network. Simulation is set to 300 episodes. After every 20 episodes, the rewards are calculated to update the network with cumulative rewards. The results says that the DQSP has good convergence as expected by authors. The reward factor can further be improved by varying the learning rate and discount factor. For instance, if we have some discount factor $\gamma = 0.7$ for which it yielded a reward. It was experimentally observed that peak reward can be reached in lesser number of training episodes if the value of γ is increased to 0.9. Similar results were observed even in case of learning rate δ. With increasing value of δ, better results were observed in the subsequent training episodes and peak reward was achieved in lesser number of episodes [1].

4.2 Performance evaluation

The performance of the routing protocols were evaluated and compared by taking the criteria like Packet Delivery Ratio(PDR), End-to-End delay and probability of routing path passing through the attacked nodes into account [1].

As per the results in an attack free network, the PDR value was almost the same in both OSPF and DQSP. But in an attacked model, the PDR value was higher in DQSP compared to OSPF in case of both gray hole and DDOS attacks as shown in Figure 2 and it is mainly because OSPF is primitive and has poor routing policies which resulted in packet loss whereas in DQSP, constant improvement of reward function was made in every episode which

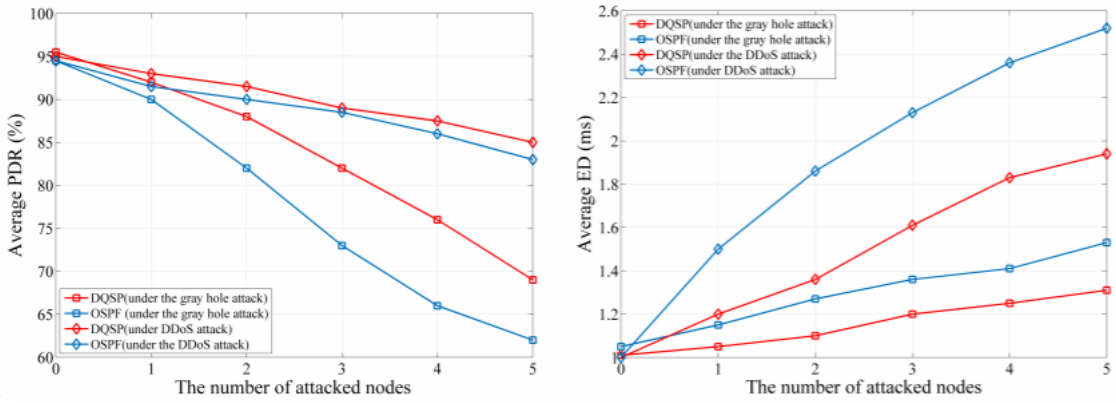

Figure 2: Packet Delivery Ratio and End-to-End delay in an attacked network [1].

resulted in secure and intelligent routing decisions. When End-to-End delay was considered as a performance metric, DQSP still outperformed OSPF in an attacked network and was proved to be more efficient as the delay in it was at least 10 percent less when compared with OSPF. This improvement was a result of intelligent routing decisions which were made by deep reinforcement learning process depending.

Furthermore in case of a gray hole attack, Higher values of PDR and End-to-End delay were observed when the value of reward parameters were set to $\alpha = 0.6$ and $\beta = 0.4$. Where as in case of DDOS attack, the results were quite opposite as the values of both PDR and End-to-End delay were higher when the values were set to $\alpha = 0.4$ and $\beta = 0.6$. Therefore the values of both the reward parameters α and β were carefully adjusted to 0.5 in the experiment for balanced results. In general, DDOS attack had higher values of PDR and End-to-End delay when compared to the gray hole attack as observed in Figure 2. DQSP routing also senses the attacked nodes unlike OSPF and tries to avoid it by choosing an alternative path. This reduces the probability of passing through the attacked nodes thus contributing to the overall QoS of the network.

5 Conclusion

The SDN-IoT networks are prone to gray hole and DDOS attacks which is inevitable. The traditional OSPF is very much prone to QoS issues in an attacked environment. On the other hand, the proposed DQSP proved to be more promising in an attacked environment as it used reward functions of Deep Q learning for evaluating and improving routing strategies. It was experimentally proved that DQSP had atleast 10 percent performance gain when compared with the OSPF.

References

[1] X. Guo, H. Lin, Z. Li, and M. Peng. “Deep Reinforcement Learning based QoS-aware Secure Routing for SDN-IoT”. In: *IEEE Internet of Things Journal* (2019), pp. 1–1.

[2] T. Ninikrishna, S. Sarkar, R. Tengshe, M. K. Jha, L. Sharma, V. K. Daliya, and S. K. Routray. “Software defined IoT: Issues and challenges”. In: *2017 International Conference on Computing Methodologies and Communication (ICCMC)*. 2017, pp. 723–726.

[3] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra. “Continuous control with deep reinforcement learning”. In: 2015. arXiv: 1509 . 02971 [cs.LG].