A Hybrid Reinforcement Learning-Based Trust Model for 5G Networks

Israr Ahmad
Dept. of Computing and Information Systems
Sunway University
Bandar Sunway, Malaysia
israr.a@imail.sunway.edu.my

Kok-Lim Alvin Yau
Dept. of Computing and Information System
Sunway University
Bandar Sunway, Malaysia
koklimy@sunway.edu.my

Sye Loong Keoh
School of Computing Science
University of Glasgow
Glasgow G12 8RZ, U.K
SyeLoong.Keoh@glasgow.ac.uk

***Abstract*—Trust investigation in 5G, which is the next-generation wireless network, is still at its infancy. This research proposes a hybrid trust model for the selection of a legitimate (or trusted) forwarding (relay) entity and to countermeasure intelligent attacks against a route selection scheme. The hybrid trust model has a centralized entity (i.e., a centralized controller C_c) that provides the security level of the operating environment to network entities, and distributed entities that identify malicious or legitimate entities in the network. When the security level of the operating environment is high (low), the distributed entities learn more (lesser) from their respective operating environment. While legitimate entities can use artificial intelligence, such as reinforcement learning (RL), to enhance their trust models, the malicious entities can also use artificial intelligence to increase the detrimental effects of their attacks and minimize their likelihood of being detected. Our proposed trust model is feasible with the introduction of artificial intelligence and the central controller (C_c) in 5G to support the hybrid trust model. We have explained in detail our proposed model that reinforcement learning based hybridization of trust model can enhance the performance of the network interms of learning and tackling intelligent attacks, whereby achieving context awareness and detection of malicious entities.**

***Index Terms*—Trust, 5G, artificial intelligence, reinforcement learning**

I. Introduction

TRUST ensures successful communication among collaborating entities to secure data transmission, contributing to improved network performance (e.g., higher channel utilization and throughput) in traditional networks (e.g., wireless sensor networks and cognitive radio networks) and next-generation networks (e.g., 5G) [1]. Nevertheless, trust provision is still at its infancy in a 5G network, and its intrinsic characteristics, including highly dynamic and heterogeneous, and the close cooperation between centralized and distributed entities (e.g., the between the centralized controller and small cell base stations), have brought new challenges [2, 3].

Centralized trust models require network-wide information that may become stale in a highly dynamic operating environment, which is unsuitable for delay-sensitive (or real-time) schemes. This is because information exchange incurs time and affects the freshness of the data. In contrast, distributed trust models use local information only, rather than network-wide information, that may not be sufficient for making network-wide decisions.

Also in 5G networks, while packets are being forwarded along a route, forwarding (relay) entities may drop packets, contributing to a lower packet delivery rate and can be consider as a security vulnerability. The behavior of forwarding entities, which is dynamic in nature, can change from being normal to malicious and vice-versa, as time goes by. Using RL as a learning agent, an entity can observe and learn an operating environment and decide on the selection of a legitimate (trusted) forwarding entity and detection of a malicious entity [4].

Figure 1:b shows a transmitter i, as an agent, observes state $s^{i}_{n,t}$ that represents the behavior (i.e., trust value estimate) of a neighboring agent that may serve as a forwarding entity. Trust value being as one of the distinguishing factor for legitimate and malicious entities, the transmitter i (i.e., RL agent) takes action ${a_t}^{i}$ to whether select or not, a best forwarder among the neigbours. Subsequently, the transmitter i receives reward r^{i}_{t} that represents performance metrics, such as packet successful transmission rate. While legitimate entities use RL to identify and withdraw malicious entities from collaboration, malicious entities can also use RL to launch attacks against the legitimate entities with out being detected. The malicious entity percieves trust value of a legitimate entity and launches its attack with random probability and intensity. The probability and intensity of an attack represent the frequency and strength, respectively. Malicious entity that is being selected as the forwarding entity drops packets, which is widely known as black hole attacks. Thus, the operating environment or the state can be manipulated and the reward can be affected by malicious entities. Since there is lack of an entity that tell the legitimate entities whether to learn or not, the legitimate entities may continue to learn under manipulated environment. Since both kinds of entities can use RL to countermeasure each other, the network performance become unpredictable. Specifically, it is unknown which kind of entities would outperform the other. Hence, an intelligent hybrid trust model that addresses the security vulnerabilities of both centralized and distributed models is needed to cater for highly dynamic (i.e., changes in the behavior of an entity) and heterogeneous network (i.e., delay-sensitive and delay-tolerant schemes). This paper proposes a reinforcement learning (RL)-based hybrid trust

model to tackle security vulnerabilities at both local and global levels [5].

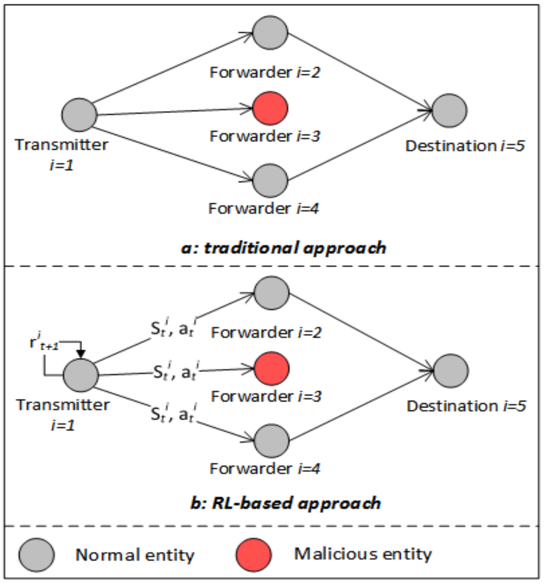

Fig. 1. Selection of forwarding entity. a: traditional-based selection of forwarding entity j in the presence of a malicious entity $i = 3$., b: RL-based selection of forwarding entity j in the presence of a malicious entity $i = 3$. Arrowed line represents a potential data transmission.

The proposed trust model (TM) is feasible with the introduction of the central controller and cloud in 5G to support the hybrid model in the presence of RL-based malicious entities. The underlying forwarding entity selection scheme also capitalizes on the new features of 5G, particularly device-to-device communication and traffic offloading. The proposed trust model also caters for the characteristics of 5G.

A. Our Contribution

This paper presents notion of a hybrid trust model that incorporates RL to ensure local- and global-level trust in 5G networks in the presence of RL-based malicious entities. RL is embedded: a) in the centralized controller (C_c) to gather network-wide information, learn, and make global-level decisions (e.g., learning about the presence of); and b) in distributed entities (i.e., nodes) to observe information from local operating environment, learn, and make local-level decisions (i.e., selecting the best forwarder (relay node) to transmit packets towards a destination. The proposed hybrid trust model is adaptive to network dynamics, whereby malicious entities, whose behavior changes as time goes by, are identified and removed from collaboration in the network.

II. System Model and Trust Model

This section presents the system model and trust model.

A. System Model

Figure 2 shows our system model. The system model consists of two planes; a) control plane that consists of macrocell, cloud, and central controller C_c that manages, controls and coordinates network-wide information to ensure system-level security; and b) data plane that consists of picocells and femtocells where legitimate and malicious distributed entities are found [6]. In Figure 2, transmitter i transfer data packets to forwarding entity j, which serves as the relay to forward packets towards the destination. The hybrid trust model consists of: a) centralized trust model embedded in the control plane to ensure global-level security; and b) distributed trust model embedded in the data plane to select a legitimate forwarder entity to forward a packet towards a destination. While a legitimate entity uses a RL model to select a forwarder entity, a malicious entity also uses a RL model to perceive the trust value assigned by a transmitter and launch successful attack (i.e., drop packets0) with out being detected by varying frequency and strength (see Section V-C). The hybrid model is further discussed in the rest of this section.

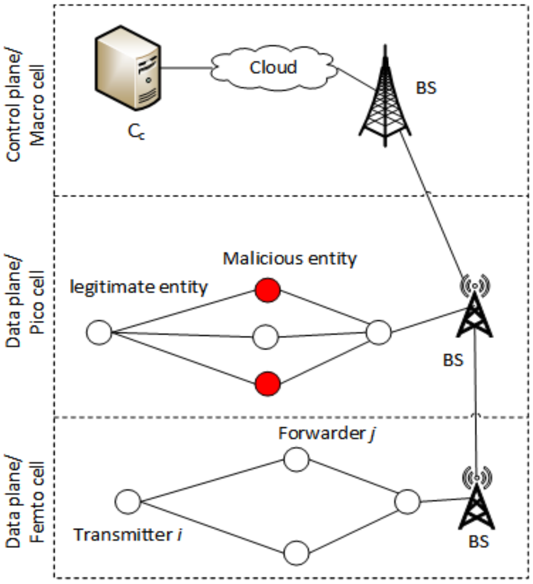

Fig. 2. Sytem model; the network is segregated into three main layers, namely macrocell, picocell, and femtocell layers, respectively. The control plane consists of the macrocell layer and the data plane consists of the picocell and femtocell layers. Transmitters and forwarding entities are located in the femto and pico cells of data plane.

III. Trust Model

Trust Model (i.e., TM) is a framework to detects and removes malicious and misbehaving entities from the cooperation to ensure trust among collaborating entities and minimize detrimental affects of attacks (i.e., packets dropping etc) in the networks [2]. The trust represents the reputation i.e., behavior of an entity over a period of time. High trust value of an entity represensts its legitimacy, while low trust value represents maliciousness. There are two types of approaches to calculate trust values i.e., 1) direct approach, whereby an agent (entity) calculates trust value by direct interaction with another entity; 2) in-direct approach, an entity considers trust information (i.e., trust experience of third entity) from other legitimate entities about a specific entity. Beta distribution, the most popular and simple method is used to calculate trust value for an entity [7, 8]. Beta distribution consists of two parameters i.e., α and β [9]. It is expressed as follows;

$$f(p|\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1} \tag{1}$$

where $0 \leq p \leq 1$ represents the probability of a behaviour of an entity, $\alpha > 0$ represents the trustworithiness of an entity, $\beta > 0$ represents the maliciousness of an entity and $\Gamma(\cdot)$ is a factorial function. Below is the probability of the expected beta distribution,

$$E(p) = \frac{\alpha}{\alpha+\beta} \tag{2}$$

Trust calculation in our paper, refers to the statistical expection of the trust function below;

$$E(T_{ij}^{t}) = \frac{P(s)+1}{P(s)+P(d)+2} \tag{3}$$

where;

- $P(s)$ represents number of packets successfully forwarded by entity (node) j at time t.
- $P(d)$ represents number of packets dropped by entity j at time t.

1) Distributed trust model: The distributed model lies in the data plane, which consists of distributed entities (i.e., nodes and edges) that can communicate and interact with each other directly without going through a centralized entity, with-in the range of communication. A source (transmitter) node i (i.e., a node that sends a packet) selects a forwarder node (i.e., a node that is supposed to receive a packet from source node and transmit it towards destination) with in his transmission range to forward a packet. As the network is highly unpredictable, the entity can behave legitimate (or malicious) at some specific point of time t, leading to various security vulnerabilities such as black hole attack etc. Trust model ensures and establishes trust among communicating entities by detecting and removing malicious entities from the network. The detection and removal of malicious entities can be achieve by observing the behavior of an entity (i.e., packets drop or forward rates) and assigning trust accordingly. For instance, an entity is dropping packet with higher rate is supposed to have low trust value (i.e., packet dropping rate $P(d)$ is inversely proportional to trust value T_N (i.e., $P(d) \propto \frac{1}{T_N}$). Each entity $i\epsilon N$ in the network is embedded with RL model that can observe and learn the highly unpredictable operating environment. *Fistly,* RL model observes the operating environment which is the state (i.e., trust value which is assigned based on their behaviour such as, packets drop or forward towards the destination). *Secondly,* takes an action (i.e., select or discard the forwarder entity for forwarding of packets towards the destination). *Lastly,* gets reward (i.e., successful packets transmission rates).

2) Centralized trust model: The centralized model lies in the control plane which consists of a central entity (or server (cloud)). The centralized entity consists of controllers that collect network-wide data from the distributed entities through edges. It processes the network-wide collected trust data to ensure system level security. Distributed entities keep on observing and learning from their operating environment with out knowing system level security status (maliciousness). The malicious entity also uses RL to launch intelligent attack whereby changing frequency and intensity during attack and avoid being detected. So, lack of any entity that tell the distributed entities to stop or continue under such manipulated environment. Using the network-wide trust information and RL, central entity can decide based on a threshold (i.e., ratio betweeen legitimate and malicious entities in the network) to whether legitimate forwarding entities stop or continue learning their operating environment.

IV. Reinforcement learning Algorithm

Artificial intelligence approaches such as RL can be incorporated in an entity of 5G network such as user equipment UE (mobile, laptop etc) or base station BS (controller) to make intelligent decisions [10]. RL enables an agent (or decision maker such as the UE or controller) to observe and learn from the operating environment [11]. In order to use the RL approach, the three main representations of RL, namely *state*, *action*, and *reward*, must be designed. The state represents the decision making factors (e.g., the estimates of trust values) that affect action selection and reward. The action represents a selected action, such as a forwarding entity (or node). The reward represents network performance (e.g., packet delivery rate, malicious node detection rate, and false alarm) achieved by the agent for taking the action under the state, which may either improve or deteriorate the network [12].

1) RL Algorithm: RL algorithm is embedded in each agent i which is shown in algorithm 1. The agent i observes its local environment i.e., the state $s_{n,t}^{i}$ and takes an action a_t^i at epoch of time t. An agent can choose to either exploit or explore the action. Following the action taken, the agent i receives delayed reward $r_{t+1}^{i} = (s_{t+1}^{i}, a_{t+1}^{i})$, whereby it represents the effect (i.e., positive or negative) from its local environment on next epoch of time $t+1$. The exploitation is a greedy approach to select an action with maximum Q-value, which shows being suitable for a state-action pair and is represented as follow;

Algorithm 1 RL Algorithm

```
procedure START PROCEDURE
    Observe current state $s_{n,t}^{i}$
    if exploration then
        select a random action $a_t^i$.
    else
        select an optimal action $a_t^{i,*}$ using Eq. 4
    Receive delayed reward $r_{t+1}^{i}$ $(s_{t+1}^{i}, a_{t+1}^{i})$
    Update Q-value $Q_{t+1}^{i}(s_t^i, a_t^t)$ using Eq. 5.
    end procedure
```

$$a_t^{i,*} = \arg\max_{a \in A_t^i} Q_t^i(S_t^i, a) \tag{4}$$

In contrary, the exploration is a non-greedy and random action selection approach to find a better action in order to update

TABLE I
General Notation

Notation	Description
N	Number of network entities
$E(T^t_{ij})$	Expected trust between *i* and *j*
t	time instance
UE	User equipment
C_c	Centralized controller
P(s)	Packets transmitted successfully by selected forwarder
P(d)	Packets dropped by selected forwarder

Q-value. The agent i updates the Q-values $Q^i_t = (st^i, at^i)$, applying Q-function (i.e., equation 5), while exploring all the state-action pairs (s^i_t, a^i_t) as the epoch time $t = 1, 2, 3...$, goes by.

$$Q^i_{t+1}(s^i_t, a^i_t) \leftarrow (1-\alpha)Q^i_t(s^i_t, a^i_t) + \alpha\left[r^i_{t+1}(s^i_{t+1}) + \gamma \max_{a\in A} Q^i_t(s^i_{t+1}, a)\right] \qquad (5)$$

where $0 < \alpha < 1$ shows the learning rate, and $0 < \gamma < 1$ shows the discount factor [13].

In our work, the distributed entities, including both legitimate and malicious entities, are embedded with RL that observes the operating environment and learns. A legitimate transmitter $i = 1$ observes the state s^i_t (i.e., trust value) of neighboring forwaring entities and selects the best possible forwarder node $a^i_t = i \in \{2, 3, \dots |N|\}$ where $i \neq 1$ and N is a set of the forwarder nodes of transmitter i. Subsequently, it receives a reward r_t (i.e., successful packet transmissions).

V. Proposed Hybrid model

Forwarding entities can drop packets when forwarding packets along a route, resulting in a lower packet delivery rate. The behavior of a forwarding entity is dynamic in nature, and so it can switch from being legitimate to malicious, and vice-versa, as time goes by.

Using RL, legitimate forwarding entities learn about the behavior of other potential forwarding entities. The malicious forwarding entity launches their attack using RL by dropping packets while avoiding being detected, which leads to a manipulated environment learnt by the legitimate entities. However, by coordination among the centralized controller and distributed entities whereby processing both fresh and historical information in the network, it can detect malicious forwarder entity and the centralized controller can disseminate the information on the system level security to continue learning.

As shown in Figure 3, RL is embedded in both centralized and distributed entities. A transmitter, being the distributed agent, observes state s^i_t, takes action a^i_t, and receives reward r^i_t. However, the malicious forwarding entity also uses RL to launch desired attacks and avoid being detected. In such situation the legitimate entities learn the operating environment which is manipulated by such malicious entities. Since there is lack of an entity to tell the legitimate entities whether to learn or not, the legitimate entities may continue to learn under manipulated environment. Under such circumstances, our proposed hybrid model ensures local and global level trust among the distributed entities. The C_c embedded with RL, being an agent, collects the network-wide information (i.e., trust values) from the distributed entities as state (i.e., ratio of legitimate and malicious entities), takes an action whether to stop or continue learning, gets reward (i.e., the rate of successful packets) and updates the knowledge (i.e., the Q-values in the Q-table of the RL). The legitimate forwarding entities identify legitimate next-hop forwarding entities and stop sending packets to malicious forwarding entities.

Meanwhile, using RL, the malicious forwarding entities launch attacks against the legitimate forwarding entities while avoiding being detected, which results in the manipulation and deterioration of the network. Since both forwarding and malicious entities use RL to countermeasure each other, network performance becomes unpredictable. Specifically, it is unknown which kind of entities outperform the other.

In our proposed model, each entity (e.g., UE, node) can transmit data packets to another entity in the vicinity. Each forwarding entity is a transmitter node that forwards packets to a receiver node over a link. The forwarding entities $N = \{1, 2, \dots, n, \dots, |N|\}$ can change their behaviors (i.e., legitimate and malicious) as time goes by under a dynamic operating environment in 5G. Denote a set of potential forwarding entities of node n by $N_n \subset N$. A transmitter $i \in N$ selects a receiver $j \in N_i$ to transmit packets based on the trust value of each potential forwarding entity in N_i. The trust value represents the behavior of a particular forwarding entity, and it is estimated using RL. Hence, the transmitter i selects a receiver j that has the highest trust value out of the set N_i. As this applies to every transmitter along a route, each selected forwarding entities has highest trust value.

A. *RL-based Distributed Trust Model*

As shown in Figure 4, RL model is embedded in network-wide entities i.e., legitimate and malicious entities. The RL model can be represented using *state*, *action*, and *reward* as; The state $s^i_{n,t} = (s^i_{1,t}, s^i_{2,t}, \dots, s^i_{|N_i|,t}) \in S^i_{N_{i,t}}$ where $n = 1, 2, \dots, |N_i| \in N_i$ represents the behavior (i.e., trust value estimate), which is the decision making factor, of a particular forwarding entity n of a transmitter i at time t. The action $a^i_t \in A^i_t = N_i$ represents a forwarding entity. The reward $r^i_t \in R^i_t$ represents performance metrics, such as packet successful transmission rate. The transmitter, as an agent, observes the state $s^i_{n,t}$, takes action $a_t{}^i$, receive reward r^i_t, and learn by updating its knowledge (i.e., Q-values in the Q-table) as shown in Figure 4 and Table II. Therefore, a transmitter learns about the best possible forwarding entity out of its potential set of forwarding entities based on the state as time goes by. This helps transmitters to select forwarding entities with high trust value. As time goes by, the accumulated reward is maximized. Since the same RL model is embedded in each transmitter,

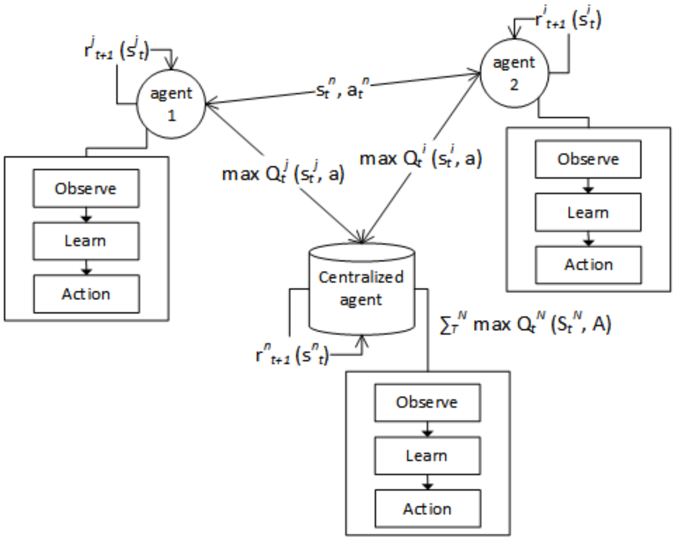

Fig. 3. RL-based hybrid trust model. $maxQ^j_t(s^j_t, a)$ is a specific Q-value in a Q-table at agent 1 which represents a specific policy (action) under a specific state, $r^j_{t+1}(s^j_t)$ represents a received immediate reward at an agent 1 (i.e., rate of transmitted packets or rate of dropped packets), $\sum_T^N maxQ^N_t(S^N_t, A)$ represents received future cumulated reward received from network-wise agents (i.e., global/system level information) learn to take system level action i.e., to continue (stop) learning by network entities at specific time instance t.

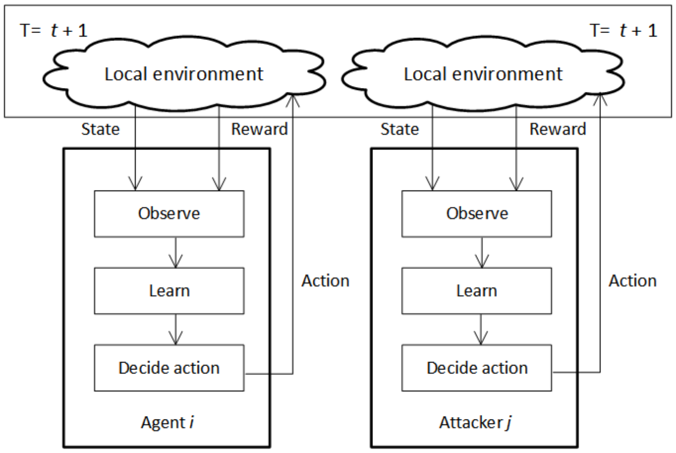

Fig. 4. An abstract view of RL-based legitimate and attacker entities.

packets are forwarded along a route comprised of trasmitters with high trust values.

TABLE II
RL MODEL FOR EACH TRANSMITTER i.

State	$s^i_{n,t} \in S = S^i_{N_t}$ represents the behavior of a particular entity n at time t.
Action	$a^i_t \in A^i_t$ represents the forwarding entity which is selected at time t.
Reward	$r^i_t \in R^i_t$, represents the performance metric, particularly the successful packet transmission rate at time t.

B. RL-based Centralized Trust Model

The RL trust model is embedded in the centralized controller. The controller has network-wide trust information so it can manage and control network activities, such as rewarding legitimate entities and punishing malicious entities (i.e., entity with low trust value). Similar to RL in distributed trust model, RL in centralized controller has *state*, *action*, and *reward* representations. The state s_t represents the ratio of legitimate to malicious entities surrounding forwarding entities in the network. The action a_t is to continue or stop learning. The reward r_t represents network-wide performance metrics, such as packet successful transmission rate. The centralized controller, being an agent, collects network-wide information (i.e., trust values) from distributed entities, which is the state s_t and takes action a_t. Subsequently, the centralized controller receives reward r_t and updates its knowledge (i.e., Q-values stored in Q-table). This enables the centralized controller to learn based on network-wide information in order to select the best possible forwarding entities to the destinations.

TABLE III
RL MODEL FOR CENTRALIZED ENTITY (CONTROLLER).

State	$s_t \in S = S_t$ represents the ratio of legitimate to malicious entities surrounding forwarding entities in the network.
Action	$a_t \in A$ represents the decision to continue or stop learning.
Reward	$r_t \in R_t$ represents the network-wide performance metrics, such packet successful transmission rate.

C. RL-based Attack Model

The malicious entities launch intelligent attack i.e., intend to be selected as forwarder node and drop the packets which is widely known as black hole attack against the legitimate entities. Each malicious entity i varies its *probability (frequency) of attack* $0 \leq P^i_t \leq 1$ and the *intensity (strength) of attack* $0 \leq I^i_t \leq 1$ at time t. The malicious entity uses RL to launch attacks by dropping packets to be forwarded, which reduce the successful transmission rate, by legitimate entities while avoiding detection. The malicious entities vary their probability and intensity of attack based on their respective perceived trust values from the legitimate entities. Table IV summarizes the RL model for the malicious entity. The state s^i_t represents the perceived trust value of the malicious entity i at time t. The action $a^i_t = (P^i_t, I^i_t)$ represents the probability and intensity of attack of the malicious entity. the reward r^i_t represents the packet drop rate and the success of detection avoidance from legitimate entities.

VI. Conclusion and Future Work

This section presents conclusion and future work for our proposed model.

A. Conclusion

5G networks paradigm shift splits traditional hardware communication . The control plane consists of centralized controller that can collect and control network-wide information

TABLE IV
RL MODEL FOR EACH MALICIOUS ENTITY i.

State	$s_t^i \in S = S^i_{N_{i,t}}$ represents its perceived trust value of the at time t. States $s_t^i = 1$ and $s_t^i = 4$ indicate that it has the worst and the best performances, respectively.
Action	$a_t^i \in A = (P_t^i, I_t^i)$ represents the probability and intensity of its attack.
Reward	$r_t^i \in R_t^i$ represents the packet drop rate and successful detection avoidance rate.

i.e., trust values of all the network-wide entities to ensure system level security. The data plane consists of distributed entities that can communicate directly with each other to transmit data packets towards the destination. However, these entities can be legitimate (malicious) as time progresses. The malicious entities can deteriorate the network and manipulate the environment. Thus, to tackle this problem, RL will be embedded in each legitimate entity to detect and remove the neighboring malicious entities and ensure trust among the collaborating entities. However, the malicious entities can also use RL to intelligently launch their attack and avoid being detected. So, there is lack of any entity that tells the legitimate entities to stop or continue learning. Therefore, the intelligent hybrid trust model will ensure the system level security by collecting network-wide information, get aware and tells the distributed legimate about the situation (i.e., to stop or continue) learning.

B. Future work

We shall simulate and investigate our proposed model using python library i.e., Tensorflow and tensorforce to simulate 5G scenario and RL algorithm. We shall evaluate our model by showing malicious entities vs throughput, malicious entities vs accumulated reward, malicious vs legitimate entities ratio and packets transmission rate. Furthermore, we can investigate using deep reinforcement learning in the C_c and distributed entities to make the learning process feasible.

REFERENCES

[1] Shen Su, Zhihong Tian, Siyu Liang, Shuang Li, Shasha Du, and Nadra Guizani. A reputation management scheme for efficient malicious vehicle identification over 5g networks. *IEEE Wireless Communications*, 27(3):46–52, jun 2020.

[2] Han Yu, Zhiqi Shen, Chunyan Miao, Cyril Leung, and Dusit Niyato. A survey of trust and reputation management systems in wireless communications. *Proceedings of the IEEE*, 98(10):1755–1772, oct 2010.

[3] Israr Ahmad, Kok-Lim Alvin Yau, Mee Hong Ling, and Sye Loong Keoh. Trust and reputation management for securing collaboration in 5g access networks: The road ahead. *IEEE Access*, 8:62542–62560, 2020.

[4] Miaojiang Chen, Tian Wang, Kaoru Ota, Mianxiong Dong, Ming Zhao, and Anfeng Liu. Intelligent resource allocation management for vehicles network: An a3c learning approach. *Computer Communications*, 151:485–494, feb 2020.

[5] Sohan Gyawali, Yi Qian, and Rose Qingyang Hu. Machine learning and reputation based misbehavior detection in vehicular communication networks. *IEEE Transactions on Vehicular Technology*, pages 1–1, 2020.

[6] Jordan Lam and Robert Abbas. Machine learning based anomaly detection for 5g networks. *arXiv preprint arXiv:2003.03474*, 2020.

[7] Weidong Fang, Chuanlei Zhang, Zhidong Shi, Qing Zhao, and Lianhai Shan. BTRES: Beta-based trust and reputation evaluation system for wireless sensor networks. *Journal of Network and Computer Applications*, 59:88–94, jan 2016.

[8] Saurabh Ganeriwal, Laura K. Balzano, and Mani B. Srivastava. Reputation-based framework for high integrity sensor networks. *ACM Transactions on Sensor Networks*, 4(3):1–37, may 2008.

[9] Reyhaneh Changiz, Hassan Halabian, F. Richard Yu, Ioannis Lambadaris, Helen Tang, and C. Mason Peter. Trust establishment in cooperative wireless networks. In *2010 - MILCOM 2010 MILITARY COMMUNICATIONS CONFERENCE*. IEEE, oct 2010.

[10] Yanee Naputta and Wipawee Usaha. RL-based routing in biomedical mobile wireless sensor networks using trust and reputation. In *2012 International Symposium on Wireless Communication Systems (ISWCS)*. IEEE, aug 2012.

[11] Mee Hong Ling, Kok-Lim Alvin Yau, Junaid Qadir, and Qiang Ni. A reinforcement learning-based trust model for cluster size adjustment scheme in distributed cognitive radio networks. *IEEE Transactions on Cognitive Communications and Networking*, 5(1):28–43, mar 2019.

[12] K. Maneenil and W. Usaha. Preventing malicious nodes in ad hoc networks using reinforcement learning. In *2005 2nd International Symposium on Wireless Communication Systems*. IEEE.

[13] Yanlong Zhai, Tianhong Bao, Liehuang Zhu, Meng Shen, Xiaojiang Du, and Mohsen Guizani. Toward reinforcement-learning-based service deployment of 5g mobile edge computing with request-aware scheduling. *IEEE Wireless Communications*, 27(1):84–91, feb 2020.