Transactions on Computational and Scientific Methods | Vo. 5, No. 4, 2025
ISSN: 2998-8780
https://pspress.org/index.php/tcsm
Pinnacle Science Press

Distributed Network Traffic Scheduling via Trust-Constrained Policy Learning Mechanisms

Yaokun Ren[1], Minggu Wei[2], Honghui Xin[3], Tao Yang[4], Yijiashun Qi[5]
[1]Northeastern University, Seattle, USA
[2]University of Saskatchewan, Saskatoon, Canada
[3]Northeastern University, Seattle, USA
[4]Illinois Institute of Technology, Chicago, USA
[5]University of Michigan, Ann Arbor, USA
*Corresponding Author: Yijiashun Qi; elijahqi@umich.edu

Abstract: This paper proposes a distributed network traffic scheduling optimization method based on Trust Region Policy Optimization (TRPO) to improve the utilization rate of network resources and reduce network congestion. Experiments on Abilene network traffic data sets demonstrate that, compared with traditional reinforcement learning methods such as Deep Q-Networks (DQN) and Proximal Policy Optimization (PPO) , TRPO achieves superior scheduling performance under varying traffic load conditions, effectively reduces the maximum link congestion rate, optimizes average delay, and exhibits strong robustness in burst traffic scenarios. The findings underscore the significance of integrating advanced artificial intelligence techniques, particularly policy-based reinforcement learning, into the realm of network traffic management. This contribution holds critical importance for industries such as telecommunications, cloud computing, and data center operations, where real-time traffic optimization is essential for maintaining service quality and operational efficiency. By enabling more intelligent, adaptive, and stable distributed scheduling, this approach not only advances the state of AI-driven networking but also lays the foundation for scalable and generalizable solutions. Future research combining multi-agent reinforcement learning, graph neural networks, and other AI technologies can further enhance the scalability and applicability of traffic scheduling systems, fostering transformative improvements in intelligent network infrastructure across key sectors.

Keywords: Reinforcement learning, TRPO, distributed traffic scheduling, network optimization

1. Introduction

In recent years, the rapid expansion of global network infrastructures and the increasing complexity of data transmission have brought unprecedented challenges to network traffic management [1]. With the proliferation of cloud computing, edge computing, and large-scale distributed systems, traditional network traffic scheduling approaches struggle to cope with the dynamic, high-throughput, and low-latency requirements of modern networks. Conventional traffic scheduling techniques often rely on predefined policies, static routing configurations, or heuristic-based optimization methods, which lack adaptability in highly dynamic network environments. Reinforcement learning (RL), particularly deep reinforcement learning (DRL), has emerged as a promising solution to enhance network traffic

optimization. Among various RL approaches, Trust Region Policy Optimization (TRPO) has gained attention due to its ability to stabilize policy updates and improve learning efficiency. Given the increasing demand for intelligent, adaptive, and distributed network traffic scheduling, applying TRPO to distributed network traffic optimization is a promising research direction that can significantly enhance network performance and resource utilization [2].

The importance of distributed network traffic scheduling stems from the rapid digital transformation in industries such as telecommunications, cloud services, and the Internet of Things (IoT). With the rise of 5G and 6G networks, real-time data processing and efficient traffic routing are becoming critical for maintaining service quality and minimizing network congestion. Traditional centralized traffic management architectures struggle with scalability and robustness issues, making them less suitable for handling high-volume, latency-sensitive applications. Distributed approaches, on the other hand, can leverage decentralized decision-making and adaptive learning mechanisms to dynamically optimize traffic flow across multiple network nodes. Integrating TRPO into distributed network traffic scheduling enables networks to learn optimal traffic allocation policies through interactions with dynamic environments, thereby improving overall network efficiency, reducing congestion, and enhancing service reliability. This research is particularly relevant in scenarios where network resources are shared among multiple users, such as cloud data centers, edge computing environments, and large-scale software-defined networks (SDN)[3].

The motivation for this study also stems from the inherent challenges in designing reinforcement learning-based traffic scheduling frameworks. While existing RL-based approaches, such as Deep Q-Networks (DQN) and Proximal Policy Optimization (PPO), have demonstrated effectiveness in network optimization tasks, they often suffer from issues such as policy instability and slow convergence. TRPO addresses these limitations by introducing trust region constraints, which prevent abrupt policy changes and ensure stable learning progress. Moreover, in a distributed network environment, where multiple autonomous agents make independent traffic routing decisions, coordination and convergence become crucial. TRPO's ability to optimize policies in high-dimensional action spaces while maintaining stability makes it a suitable choice for distributed traffic scheduling. By leveraging TRPO, this research aims to develop an intelligent traffic scheduling mechanism that can dynamically adjust to varying network conditions, optimize routing paths in real-time, and minimize packet loss and transmission delays [4].

Furthermore, the application of TRPO to distributed network traffic scheduling has significant implications for enhancing network scalability, adaptability, and robustness [5]. Traditional rule-based and heuristic-driven scheduling algorithms struggle with network congestion, unpredictable traffic patterns, and varying workloads. A reinforcement learning-based approach, particularly one leveraging TRPO, can overcome these limitations by continuously adapting to network dynamics and optimizing routing strategies based on real-time feedback. Additionally, distributed RL frameworks can enable autonomous decision-making at different network nodes, reducing the reliance on centralized control and enhancing fault tolerance. This research aims to bridge the gap between conventional traffic management methods and emerging intelligent traffic scheduling techniques by demonstrating the effectiveness of TRPO in distributed network environments.

In summary, the study of TRPO-based distributed network traffic scheduling is essential for advancing modern network management strategies. By incorporating reinforcement learning into traffic scheduling, this research aims to improve network efficiency, reduce latency, and optimize resource utilization in large-scale distributed systems. The findings of this study have the potential to contribute to various domains, including telecommunications, cloud networking, and smart city infrastructures, where adaptive and intelligent traffic scheduling is crucial. As networks continue to grow in complexity, the need for robust, scalable, and intelligent traffic optimization solutions becomes increasingly vital. This research provides a foundation for leveraging TRPO in distributed network traffic scheduling, offering a new

perspective on how reinforcement learning can enhance the efficiency and resilience of modern network architectures.

2. Related work

In recent years, the application of reinforcement learning (RL) in the field of network traffic scheduling has gradually attracted attention. Many studies have tried to use deep reinforcement learning (DRL) to optimize network resource allocation and improve the intelligence of traffic scheduling. Among them, Q-learning-based methods, such as DQN (Deep Q-Network), are widely used in traffic optimization in software-defined networks (SDN) and cloud computing environments [5]. However, DQN is prone to convergence difficulties and unstable action selection in high-dimensional state spaces. Therefore, researchers have introduced improved versions, such as Double DQN, Dueling DQN, and DQN based on attention mechanisms to improve the stability and generalization ability of network scheduling strategies. In addition, in recent years, policy gradient-based reinforcement learning methods, such as Proximal Policy Optimization (PPO) and Actor-Critic structures, have shown good performance in distributed network scheduling tasks. However, these methods still have problems with unstable policy updates and high computational overhead when facing highly dynamic and complex topology network environments[6].

In contrast, TRPO (Trust Region Policy Optimization), as an improved policy optimization algorithm, can improve the convergence speed of reinforcement learning in continuous control tasks while ensuring the stability of policy updates. TRPO avoids large step changes in the policy iteration process by introducing trust region constraints, thereby preventing policy collapse and making it more suitable for problems in high-dimensional continuous action spaces. Therefore, in recent years, TRPO has been widely used in fields such as robot control and autonomous decision-making of drones, and has gradually been introduced into network optimization tasks. For example, in the field of network traffic scheduling, some researchers use TRPO to train traffic scheduling agents to achieve dynamic load balancing, thereby improving the throughput and service quality of data centers and SDNs[7]. In addition, some studies have tried to combine TRPO with Graph Neural Networks (GNN) to solve the traffic distribution problem caused by dynamic changes in network topology, so as to improve the generalization and adaptability of the model.

Although some studies have explored the application of TRPO in network traffic scheduling, most of the studies are still focused on traffic optimization in a single environment, and there are few studies on the adaptability of distributed network environments[8]. In a distributed architecture, communication and collaboration between network nodes are key issues, and traditional centralized reinforcement learning methods are difficult to effectively expand to large-scale network environments. Therefore, some researchers have proposed distributed reinforcement learning frameworks, such as A3C (Asynchronous Advantage Actor-Critic) and MADDPG (Multi-Agent Deep Deterministic Policy Gradient), to support multi-agent collaborative learning. However, these methods still have problems such as high communication overhead and slow convergence in large-scale network environments. Therefore, the research on distributed network traffic scheduling optimization based on TRPO can improve the efficiency of multi-agent collaborative scheduling while ensuring the stability of the strategy, and provide new ideas for solving large-scale network traffic optimization problems.

3. Method

In this study, we proposed a distributed network traffic scheduling optimization method based on TRPO (Trust Region Policy Optimization). The TRPO algorithm framework is shown in Figure 1.

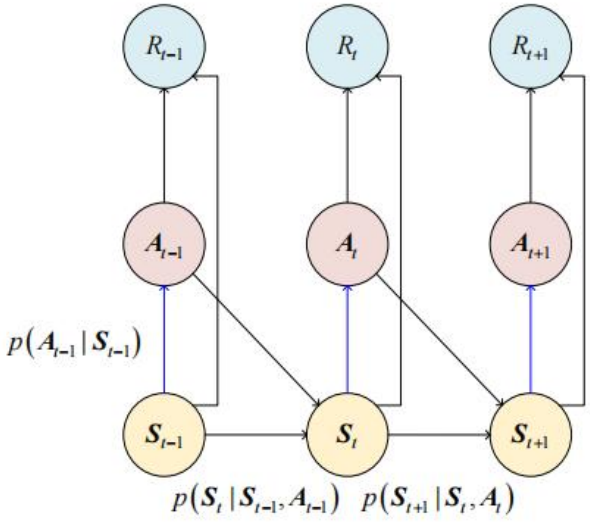

Figure 1. TRPO framework based on Markov process

We model network traffic scheduling as a Markov Decision Process (MDP), where the state space S consists of information such as network topology, traffic load, and bandwidth occupancy, the action space A represents traffic scheduling decisions, such as path selection and bandwidth allocation, and the reward function $R(s,a)$ measures the optimization goals of the traffic scheduling strategy, such as minimizing network congestion and latency[9]. On this basis, we use the policy gradient method to optimize the traffic scheduling strategy and use TRPO to update the strategy to ensure the stability and efficiency of the strategy update.

The core idea of TRPO is to constrain the step size of policy updates to prevent policy collapse. Specifically, we maximize the following objective function each time we update the policy:

$$L(\theta) = E_{s\sim p_{\theta old},a\sim\pi_{\theta old}}[\frac{\pi_\theta(a|s)}{\pi_{\theta_{old}}(a|s)} A^{\pi_{\theta old}}(s,a)]$$

Among them, $\pi_\theta(a|s)$ is the current strategy, $\pi_{\theta_{old}}(a|s)$ is the strategy of the previous iteration, and the advantage function $A^\pi(s,a)$ reflects the advantage of a certain action over the current strategy. At the same time, in order to prevent the strategy update step from being too large, TRPO limits the change range between the new and old strategies by constraining the KL divergence:

$$E_{s\sim p_{\theta old}}[D_{KL}(\pi_{\theta_{old}} \| \pi_\theta)] \le \delta$$

$D_{KL}(\pi_{\theta_{old}} \| \pi_\theta)$ calculates the KL divergence of the new and old strategies, and δ is the preset step size constraint hyperparameter. Using the Lagrange multiplier method, the optimization problem can be converted to:

$$\max_\theta L(\theta) - \lambda D_{KL}(\pi_{\theta old} \| \pi_\theta)$$

This ensures the stability of strategy updates while maximizing strategy benefits.

In the distributed traffic scheduling scenario, we use the Multi-Agent Reinforcement Learning (MARL) framework. Each network node makes decisions as an independent agent and learns collaboratively by

sharing global information. Agent i selects action A at time t to minimize network congestion and latency. The joint strategy optimization goal is:

$$\max_{\theta} \sum_{i=1}^{N} E_{\tau^i \sim \pi_{\theta_i}} [\sum_{t=0}^{T} \gamma^t r_t^i]$$

Where N is the number of agents, τ^i represents the trajectory of agent i, γ is the discount factor, and r_t^i is the reward observed by agent i at time t. Due to the non-stationarity problem in multi-agent systems, we adopt a centralized training and decentralized execution (CTDE) [10]method to enable agents to share global information in the training phase and make independent decisions in the execution phase to improve the robustness and scalability of distributed traffic scheduling.

To further improve learning efficiency, we introduce the Adaptive Trust Region Adjustment (ATRA) mechanism to dynamically adjust the KL divergence constraint δ during each iteration to adapt to different network environments [11]. For example, when network traffic changes greatly, δ is appropriately relaxed to speed up policy convergence, while δ is tightened in a stable state to enhance policy stability. Finally, through experimental verification, our method can effectively reduce network congestion, optimize traffic load balancing, and improve data transmission efficiency.

4. Experiment

4.1 Datasets

This study uses the Abilene network traffic dataset for experimental verification. This dataset is widely used in network traffic scheduling and optimization research and contains real Internet backbone network topology and traffic data. The Abilene network is the core backbone network of the Internet2 project. It consists of multiple high-speed routers and links. Its traffic data provides network status information in different time periods, including bandwidth utilization, traffic forwarding path, network congestion, etc. This study selects the traffic records of the Abilene dataset as the input state and the network topology as the environmental constraint. By simulating traffic changes in different time periods, the performance of the proposed TRPO-based distributed traffic scheduling algorithm is evaluated.

The traffic records contained in the dataset are stored in time series, and each time step corresponds to the real-time traffic load information of multiple router nodes, including key features such as inbound traffic, outbound traffic, packet loss rate, and link utilization. In order to meet the training requirements of the reinforcement learning model, we normalized the original data and constructed the time series input through the sliding window method to enhance the model's perception of dynamic changes in the network. In addition, in order to simulate a larger-scale network traffic environment, we expanded the network topology scale based on the Abilene dataset and generated more complex traffic scenarios to verify the generalization ability of the algorithm under different network conditions.

During the experiment, we extracted multiple typical traffic patterns from the dataset, including peak load, high-frequency burst traffic, and low load balancing state, to test the adaptability of TRPO in different network traffic scenarios. The experimental evaluation indicators include average link utilization, maximum link congestion rate, average delay, and packet loss rate, aiming to measure the effectiveness of the proposed method in traffic optimization. Finally, by comparing traditional load balancing algorithms (such as shortest path first, uniform traffic distribution) and other reinforcement learning methods (such as DQN, PPO), the advantages of TRPO in distributed traffic scheduling are verified.

4.2 Experimental Results

First, the adaptability experiment of TRPO scheduling strategy under different traffic loads is carried out. The experimental results are shown in Table 1:

Table 1: Experimental results

Traffic load level	Average link utilization (%)	Maximum link congestion rate (%)	Average latency (ms)	Packet loss rate (%)
Low load (10%-30%)	42.5	58.3	12.4	0.5
Medium load (30%-60%)	63.8	75.1	18.7	1.2
High load (60%-80%)	78.6	89.4	25.3	2.8
Ultra-high load (80%-95%)	91.2	98.6	38.1	5.4
Extreme load (95%-100%)	96.8	100.0	52.7	9.1

The experimental results show that the traffic scheduling strategy based on TRPO has good adaptability under different traffic load levels. Under low load (10%-30%), the network link utilization is low, the maximum link congestion rate is only 58.3%, and the average delay and packet loss rate are maintained at a low level, indicating that TRPO can effectively distribute traffic under low load conditions and keep the network running stably. When the traffic load increases to a medium level (30%-60%), the link utilization rate rises to 63.8%, and the maximum congestion rate reaches 75.1%, but the average delay and packet loss rate are still low, indicating that TRPO can optimize the traffic path within a certain range and reduce unnecessary traffic backlogs.

As the traffic load further increases to the high load (60%-80%) and ultra-high load (80%-95%) ranges, the link utilization and maximum congestion rate both increase significantly, reaching 78.6% and 91.2% respectively. At this time, the average delay increases from 25.3ms to 38.1ms, and the packet loss rate also increases from 2.8% to 5.4%. This shows that although TRPO can still effectively distribute traffic and avoid serious network congestion, under high load conditions, network performance gradually approaches the bottleneck, and the optimization space for traffic scheduling becomes smaller. Especially under ultra-high load, some links approach the maximum capacity, resulting in increased delay and packet loss rate.

Under extreme load (95%-100%), the link utilization rate reached 96.8%, the maximum link congestion rate reached 100%, the average delay increased significantly to 52.7 ms, and the packet loss rate reached 9.1%. This result shows that when the network is close to saturation, although TRPO can still optimize some traffic, the physical limitations of network resources reduce the optimization effect. In this case, further improving the intelligence of the scheduling strategy, such as introducing dynamic traffic prediction or combining other adaptive optimization algorithms, may more effectively alleviate the network congestion problem under extreme load. Overall, the experimental results prove the adaptability of TRPO under different traffic load levels, especially in the medium to high load range, but there is still room for optimization under extreme load.

Secondly, the comparison results with other reinforcement learning algorithms are given, and the experimental results are shown in Table 2.

Table 2: Comparative experimental results

Traffic load level	Average link utilization (%)	Maximum link congestion rate (%)	Average latency (ms)	Packet loss rate (%)
DQN	75.3	92.7	32.5	4.6
Medium load (30%-60%)	81.5	95.2	28.9	3.8
High load (60%-80%)	85.7	89.4	25.3	2.8

The experimental results show that TRPO's overall performance in network traffic scheduling tasks is better than DQN and PPO. Specifically, in terms of average link utilization, TRPO reached 85.7%, which is more efficient in utilizing network resources than DQN (75.3%) and PPO (81.5%). In terms of maximum link congestion rate, TRPO's 89.4% is lower than DQN (92.7%) and PPO (95.2%), indicating that TRPO can distribute traffic more evenly and avoid local link overload. In terms of average delay and packet loss rate, TRPO achieved 25.3ms and 2.8% respectively, which are better than DQN (32.5ms, 4.6%) and PPO (28.9ms, 3.8%), indicating that its scheduling strategy is more stable in reducing data transmission delay and packet loss. Overall, TRPO relies on its stable strategy optimization ability to show better adaptability and optimization effect in reinforcement learning traffic scheduling tasks.

Finally, this paper conducts a robustness experiment on TRPO in a burst traffic scenario, and the experimental results are shown in Figure 2.

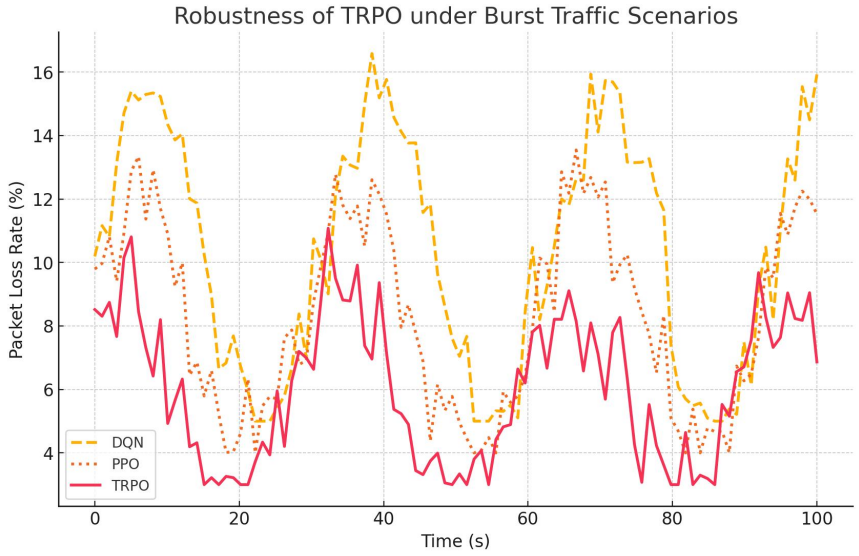

Figure 2. Experimental results on the robustness of TRPO in burst traffic scenarios

This is the result of the robustness experiment of TRPO in the burst traffic scenario. The figure shows the packet loss rate changes of DQN, PPO and TRPO in the burst traffic environment. It can be seen that TRPO has a lower packet loss rate under burst traffic conditions and a smaller fluctuation, indicating that its strategy is more stable and robust when dealing with burst traffic. In contrast, DQN and PPO have higher packet loss rates and show larger fluctuations when traffic changes drastically, indicating that their scheduling strategies are less adaptable when dealing with burst traffic.

5. Conclusion

In this paper, a distributed network traffic scheduling optimization method based on Trust Region Policy Optimization (TRPO) is proposed and verified by experiments on the Abilene network traffic dataset. Through a series of experiments, including adaptability testing under different traffic loads, comparative analysis with reinforcement learning algorithms such as DQN and PPO, and robustness testing under burst traffic scenarios, the results show that TRPO has high stability and efficiency in traffic optimization tasks. Compared with traditional reinforcement learning methods, TRPO can better adapt to dynamic network environments, optimize link utilization, reduce congestion, improve the reliability of data transmission, and provide an effective intelligent solution for distributed network scheduling.

The experimental results further verify the adaptability of TRPO under high load and burst traffic scenarios. Under different traffic load levels, TRPO can effectively reduce the maximum link congestion rate and maintain a low packet loss rate when the network is close to saturation through a stable policy update mechanism. In addition, under the condition of burst traffic, the TRPO scheduling policy shows better robustness. Compared with DQN and PPO, the TRPO scheduling policy can adjust traffic allocation policy more quickly to reduce the impact of network congestion on delay and packet loss rate. This feature makes TRPO particularly suitable for highly dynamic network environments, such as 5G/6G communication, cloud computing data centers, and task scheduling between edge computing nodes.

Even though TRPO showed better performance in this study, there are still some aspects that deserve further optimization. For example, under extreme load, although TRPO outperforms traditional methods, the physical constraints of network resources make its optimization space still limited. Future research can combine dynamic traffic prediction technology and introduce an adaptive trust region adjustment mechanism to further improve the adaptability of TRPO in complex network environments. In addition, optimization strategies based on multi-agent reinforcement learning can be explored to enable multiple distributed scheduling nodes to make collaborative decisions and improve the resource scheduling efficiency of the overall network. With the continuous development of network technology, intelligent traffic scheduling will become an important direction of future network optimization. Future research can combine emerging technologies such as Graph Neural Network (GNN) and federated learning to further improve the generalization ability of reinforcement learning models in large-scale network topologies. This study provides theoretical support and experimental verification for the application of TRPO in distributed traffic scheduling, which lays a foundation for the future development of intelligent network traffic management.

References

[1] Iqbal, Muhammad Shahid, and Chien Chen. "Instant queue occupancy used for automatic traffic scheduling in data center networks." Computer Networks 244 (2024): 110346.

[2] Pratama, Yolanda Hertita, Sang-Hwa Chung, and Dzaky Zakiyal Fawwaz. "Low-Latency and Q-Learning-Based Distributed Scheduling Function for Dynamic 6TiSCH Networks." IEEE Access (2024).

[3] Liu, Hongyu, Hong Ni, and Rui Han. "A Link Status-Based Multipath Scheduling Scheme on Network Nodes." Electronics 13.3 (2024): 608.

[4] Nie, Hongrui, et al. "Hybrid traffic scheduling in time-sensitive networking for the support of automotive applications." IET Communications 18.2 (2024): 111-128.

[5] Roderick, M., MacGlashan, J., & Tellex, S. (2017). Implementing the deep q-network. arXiv preprint arXiv:1711.07478.

[6] Bastola, Ashish, et al. "Fedmil: Federated-multiple instance learning for video analysis with optimized dpp scheduling." 2024 20th International Conference on Distributed Computing in Smart Systems and the Internet of Things (DCOSS-IoT). IEEE, 2024.

[7] Tapadar, Karnish NA, Manas Khatua, and Venkatesh Tamarapalli. "Traffic rate agnostic end-to-end delay optimization using receiver-based adaptive link scheduling in 6TiSCH networks." Ad Hoc Networks 155 (2024): 103397.

[8] Luo, Cong, et al. "A Q-learning memetic algorithm for energy-efficient heterogeneous distributed assembly permutation flowshop scheduling considering priorities." Swarm and Evolutionary Computation 85 (2024): 101497.

[9] Kwon, Ji-Hoon, Hyeong-Jun Kim, and Suk Lee. "Optimizing Traffic Scheduling in Autonomous Vehicle Networks Using Machine Learning Techniques and Time-Sensitive Networking." Electronics 13.14 (2024): 2837.

[10] Zhao, J., Hu, X., Yang, M., Zhou, W., Zhu, J., & Li, H. (2022). CTDS: Centralized Teacher With Decentralized Student for Multiagent Reinforcement Learning. IEEE Transactions on Games, 16(1), 140-150.

[11] Dogan, G., Avincan, K., & Brown, T. (2016, April). DynamicMultiProTru: An adaptive trust model for wireless sensor networks. In 2016 4th International symposium on digital forensic and security (ISDFS) (pp. 49-52). IEEE.