MULTIPATH ROUTING IN 10T USING REINFORCEMENT LEARNING

T.C.Jermin Jeaunita¹, Papareddy Rajeevini Reddy², Sahithi G M³, Subbireddigari Usha Kiran⁴

^{1,2,3,4} R.M.K. Engineering College, Tamil Nadu, India Email: ¹ tcj.cse@rmkec.ac.in, ² papa17225.cs@rmkec.ac.in, ³ sahi17236.cs@rmkec.ac.in, ⁴ subb17225.cs@rmkec.ac.in

Abstract

Wireless Sensor Networks is generally described as a large network of nodes for data monitoring, collection and dissemination. A Large number of routing protocols have been used to achieve actual output and reliability and also involved many numbers of nodes to obtain enormous count of paths which therefore reduces the reliability ofIoT services.Here, a multipath routing protocol is introduced using reinforcement learning.Using this reinforcement method best possible paths from source to destination can be found. The algorithm we use here is Q-learning. Advantage of using this technique is that there is no necessary for this method to depend on environments and they can find the solution for the problem using rewards and transitions. This algorithm tells the representative what action should be taken under what situations. Q-learning algorithm finds the solution to the problem for choosing the efficient path from the source node to the destination node in very less time than the path bridging technique. It is designed to explore possible routes and adapts a network routing based on information gathered and converges towards an efficient solution.

Key words: Reinforcement learning, Multipath routing, Internet of Things

Introduction

The Internet of Things is acquiring everyday things which are enclosed with some electronics, sensors etc., to the network to start collecting some data and also for exchange without any human interference could be a new idea allowing a large number of machines like some sensors to connect to the network. IoT is rapidly raising the range of some areas like agriculture, cities and energy saving. Random Node failures and link losses are some of the major issues for reliable data transmission in Wireless sensor networks. Hence, the routing protocol designing is one of the challenging issues for Wireless Sensor Networks(WSN). Wireless Sensor Network also has some sensor devices for watching and recording some parameters. They also collect and organize the data from the devices. We use WSN for satisfying some requirements like link quality, energy efficiency and also reliability. In the real world, reliability and packet delivery ratio are degraded due to various cases.

For solving the issues based on the reliability there are many possible routing protocols that are already in existence. We perform multipath routing to deliver the packets from source to destination and we use multiple routes for the transmission of the packets in the particular interval regularly to reach the sink node. However, if we include a large network transmitting the packets, it needs more energy and link quality for packet transmission. Hence transmitting the packets becomes difficult and also there will be excessive energy consumption that leads to shortening the lifetime of the network. Therefore, the already available routing protocols are not efficient for some applications of IoT.

We propose an efficient reinforcement learning, Q-learning algorithm, which takes lesser time and energy than the already available routing protocols. The suggested protocol uses the link quality between every two neighbouring nodes and energy of every node to select the best path in the graph. The existing algorithm constructs the main route which is the better path of each node to the destination. The path is chosen as the best path in the network. Secondly, the sub-path is constructed. For construction of the sub-path, we will consider the nodes that are not considered in the main path. Using the main path and sub-path bridge node is found. If a node failure occurs then a new node that is neighbour to both the main path node and also the sub-path node is the bridge node. Now we will find the route from the supplyedge to destination edge in the multipath routing using reinforcement learning. We use Q-learning technique to find the best path in the multipath routing protocol. Through Q-learning we find the best optimal path in the multipath routing protocol.

Related Works

This part of the section explains one of the latest multipath routing protocols developed. Opportunistic Multipath Routing.[1] in long-hop WSN is planned to boost the packet delivery magnitude relation in wireless sensing element networks. The planned technique predicts the specified range of methods, further as bifurcation supported expedient routing in step with the fickleness demand. In this method, an intermediate node is chosen as a unique node for every transmission and conjointly handles path failure. The planned theme achieves a better packet delivery magnitude relation and reduces the energy consumption by a minimum of some thirty third and up to some sixty fifth compared with existing routing protocols, below the condition of associate eightieth link success magnitude relation within the long-hop sensing element network.

A unified single and multipath routing algorithm.[2] for wireless device networks with source location privacy which might handle each event-driven and question driven traffic patterns to deliver packets as shortly as potential. In each single and multipath routing handles traffic at identical times in the same intervals within unified protocol. It uses NS-2 simulations for potency using these simulations which is difficult to search out economical intermediate nodes to deliver packet immediately to destination.

An effective Quality of Service multipath routing protocol[3]In this case, consider the IoT based Wireless Sensor Networks which has shown us an effective QoS-mindful multipath steering convention for IoT dependent on the Wireless Sensor Networks. The proposed protocol estimates the path from supply node to the sink node by estimating the cost of the path, where it considers some parameters that is duration in a node and excess of traffic in a node. This routing protocol has been executed to give the efficient performance than the previous routing algorithm

Energy economical load balancer for ECMP in networks [4] is for congestion management address load levelling, route maintenance problems. In this work, traffic is equally distributed based mostly on packet identifier among multipath routing protocols.it achieves improvement and packet delivery quantitative by NS a pair of 2.35 it improves 11.11% and packet delivery ratio is 43.95%. In case of route failure, it has to notice a third different path which is optimistic, so it takes time and delay to deliver the packet.

MP-TCP-BasedIoT exchanging data Evaluation using Machine Learning [5] establishes machine learning algorithms, and proposes an instinctive learning choice path mechanism supported byMultipath TCP, which might decide some higher-quality methods and transfer some information at constant time. The standard transmission management mechanism supported static mathematical model will now not intersect the completeness and accuracy necessities of the long run Internet of Things. So as to form complete use of the advantages of recent exchanging of Data discourse, researchers have come across some of the deep research on the appliance of machine learning technology. Hence this method makes us understand for checking the transmission algorithms which have a high intellectual learning and have many other useful purposes for the verification of multipath quality and it effectively executes multipath quality. [6]Reliable and Capacity-related Multipath Routing which is planned to decrease the energy capacity requirements of sensor networks, but here it provides suitable knowledge transmission through maintaining a substitute path from each and every source node towards the destination node.

The same to a portion of the generally given multipath conventions, the execution of steering during this convention is moreover begun by the objective hub. While execution of this strategy, at whatever point the objective hub gets a partner message from a source hub and afterward when there is no active course to the source hub, it begins to initialise a functioning pathway disclosure technique and through overpowering a functioning pathway demand message. Subsequent to getting the dynamic pathway demand message at the specific source hub, the recipient hub sends a functioning pathway reservation message towards the sink to substitute the way that is found. At that point as regular the dynamic pathway saved message goes from the source hub towards the objective hub, at whatever point a hub on the converse way gets the message, it restricts

a close by spot to its battery level for commonality transmission this path. The dynamic pathway building technique is finished by getting the dynamic pathway reservation message at the source hub. Despite the fact that this convention gives data about the high energy and adaptable information in the transmission, it additionally experiences the terrible impacts of the decision in multipath directing methodology: the one finish to the next ability is confined to the capacity of one course. Most essentially, multipath convention overlooks the aftereffects of remote and connections unreliability to the ideal energy for information out sending.

The resource limitations of wireless sensing are discussed in [7]. Here, according to the element nodes, very high congestion masses in data rate reliable to network traffic, that extremely has some effect on the graph parameters. A solution to this situation, information scattering pseudocode will make the most of the higher densities of sensing element graphs to extend the graph capability by using a lot of informational techniques. Multipath routing methods will give the simplest answer to take the information measure needed for various applications and cut back the chance of network traffic through cacophonic network congestion ahead of many methods. Also, an exceedingly mono wireless sensor network, sensing element nodes uses a shareable connection to speak with one another co-occurring utilization of adjacent methods leads to inter-path communication, that will raise the opportunity of the data packet colliding at the edges on the live method. This downside is named the edge grouping result and importantly it has also limited the productivity of MP routing protocols. The problem that also has an enormous difficulty in planning economical MP routing algorithms. So the given method needs the property in the network to pick out the small busy ways, victimization the method in relatively high density wireless detector networks produces high procedure.

[8]QoS Improvement: support as far as organization yield, start to finish idleness and information conveyance extent connection might be a fundamental goal in emerging with multipath directing conventions for different sorts of organizations. Found ways with various attributes are regularly needed to disperse network traffic that upholds QoS requests of the machine that the multipath steering convention has been planned. For instance, important data parcels are frequently sent through higher capacity ways with least postponement though defer heartless non-basic data bundles are regularly sent through non-ideal ways with better quality to-end delay. Additionally, in differentiation with the single-way directing methods, multipath steering approaches will safeguard QoS requests of the alleged application inside the instance of way disappointments through controlling organization traffic to another dynamic way. All things considered, because of the connection layer issues in mono remote organizations, rising organization yield and information conveyance greatness connection through synchronic multipath steering in identifier networks will not be just about as clear as wired organizations.

Methodology

In this section, we describe the path bridging method to satisfy the wants like reliability. we'll discuss the general function of the proposed system. consistent with the IoT the appliance may sometimes require just one requirement, but it's necessary to satisfy multiple requirements for best performance.

Path bridging:

Within the path bridging process, we consider the nodes within the network with the link quality between every two neighbouring nodes and energy of each node. As in fig.1, we start from the source node and find the simplest path that has the largest energy within the node and also the nodes with the higher link quality. once we traverse whenever the link quality between the nodes decreases simultaneously. If the link quality fails, we elect the subsequent best node for transmission. If there are quite two nodes with the same energy, select the participating node with better link quality. Thus, a far better path is chosen within the network. This path from source to destination is taken into account because the main path. After the development of the best path, we'll build the sub-path. For sub-path construction from the source to destination, the nodes that were involved within the main path construction aren't considered within the sub-path construction, that's no participation of the most path nodes. that's calculated by subtracting the top time and begin time and therefore the values are noted. Thus, the sub-path construction is completed.



Fig. 1 Finding path through bridging.

After the completion of the main path and sub-path construction, we'll build the bridging path to get reliability. Within the path construction processes the bridge nodes are needed for the inter path communication between the sub-path and therefore the main path. The bridge path acts because the path satisfies the reliability and other real-time requirements. within the multipath construction process if there's any node failure or whether it's poor link quality, we cannot choose that path. So we'll find the node that's the neighbour to both the most path nodes and also the sub-path nodes, this node acts because the bridge node. Hence we'll choose this new path consisting of the bridge node because this path has more reliability and it's the simplest path than the opposite proposed routing protocols.

Reinforcement Learning

A major scope of learning issues is regularly projected into support learning outline work.Broadly expressed, support learning is the trouble of figuring out how to understand an objective through communication during a restricted capacity of climate. The preparation element which is obligated for making moves is named by someone else. The specialist ceaselessly associates with the climate by making moves, and getting prizes and state data. The objective of the specialist is to explore different avenues regarding diverse activity groupings to augment the prize after some time.

A significant part of support learning calculations is that they're prepared to gain from postponed rewards. In certain issues, a specialist must execute a chosen grouping of activities before it gets a blessing, to discover such an arrangement, a specialist must defeat the matter of a fleeting credit task, for example a specialist must choose which states inside the activity arrangement were dependable for the gotten reward. Support learning calculations along these lines are worried about tracking down the ideal grouping of activities through experimentation associations in a climate that expands the reward overtime. Reinforcement taking in calculations contrast from heading learning calculations in that they're not prepared on input/yield sets determining which activity is the awesome each state. All things considered; they're guided to the objective by the prizes they get. In an alternate manner, the prize obtained after each activity completely determines the difference to be tackled. Another distinction to regulated learning is that an undertaking regularly has no different preparing stages. All things considered; a few errands require consistent learning all through a specialist's life.Q-Learning is named an off-arrangement learning calculation since it unites to the ideal worth capacity autonomous of the investigation strategy being followed. All in all, the important part of the genuine investigation technique doesn't impact the value work, yet just the speed of union. there's additionally an onarrangement Q-Learning calculation called SARSA [30]' during which the investigation procedure is contemplated. Be that as it may, the two calculations merge to a comparable worth capacity when E, the probabilities of investigation, diminishes towards nothing.

Q-Learning

In this segment, we will in general inventory a transient prologue to the basic Q-learning procedure, that is the fundamental principle hypothesis of our calculation. Then, we will in general legitimize our Q-learning based

framework model. Q-learning is an Associate in off strategy support learning calculation that tries to peer out the lone activity which needs the given present state. It's about off-approach for the consequences of q-taking in which gains from activities that are outside of this arrangement, such as making arbitrary moves, so a strategy isn't needed. Furthermore, to all the more explicitly, q-learning looks to search out an approach that expands the whole reward.The 'q' in q-learning represents quality. At the point when q-learning is performed we will in general inventory a q-table or lattice that follows the state of [state, action] then we introduce our qualities to nothing. we will in general at that point update and store our q-values when a scene is finished. This q-table turns into a reference table for our representative to choose the solitary activity on refreshed q-esteem.

Initialize q-table qualities to nothing

Q = np.zeros((state_size, action size))

The following stage is exclusively for the specialist to act with the setting and make updates to the state activity sets in our q-table Q[state, action].

Q-Routing:

Q-Routing might be a disseminated, versatile, on-line steering calculation that utilizes the Q-Learning structure.

In Q-Routing, every hub chooses which adjoining hub to advance bundles. in order to weaken the run of the mill parcel conveyance time dependent on neighbourhood data. This is frequently significant on the grounds that, as organizations fill in size, the amount of directing data to be sent gets critical, and should affect the overall organization execution. Every hub must store directing data to frame the steering choice. This steering data should be refreshed consistently to mirror the current organization

Taking Action: Explore or Exploit

At the point when a specialist must demonstrate ideally in an obscure climate, it's to adjust the 2 restricting destinations of investigation and misuse. In order to search out an ideal control strategy, the specialist must adequately investigate the climate. Be that as it may, exploration is typically an upscale activity; in this way, it's reasonable for the specialist to exploit the information it's now educated. On the off chance that the specialist just adventures, it's going to not track down the ideal arrangement; in the event that it simply investigates, it burns through an inordinate measure of time investigating portions of the state space which are unimportant to the main job. All in all, we've to both investigate and adventure. The inquiry emerges of when to investigate and when to exploit, which characterizes the compromise which most support learning calculations need to tackle. The looks at this compromise and presents strategies for adjusting each objective. There are two significant kinds of investigation techniques: coordinated and undirected investigation. Undirected investigation is predicated on arbitrariness, for example a specialist may make an irregular move with a reliable likelihood dissemination. On the contrary hand, coordinated investigation techniques gauge the investigation utility of activities, and pick activities which boost the normal data acquired by investigation.

The least difficult strategy for undirected investigation chooses activities from a steady probability dispersion. This prompts a stochastic interaction over the state space and utilizes recently scholarly information. By choosing E-ravenous activity, a specialist chooses activity with greatest worth capacity more often than not, however with a little likelihood it chooses unintentional activity in uniform likelihood appropriation.

Q-Learning can perform coordinated investigation and upheld a simply manipulative arrangement by setting all underlying qualities to an overestimate of the value of each state-activity pair. Misuse will at that point select neglected activities in light of high worth making such a move. Over the long haul, this guarantees that every one state-activity successions are followed. The investigation is progressively diminished in light of the fact that the assessed esteem work combines to truth esteem work. The sole issue in the framework is that it functions admirably just in fixed conditions. Since the award work changes during a unique climate, union to at any rate one approach isn't alluring; hence, this strategy is normally used in blend with other investigation methods.

Updating the q-table:

The updates happen after each and each progression or activity closes once a scene is done. During this case the completed activity is arriving at some terminal state by the specialist. A terminal state as an illustration is generally something like arriving on a checkout page, arriving at the tip of some game, completing some ideal target, and so forth the specialist will not learn after one scene, anyway at last with enough investigating (steps and scenes) it'll merge and gain proficiency with the ideal q-qualities or q-star (Q*).

Here zone unit the three essential advances:

Specialist begins in during a state (s1) makes a move (a1) and gets a reward(r1)

Specialist chooses activity by referring to Q-table with most noteworthy worth (max) OR by arbitrary (epsilon, ϵ)

Update q-values

Update Q esteems

Q[state, action] = Q[state, action] + lr * (reward + gamma * np.max(Q[new state, :]) - Q[state, action])

In the report on top of 1 or two factors that we haven't referenced notwithstanding. What's occurring here is we will in general differ our q-values upheld by differentiation between the limited new qualities and furthermore the past qualities. we will in general limit the new qualities exploitation gamma which alter our progression size exploitation utilizing learning rate (lr). The following are a few references.

Learning Rate: It or learning rate, typically seen as alpha or α , will just be illustrated as what amount you need to acknowledge the new worth versus the past worth. above we are taking greatness among new and previous at that point increasing that value by the preparation rate. This value at that point gets extra to our past q-esteem which principally moves it inside the heading of our most recent update.

Gamma: gamma or γ may be a decrease issue. it's wont to adjust quick and potential compensation. From our update rule on top, you will see that we will in general utilize the rebate to the since quite a while ago run reward. generally, this value will differ wherever from zero.8 to 0.99.

Prize: reward is the cost taken when completing a particular activity at a given state. a blessing will occur at some random time step or exclusively at the terminal time step.

Max: np.max () utilizes the NumPy library and is taking the chief of the since quite a while ago run reward and applying it to the prize for this state. What this may do is sway this activity by the conceivable potential compensation. This is generally the pleasantness of q-learning. We're allotting potential compensation to current activities to assist the specialist with picking the most flawlessly awesome come activity at some random state.

Well, that is it, quick and painless (ideally). we tend to reference that q-learning is Associate in off-approach support learning algorithmic standard. We will in general show the basic update rule for q-learning exploitation utilizing some fundamental python pseudocode and that we checked on the particular contributions to the algorithmic principle. we tend to discover that q-learning utilizes potential compensations to impact the current activity given a state and afterward assists the specialist with picking activities that boost complete prize.



Fig 2. System Architecture

System Architecture

As all the information exchanges between the network nodes with each other. All the information gathered by the network topology like link quality, energy consumption, link quality will be evaluated by taking action in Q-algorithm using the reward function. Using Q-learning algorithm in reinforcement learning to implement multipath routing needs Q table. After constructing q- table initialize all the values by zeros. After taking action, select any one among the possible options to find the current state based on the highest Q value. Then update the new q value. The process will be repeated again until to reach the goal state. Based on the updated q value the optimal route will be selected.



Fig. 3: Flow of the multipath routing algorithm

As shown in the above diagram we will initialise the nodes with the energy and link quality and we exchange information to all the nodes and transfer the information to the Q-value algorithm. In this Q-value algorithm we have state, action and reward function. From this we choose the best path and update the Q-value based on the reward function that is obtained by using this algorithm. Then finally we select the most optimal path in the

network with the updated Q-value. We can compare the path bridging technique and the Q-learning algorithm. If we compare the reliability Q-learning algorithm works better than the path bridging technique. Hence Q-learning is most efficient among the proposed multipath routing protocols.

Simulation and Results

Simulations have been performed with multiple nodes.



Fig. 4. Network graph of 27 nodes. With node 0 as source and node 26 as destination main path and sub path was found using Multipath, Sub path and Bridge(MSB) based routing algorithm.

[0, 1, 3, 5, 9, 11, 13, 15, 17, 21, 23, 25, 26] [0, 2, 4, 6, 8, 10, 14, 16, 18, 20, 22, 24, 26]

Fig 5. Main path and sub path of the graph of fig.4

Fig. 4 and 5 shows the graph of 27 nodes



Fig.5 Elapsed time of multipath route identification using MSB protocol and Q-Learning multipath routing algorithm

Fig 5 shows the elapsed time to find multipath through the multipath, sub path and bridge node finding algorithm and the Q-learning based algorithm. The graph shows the path finding time is lesser in Q-learning algorithm compared to the existing algorithm.

Conclusion

In this paper we built an efficient optimal multipath routing protocol using Q-learning technique. We have seen many existing multipath routing protocols which achieves reliability and some real time requirements of wireless sensor networks in the IoT. Here we considered an existing system, that is, path bridging technique and also implemented the multipath routing using Q-learning. In this paper we conclude that Q-learning is more efficient than the existing multipath routing protocols.

References

- 1. H. Chan, A. Perrig, D. Song, "Random key predistribution schemes for sensor networks," IEEE Symposium on Security and Privacy (SP'03), Oakland, CA, May 2003.
- 2. S.K. Das, A. Mukherjee, et al., "An adaptive framework for QoS routing through multiple paths in ad hoc wireless networks," J. Parallel Distributed Computing, 63(2003)141–153.
- AI-Karaki J. N., Kamal A. E. (2004) Routing techniques in wireless sensor networks: A survey. IEEE Wireless Communications 11(6): 6–28
- Akkaya K., Younis M. (2004) A survey on routing protocols for wireless sensor networks. Ad Hoc Networks 3: 325–349
- 5. Heikalabad, S. R., Rasouli, H., Nematy, F., & Rahmani, N. (2011). Gempar: Qos and energy aware multipath routing algorithm for real-time applications in wireless sensor networks. CoRR, abs/1104.1031.
- 6. Wang, Y.-H., Mao, H.-J., Tsai, C.-H., & Chuang, C.-C. (2005). Hmrp: Hierarchy-based multipath routing protocol for wireless sensor networks. In Proceedings of EUC Workshops'05, (
- Almalkawi I. T., Zapata M. G., Al-Karaki J. N. (2011) A secure cluster-based multipath routing protocol for wmsns. Sensors 11(4): 4401–4424
- 8. Chen, Y., Chan, E., & Han, S. (2005). Energy efficient multipath routing in large scale sensor networks with multiple sink nodes. Advanced Parallel Processing Technologies, 3756, 390–39
- 9. De S., Qiao C., Wu H. (2003) Meshed multipath routing with selective forwarding: An efficient strategy in sensor networks. Computer Networks 43: 481–497
- 10. Li, S., Neelisetti, R. K., & Liu, C. (2010). Efficient multi path protocol for wireless sensor networks. International Journal of Wireless and Mobile Networks, 2(1).
- 11. Lu Y. M., Wong V. W. S. (2007) An energy-efficient multipath routing protocol for wireless sensor networks. International Journal of Communication Systems 20: 747–766
- 12. Pratheema S., Srinivasagan K. G., Naskath J. (2011) Minimizing end-to-end delay using multipath routing in wireless sensor networks. International Journal of Computer Applications 21(5): 20–26
- Fragouli C., Boudec J.-Y. L., Widmer O. (2006) Network coding: An instant primer. SIGCOMM Comput. Commun. Rev. 36: 63–68
- Ganesan D., Govindan R., Shenker S., Estrin D. (2001) Highly-resilient, energy-efficient multipath routing in wireless sensor networks. SIGMOBILE Mob. Comput. Commun. Rev. 5: 11–25
- 15. Garcia-Hernandez C. F., Ibarguengoytia-Gonzalez P. H. et al (2007) Wireless sensor networks and application: A survey. International Journal of Computer Science and Network Security 7: 264–273
- 16. Han, S., Zhong, Z., & Li, H. (2008). Coding-aware multi-path routing in multi-hop wireless networks. Proceedings of IPCCC'08.
- 17. Heikalabad, S. R., Rasouli, H., Nematy, F., & Rahmani, N. (2011). Qempar: Qos and energy aware multipath routing algorithm for real-time applications in wireless sensor networks. CoRR, abs/1104.1031.
- 18. Hind, A., & Anjali, A. (2009). A survey on fault tolerant routing techniques in wireless sensor networks. In Proceedings of the 2009 third international conference on sensor technologies and applications.
- 19. Karlof, C., & Wagner, D. (2003). Secure routing in wireless sensor networks: Attacks and countermeasures. Proceedings of first IEEE international workshop on sensor network protocols and applications.
- 20. Li, S., Lim, A., Kulkarni, S., & Liu, C. (2007). A routing algorithm for maximizing throughput and minimizing delay in wireless sensor networks. In Proceedings of MILCOM 2007.
- 21. Lou, W. (2005). An efficient n-to-1 multipath routing protocol in wireless sensor networks. In Proceedings of mobile adhoc and sensor systems conference.
- 22. Mueller S., Tsang R., Ghosal D. (2004) Multipath routing in mobile ad hoc networks: Issues and challenges. In Performance Tools and Applications to Networked Systems 2965: 209–234
- 23. Saleem, K., Fisal, N., & Hafizah, S. (2009). A self-optimized multipath routing protocol for wireless sensor networks. International Journal of Recent Trends in Engineering (IJRTE), 2(1).
- Stavrou E., Pitsillides A. (2010) A survey on secure multipath routing protocols in wsns. Computer Networks 54(13): 2215–2238
- 25. Wu, J., Dulman, S., Havinga, P., & Nieberg, T. (2004). Multipath routing with erasure coding for wireless sensor networks. In Proceedings SAFE and ProRISC, pp. 181–188.