Improving the Performance of P2P Routing using Adaptive Machine Learning

Amruta Deshmukh, M. A. Pund

Abstract*:*Effective Peer to Peer (P2P) routing is a very complex problem to solve, due to the fact that the nodes are not location aware, and the network is constantly changing in terms of number of nodes, link quality and other network parameters. For such adhoc networks protocols like adhoc on demand distance vector routing (AODV), AODV with multi-channel (AOMDV), viceroy and others have been proposed. These protocols show a moderate to low quality of service (QoS) for the network when routing is concerned, in some scenarios these protocols have low delay, while in some other scenarios these protocols have better energy efficiency than others. In this paper, we propose a delay and energy efficient machine learning based routing protocol for P2P networks, which allows the network designers to improve upon the network QoS without compromising on the routing efficiency of the network. The results indicate that our proposed protocol is better in terms of both delay and energy efficiency when compared to the existing ones, and provides alternate routing paths, in case of node or link down scenarios. We also propose some further work which can be taken up by using the proposed protocol to make the network more secure and efficient in terms of overall privacy.

Index Terms*: *delay, energy, machine learning, P2P, QoS.

I. INTRODUCTION

Peer to Peer (P2P) routing is generally performed on networks which are co-operative in nature. This kind of routing requires careful design of the routing protocol, which is often dependent on the application of the network. For example, for P2P file sharing applications like uTorrent, the network design is based on leachers and seeders, here the network doesn't provide any delay constraints for delivery of the packets from the seeders to the leachers, but instead depends on the reliability of the file transfer, and then there are applications like P2P sensor networks, where the designers have to apply delay and value constraints for the network so that reliable data can be communication between peers in a restricted amount of time. Thus, to design a P2P routing protocol, the following design steps are needed,

- Initialize the input network parameters like the network size, number of nodes, the configuration of each of the nodes, the channel type for which the design is being done, and the micro communication infrastructure for each of the node or peer.
- Once the network is initialized, then the nodes need to be configured based on the constraints on delay, energy efficiency, packet delivery efficiency, consistency of packet delivery and other output related parameters. Setting these parameters is of utmost importance, so that the resulting network's behaviour is the way how it is expected by the network designers.
- After setting up the nodes, there is need of a peer discovery protocol, where each node or peer can discover the approximate to actual location of the peers nearby. This is generally done with the help of beacon-based techniques, which are low power and evaluate the node position based on the delay with which the acknowledgement of the beacon is received at the receiver side.
- The peer discovery module needs to re-run in order to update the locations for each node, thus there is a need of a synchronization layer which will time synchronize each of the node's beacon sending rate, so that none of the beacons are dropped, and a new rate is assigned to each new node which joins the network. This allows all the nodes to be in approximate synchronization w.r.t. a single clock period value.
- Upon discovery, protocols like AODV, AOMDV, VICEROY and others are applied at each of the nodes to obtain the routing path between the sender and receiver. These paths are optimized based on the constraint's setup by the designers, and usually have good level of packet delivery and data throughput
- The selected paths might break down due to a many reason like node energy being too low, the link between the nodes break, and many others. To avoid this, the designers need to provide backup paths for the nodes, so that the nodes can reconfigure themselves in runtime during breakdowns. Protocols based on node recovery, alternate path provision, link recovery and others are used by the network designers to perform this task.

In this paper, we propose a new machine learning based protocol which follows all the above procedures and produces an effective routing protocol to be used in P2P routing. This protocol is described in section III of this paper, the next section describes the standard P2P routing protocols, and finally we conclude this text with a comparative analysis between the standard P2P routing protocols and the proposed routing protocol, in order to evaluate its efficiency.

Revised Manuscript Received on May 24, 2019.

Ms. Amruta Deshmukh, PhD Student, Computer Science, Sant Gadge Baba Amravati University Amravati, Maharashtra (India)
Dr. M. A. Pund, Dept. of Computer Science,Prof. Ram Meghe Institute of Technology and Research Amravati, Maharashtra (India)

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

II. LITERATURE REVIEW

As of late, numerous creators have perceived the potential edges of P2P Cloud models. In [1] the creators outlined a general structure to help completely disseminated applications running over an extensive scale and dynamic pool of assets. The creators list many tattle-based conventions which will be connected to make the sub mists and to execute bootstrapping, recognition and the executive's administrations. Expanding on the most arrangement of [2], amid this paper we will in general blessing a reasonable structure, with a worldview usage, of a P2P Cloud. Framework which blends choices from the VC model and Cloud figuring worldview. It should be found that the Cloud@Home configuration relies upon brought together components, though allowing completion clients to contribute additional assets. On the contrary hand, our proposition is completely suburbanized, and needn't bother with any focal bookkeeping administration. At the season of composing we tend to aren't checked out any execution of a Cloud@Home worldview in [3] an extraordinary bearing is taken: the creators propose an appropriated registering stage known as Nano learning Centers (NaDa). Nada utilizes home portals, constrained by ISPs, to deliver registering and capacity administrations. Utilizing an oversaw shared model, nothing types a disseminated learning focus framework. We will in general at last notice Wuala10 as Associate in Nursing case of Cloud based for the most part stockpiling administration. Wuala licenses clients to exchange house on their difficult circles to get scrambled lumps of records transferred by various clients. Another dispersed Cloud stockpiling framework is spoken to in [19]. These territory unit carefully capacity administrations, so they give no help to beating process undertakings. Our plan, on the contrary hand, goes for giving every calculation and capacity administrations. In the IaaS display, the apparatus is given by the customer through a little interface, that likely could be an online program, and in this way the buyers have Associate in Nursing outright obligation to deal with their records exclusively. The entire application is that the duty of the supplier [5] and subsequently the Cloud customer has no administration over this foundation. A PaaS Cloud gives virtual machines, in task frameworks, application administrations, improvement systems, exchanges and the board structures for applications created by the Cloud customer. Amid this model, the clients or clients will build up their applications at interims the cloud framework or utilize their applications customary. The administration provider deals with the cloud framework, programming bundle, PC code or native [4]. Be that as it may, the buyer's region unit responsible for the establishment and upkeep of the applications they're creating. At last, Associate in Nursing IaaS Cloud gives basic registering capacities like virtual machines, virtual memory, virtual framework and diverse equipment similarity as an arrangement for buyers. IaaS provider deals with all framework though the customer is responsible for the occasion viewpoints [4], [5]. Importance the occasion demonstrates we've the ensuing sorts [5]: non-open, open, network and half breed. As of late has developed the P2P cloud thought, which blends cloud and P2P systems. This sort of Cloud processing is predicated on a completely disseminated Cloud structure and might be useful for a couple of utilization outcomes. Creators in [4] express that a P2P Cloud grants association or perhaps individual to make a figuring framework out of existing assets, which might be basically apportioned among totally extraordinary undertakings. Potential edges of P2P Cloud processing have perceived all through the most recent year and various other associated work are arranged. Cloud@Home is presented in [6], that could be a half breed framework that blends qualities from volunteer workstation and distributed computing ideal models. In [7] proposed another disseminated P2P Cloud framework that is planned to deliver capacity administration exclusively. Partner in Nursing configuration Associate in Nursing its worldview to deliver a framework and fix through a P2P cloud territory unit presented in [4]. Creators in [7] blend P2P and distributed computing to get a cross breed and appropriated structure for sight and sound framework gushing administration.

The accompanying systems will be wont to actualize the P2P cloud models,

1. Centralized procedure
2. Hierarchal procedure
3. Flooding procedure
4. DHT based procedure
5. Gossip based procedure

The concentrated strategy could be an answer utilized mainly in early P2P gushing frameworks. Amid this system, the learning concerning all hubs, e.g., their location or offered data measure, is whole in an extremely concentrated index and in this way the unified catalog is responsible to build and keep up the topology. CoopNet [13] and DirectStream [14] zone unit 2 test frameworks that utilization the focal procedure. Since the focal server includes a world read of the overlay organize, it will deal with hubs change of uprightness and going appallingly rapidly. one among the contentions against this model is that the server winds up one motivation behind disappointment, and on the off chance that it crashes, the same hub will be a piece of the framework. The quantifiability of this model, additionally, is another disadvantage. Notwithstanding, these issues will be settled if the focal server is supplanted by a gathering of conveyed servers. Following response for finding action hubs is utilizing a hierarchal approach. This methodology is utilized in numerous frameworks, like Nice [15], ZigZag [16], and BulkTree [17]. In Nice and ZigZag, for instance, assortment of layers zone unit made over the hubs, indicated unequaled low layer contains every one of the hubs. The hubs amid this layer territory unit arranged into certain bunches, in accordance with a property sketched out inside the algorithmic program, e.g., the idleness between hubs. One hub in each bunch is picked as a head, and subsequently the picked head for each group turns into an individual from 1 higher layer. By bunch, the hubs amid this layer and picking a head in each group, they type resulting layer, etc., till it at long last winds up in a very layer comprising of one hub. This single hub, that could be an individual from all layers is named the meeting reason. At whatever point a substitution hub comes into the framework, it sends its be a piece of solicitation to the meeting reason. The meeting hub restores a stock of every single associated hub on resulting down layer inside the order.

The new hub tests the rundown of hubs and finds the chief right one and sends its be a piece of solicitation thereto hub. The strategy rehashes till the new hub finds a balance inside the structure wherever it gets its ideal substance. Even though this answer tackles the quantifiability and thusly the single motivation behind disappointment issues inside the focal strategy, it's a moderate assembly time. The third approach to discover hubs is that the controlled flooding, that is initially arranged by Gnutella [18]. GnuStream [19] could be a framework that utilizes this plan to look out movement hubs. Amid this framework, each hub includes a neighbor set, that could be a halfway rundown of hubs inside the framework. At whatever point a hub looks for a provider, it sends its inquiry to its neighbors. Each hub advances the solicitation to any or most of its own neighbors aside from the one World Health Organization has sent the solicitation. The inquiry includes an opportunity to-live (TTL) value, that diminishes when each rebroadcasting. The telecom proceeds till the TTL ends up zero. If a hub that gets the solicitation fulfils the hub decision limitations, it'll answer to the underlying sender hub. This strategy has 2 fundamental disadvantages. To start with, it produces a noteworthy amount of system traffic and second, there's no assurance for finding relevant providers. Another response for finding the action hubs is to utilize Distributed Hash Tables (DHT), e.g., Chord [20] and Pastry [21], SplitStream [22] and [23] zone unit 2 tests that activity over a DHT. In these frameworks, each hub keeps a steering table together with the location of different hubs inside the overlay organize. The hubs, at that point, will utilize these directing tables to seek out movement hubs. This philosophy is adaptable, and it finds right providers rather rapidly. It ensures that if right providers region unit inside the framework, the algorithmic program discovers them. Be that as it may, it needs further exertion to oversee and keep up the DHT. The last way to deal with pursuit out action hubs is that the tattle-based strategy. A few calculations zone unit arranged upheld this model, e.g., NewCoolstreaming [24], DONet/ - Coolstreaming [25], PULSE [26], gradienTv [9] and [10] utilize a tattle produced arbitrary overlay system to search for the action hubs. We will in general utilize the tattle produced Gradient overlay [7, 8] for hub revelation in Sepidar and Glive. inside the tattle-based philosophy, each hub sporadically sends its information accommodation information to its neighbors, a halfway read of hubs inside the framework, to modify them see appropriate providers, World Health Organization have learning they're longing for. This system is versatile and disappointment tolerant, anyway because of the irregularity property of neighbor decision, for the most part the appropriate providers aren't found in sensible time.

III. MACHINE LEARNING BASED P2P ROUTING (MLPR) PROTOCOL

The proposed MLPR protocol has the following input and output structure,

Inputs:

No of Nodes = n
Energy Level of each node = E_i
Link Quality of data paths = LQ_i
Locations of Nodes= L_i
Optimization parameters = power, delay, PDR
Nodes are placed /connected in a P2P Protocol Structure.

Outputs:

Paths between selected source and destination (paths outputs)
Data rates of each nodes for transmission on path (D_{ro})

Assuming the following network structure for the protocol, where we must communicate data from a source node (S) to a destination node (D), and there are multiple nodes or peers in between.

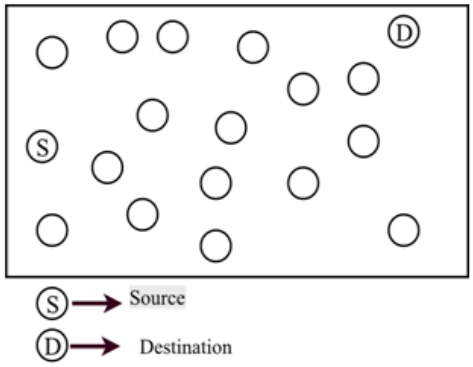

Figure 1: Network Structure

The flow of the protocol can be described as follows,

i. Discover the locations of source and destination L_{src}, L_{dst} and find Euclidean distance between them.

Let this Euclidean distance be the reference distance; D_{ref}.

$$\text{Dref} = \sqrt{(Xsrc - Xdst)^2 + (Ysrc - Ydst)^2}$$

ii. Machine Learning Step:

No. of solutions = Ns
No. of iterations = Ni
Learning convergence = Lc (0 to 1)
Max nodes in each solution = Nmax
for all nodes between source and destination, nodes are found using the simple arithmetic lemma,

D_{NS} and D_{ND}< Dref
D_{NS}= Distance of node from source
D_{ND}= Distance of node from destination

a) Generate a random path with Nmax nodes. Let the nodes in this path be,

$NP_1, NP_2, \ldots\ldots\ldots\ldots\ldots\ldots NP_{Max}$

b) Evaluate the learning matrix (LM)

$$LM = \sum_{i=2}^{Nmax} \frac{\text{Di,i} - 1}{\text{E(i} - 1)} + [LQi, i - 1]^{-1}$$

For i>1

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

$$= \sum_{i=2}^{Nmax} \frac{\text{Di, src}}{\text{Esrc}} + [LQi, src]^{-1}$$

c) Repeat steps (a) and (b) above for Ns times.

d) Evaluate the learning convergence from the threshold as,

$$L_{CTH} = \left(\frac{\sum_{i=1}^{Ns} L_{M_i}}{N_s} \right) * L_C$$

e) If a solution (i) has,

$$L_{Mi} > L_{CTH}$$

Then it is discarded and replaced by a new solution in the next iteration.

For $L_{Mi} \leq L_{CTH}$

The solution is accepted and passed to the next iteration.

f) Repeat a to e for N_i times

g) At the end of N_i iterations, select the solutions with minimum values of LM.

h) Suppose we select top three solutions LM1, LM2, LM3, then if all nodes in LM1 are working then it is used for routing.

If any node is down, then rate of LM2 is used, the LM3 and so on.

iii.AI step:

For each combination of source and destination we get top 'm' paths,

P1, P2, P3 ……...Pm

Selection of paths based on nodes state is done by AI (step 'h' of machine learning)

This ensures that the best path is selected for routing, and in case of any routing break down, we have alternate paths already in queue, so that the routing goes on, without any breaks.

We evaluated the results of the MLPR protocol and compared it with the standard VICEROY protocol, and the results shown in the next section show that the proposed MLPR protocol is better in terms of both delay and energy efficiency than the standard VICEROY protocol.

IV. RESULTS AND ANALYSIS

We tested our machine learning and AI based protocol for various numbers of communications in the network. The network parameters used are defined as follows,

Table I: Network parameters

Parameter	Value
Routing algorithm	VICEROY and MLPR
Number of nodes	30 to 100
Network type	P2P
Queue	Priority drop tail
Network size	300 m x 300 m
MAC Type	802.11
No. of communications	1-10

From the literature survey, we found that the VICEROY protocol is better in comparison with any of the existing P2P routing protocols, thus it was used as a baseline for comparison with our proposed MLPR algorithm. The variation of network delay w.r.t. number of communications can be shown as follows,

Table II: Delay v/s Number of communications

0	Delay (ms) Viceroy	Delay (ms) MLPR
1	0.00889	0.00800
2	0.00341	0.00307
3	0.00479	0.00431
4	0.00750	0.00675
5	0.00996	0.00896
6	0.01500	0.01350
7	0.02270	0.02043
8	0.03570	0.03213
9	0.04470	0.04023
10	0.05270	0.04743

From the table we can observe that the delay of the proposed MLPR algorithm is lower than almost 10% as compared to the VICEROY protocol, this is due to the fact that the proposed algorithm selects the path by minimizing the distance between the routing nodes, thus there is a minimal delay between the packet transmission and reception process. Similar comparison was done for the network's energy efficiency, and the following results were obtained,

Table III: Energy comparison

No. of communications	Energy (mJ) Viceroy	Energy (mJ) MLPR
1	3.59	1.80
2	4.75	3.17
3	5.99	4.49
4	7.23	5.78
5	8.79	7.33
6	9.81	8.41
7	10.23	8.95
8	11.22	9.97
9	12.15	10.94
10	13.02	11.84

The network's energy efficiency is improved by more than 15% when using MLPR protocol, this is again because the learning formula has energy parameter for optimization. The next comparison was done for the packet delivery ratio of the network, and the following results were obtained,

Table IV: Comparison of PDR

No. of communications	PDR (%) Viceroy	PDR (%) MLPR
1	94.5	98.5
2	94.6	98.6
3	94.54	98.54
4	94.66	98.66

5	93.32	97.32
6	94.5	98.5
7	94.6	98.6
8	94.8	98.8
9	95.1	99.1
10	94.1	98.1

From the above table we can see that the packet delivery ratio is also improved by more than 5%, this is due to the reduction in distance between the nodes, as the distance between the nodes reduces, the effectiveness with which the packets are being transmitted improves, thereby improving the overall delivery ratio of the network. Thus, we observe that the overall QoS of the network is improved using the MLPR protocol for P2P routing.

V. CONCLUSION

The observed results demonstrate that the QoS of the MLPR protocol is superior than the existing standard P2P routing protocols like VICEROY. The AI layer also reduces the delay in communication for the network and keeps a high packet delivery ratio for any type of network scenario. The proposed technique outperforms all other techniques in terms of network lifetime as well, and it also performs well with varying number of nodes and thus is suitable for any network size.

VI. FUTURE WORK

The proposed protocol demonstrates good quality of network QoS when compared with other standard protocols, and thus can be used for any real-life wireless application. In future, we plan to apply this optimization to security with the help of blockchain based techniques, so that the overall communication privacy can be enhanced without compromising on the network QoS parameters.

REFERENCES

[1] P. Francis, "Yoid: Extending the Internet Multicast Architecture," unpublished, Apr. 2000. http://www.aciri.org/yoid/docs/index.html
[2] M. Waldman, A. D. Rubin, and L. F. Cranor, "Publius: A Robust, Tamper-apparent, Censorship-safe, Web Publishing System," Proc. ninth USENIX Security Symp., Denver, CO, USA, 2000.
[3] (2002)Gnutella ultrapeers, accessible at http://rfcgnutella.sourceforge.net/designer/testing/Ultrapeers_1.0.html mlSosinsky, B.: Cloud Computing Bible. In: Wiley Publishing Inc., Indianapolis, IN, USA (2011).
[4] H. J. Siegel, "Interconnection Networks for SIMD Machines," Computer, vol. 12, no. 6, 1979, pp. 57– 65.
[5] S. Rhea et al., "Upkeep free Global Storage in Oceanstore," IEEE Internet Comp., 2001
[6] A. Rowstron and P. Druschel, "Cake: Scalable, Distributed Object Location and Routing for Large-scale Peer-to-peer Systems," Proc. Middleware, 2001
[7] P. Maymounkov and D. Mazieres, "Kademlia: A Peer-to-Peer Information System Based on the XOR Metric," Proc. IPTPS, Cambridge, MA, USA, Feb. 2002, pp. 53–65
[8] Y. Rekhter and T. Li, An Architecture for IP Address Allocation with CIDR, IETF Internet draft RFC1518, accessible at http://rfc.net/rfc1518.html, 1993.
[9] M. Castro et al., "Recorder: A Large-Scale and Decentralized Application-Level Multicast Infrastructure," IEEE JSAC (extraordinary issue on Network Support for Multicast Communications), October 2002.
[10] S. Iyer, A. Rowstron, and P. Druschel, "Squirrel: A Decentralized Peer-to-Peer Web Cache," Proc. 21st Symp. Standards of Distributed Computing (PODC), Monterey, California, USA, July 21-24 2002.
[11] C. Plaxton, R. Rajaraman, and A. Richa, "Getting to Nearby Copies of Replicated Objects in a Distributed Environment," Proc. ninth Annual ACM Symp. Parallel Algorithms and Architectures, 1997.
[12] B. Y. Zhao et al., "Embroidered artwork: A Resilient Global-Scale Overlay for Service Deployment," IEEE JSAC, vol. 22, no. 1, Jan. 2004, pp. 41-53.
[13] (2002) Gnucleus, the Gnutella Web reserving framework, accessible: at http://www.gnucleus.com/gwebcache/
[14] A. Rowstron and P. Druschel, "Stockpiling Management and Caching in Past, a Large-Scale, Persistent Peer-to-Peer Storage Utility," Proc. eighteenth ACM Symp. Working Systems Principles, Oct. 2001, pp. 188-201.
[15] F. Dabek et al., "Towards a Common API for Structured Peerto-peer Overlays," Proc. second Int'l. Wksp. Distributed Systems (IPTPS 2003), Berkeley, California, USA, Feb. 20-21, 2003.
[16] L. P. Cox, C. D. Murray, and B. D. Respectable, "Pastiche: Making Backup Cheap and Easy," SIGOPS Op. Sys. Rev., vol. 36, no. SI, 2002, pp. 285-98.
[17] F. Dabek et al., "Wide-region Cooperative Storage with CFs," Proc. eighteenth ACM Symp. Working Systems Principles, 2001, pp. 202– 15.
[18] L. Peterson et al., "A Blueprint for Introducing Disruptive Technology into the Internet," SIGCOMM Comp. Commun. Rev., vol. 33, no. 1, 2003, pp. 59– 64.
[19] J. Kubiatowicz et al., "Oceanstore: An Architecture for Global-Scale Persistent Storage," Proc. ACM ASPLOS, Nov. 2002.
[20] F. Zhou et al., "Rough Object Location and Spam Filtering on Peer-to-Peer Systems," Proc. Middleware, June 2003.
[21] U. Manber, "Finding Similar Files in a Large File System," Proc. USENIX Winter 1994 Conf., Jan. 1994, pp. 1– 10.
[22] (2001) Gnutella advancement gathering, the Gnutella v0.6 convention, accessible at http://groups.yahoo.com/gathering/the gdf/records/
[23] S. Ratnasamy et al., "A Scalable Content Addressable Network," Proc. ACM SIGCOMM, 2001, pp. 161– 72.
[24] S. Q. Zhuang et al., "Bayeux: An Architecture for Scalable and Fault-Tolerant Wide-Area Data Dissemination," Proc. eleventh Int'l. Wksp. System and Op. Sys. Backing for Digital Audio and Video, 2001, pp. 11-20.
[25] B. Karp et al., "Prodding Adoption of DHTs with OpenHash, a Public DHT Service," Proc. third Int'l. Wksp. Shared Systems (IPTPS 2004), Berkeley, California, USA, Feb. 26-27, 2004.
[26] A. Rowstron et al., "Copyist: The Design of a Large-Scale Event Notification Infrastructure," Proc. third Int'l. Wksp. Organized Group Commun. (NGC2001), London, UK, Nov. 2001, pp. 30-43.

AUTHORS PROFILE

Ms. Amruta Deshmukh received M.E. degree in Wireless Communication and Computing from Rashtrasant Tukadoji Maharaj Nagpur University and B.E. degree in computer science and engineering from Santa Gadge Baba Amravati University, India. She is currently working toward the Ph.D. degree in the Department of Computer Science and engineering, Santa Gadge Baba Amravati University, India.Her research interest includes Machine Learning, Wireless Communication between network Peers, Cloud Computing and Artificial Intelligence.

Dr. M. A. Pund received the Ph.D., ME and BE degrees in Computer science and Engineering from Santa Gadge Baba Amravati University, India. He has been working as an Associate Professor at PRMIT-Prof Ram Meghe Institute of Technology and Research in Amravati, Maharashtra, India.

Dr. Pund has numerous research publications and Patents under his credentials. His Research interest includes Artificial Intelligence, E-Learning, Mobile and Cloud Computing, Machine Learning and Image Processing.