Reliability analysis of Structured P2P System

Tai Zhang, Sheng Wang, Dan Liao

School of Communication and Engineering, University
of Electronic Science and Technology of China

Email: arcaffier@163.com, wsh_keylab@uestc.edu.cn, liaodan@uestc.edu.cn

Abstract—**Reliability is most important for structured p2p systems and how to exactly evaluate the metric-"loss rate" which is tightly related to reliability is a hot topic and challenge. Many research works of the loss rate only consider the situation that the loss rate is caused by next hop node failure (NF). Indeed, the lost of queuing message (QF) on the failure node also contribute to the loss rate. This paper presents an analytical model of "loss rate" caused by NF and NQ. In order to achieve the optimal performance of system through minimizing the NQ, we propose two methods: process power enhancing (PPE) and Space Partition (SP). In the heterogeneous systems, we can obtain the minimal loss rate via improving each node's process power; in the homogeneous structured p2p systems, the partition space method can evenly partition the ID space so as to ensure the traffic load is uniformly distributed over all nodes.**

Keywords-reliability; loss rate; Chord; Optimization;

I. INTRODUCTION

Reliability is a key concern for structured p2p networks, it is a major metric to measure system's robustness and stability. Many research works for reliability of structured p2p network have been proposed in recent years. The works can be generally classified into two categories denoted by mechanism design and enhancement [1][2][3] and probability analysis of system partition[4][5]. The former works usually improve the stability mechanism through maintaining more number of states of neighbours by each node, it can dramatically enhance the stability of system, however, it also bring in more additional overhead; the later works only consider the probability of loss rate caused by next hop node failure (we abbreviate it as NF) and not consider the loss rate resulted by the lost of message queuing of node failure (we abbreviate it as QF). Indeed, in real p2p systems, peers are heterogeneous which have different process power, capacity, bandwidth and so on, thus some nodes with lower process power and bandwidth may usually deal with a large number of queries, once the arriving rate of messages exceed the threshold of the node process, the subsequent messages may queue on the node. If this node suddenly fail or leave, all queuing messages of the node may lose. This situation may heavily decrease the performance of system because of amounts of queries are lost by the node failure.

In this work, we propose an analytical model to compute the loss rate resulted by NF and QF and in this model we consider two different p2p systems: heterogeneous and homogeneous systems respectively, then we present the probability of system can not satisfy the system Qos. In order to achieve the optimization performance of system, under the heterogeneous system, we assume that each node itself can enhance the process power to improve the performance of system and we give the value of process power which make the performance of system is optimal. Under the homogeneous p2p system, we propose "space partition" method to balance the load to maximize the performance of system.

The organization of this paper is structured as follows: Section 2 presents the analytical mode to compute loss rate. Section 3 proposes the optimization problem and detailedly describes the PPE and SP methods. At last, we conclude this paper in section 4.

II. ANALYTICAL MODEL TO COMPUTE THE LOSS RATE

In this section we present an analytical model to compute the loss rate. Firstly, we give some notations used in this paper as below.

T_{am} : Time interval of two keep-alive messages of each node sends to its member lists;

T_{res} : Response time, if a node does not receive the reply message from its member list, then it judges the member is failed;

T_{at} : Time interval of two keep-alive messages from a node to its routing table entries;

λ_i : Rate of query messages arrive at node *i*;

μ_i : Processing capability of node *i*;

r : The rate of node leave;

We illustrate the definition of T_{am} and T_{at} in the figure 1. As shown in the figure, node N8 periodically send keep-alive messages to its two members (N1 and N14), after received the message, N1 and N14 immediately reply a live-message to node N8, if N8 do not receive the reply message from N1 within T_{res} , then N8 judge that N1 is failed and then disseminate this information over system.

We assume that the query messages arrive at node *i* according to a Poisson process with rate λ_i and the processing capability of node *i* according to a exponential distribution with rate μ_i, then we get that

$$\rho_i = \frac{\lambda_i}{\mu_i},$$

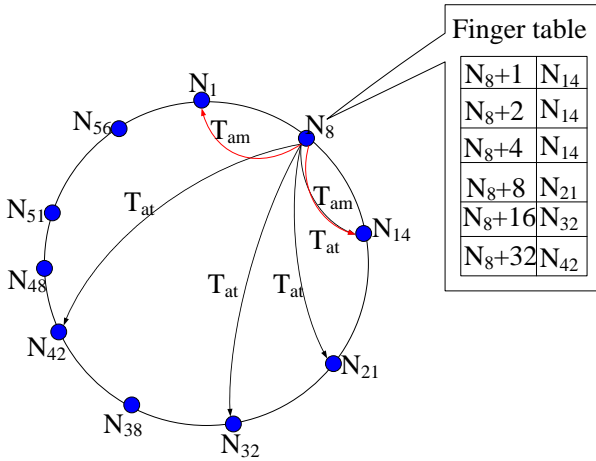and the length of messages queue on node *i* is

Finger table	
N_8+1	N_{14}
N_8+2	N_{14}
N_8+4	N_{14}
N_8+8	N_{21}
N_8+16	N_{32}
N_8+32	N_{42}

Figure 1. Keep-alive messages to members and table entries

$$\overline{N}_i = \frac{\rho_i^2}{1-\rho_i} \quad (1)$$

Note that, we use Chord [1] as the base of our analytical model. We get that the probability of forwarding a message to a failed node at each hop is:

$$P_f = 1 - \frac{1}{Tr}(1-e^{-Tr}) \quad (2)$$

Then we get the loss rate resulted by NF is:

$$R_{NF} = 1 - (1 - P_f(T_{am} + T_{res}, r))(1 - P_f(T_{at} + 2T_{res}, r))^{\log_2^N - 1} \quad (3)$$

In heterogeneous p2p system, nodes have different process power and bandwidth, node i, for example, if $\mu_i > \lambda_i$, there is no messages queue on it, otherwise there is queue on node i. We can obtain the loss rate caused by QF is:

$$R_{QF} = E(\overline{r})$$
$$= r\{E(\overline{N_1}) + E(\overline{N_2}) + ... + E(\overline{N_n})\} = r\sum_{i=1}^{n}E(\overline{N_i}) \quad (4)$$

Therefore, the loss rate of system caused by NF and QF is

$$L_{rate} = R_{NF} + R_{QF} \quad (5)$$

III. OPTIMIZATION PROBLEM OF PERFORMANCE

In order to improve the reliability of system, it should firstly minimize the loss rate. As the NF and QF are independent for each other, so we can get:

$$Min\ L_{rate} = Min\ R_{NF} + Min\ R_{QF}$$

As shown in the formula (3), we can see that, the value of NF is related to the system parameters such as T_{am}, T_{res} and T_{at}, from the figure 2, with the decrease of value of T_{am}, T_{res} and T_{at}, the probability of loss rate caused by NF is gradually approach to 1.
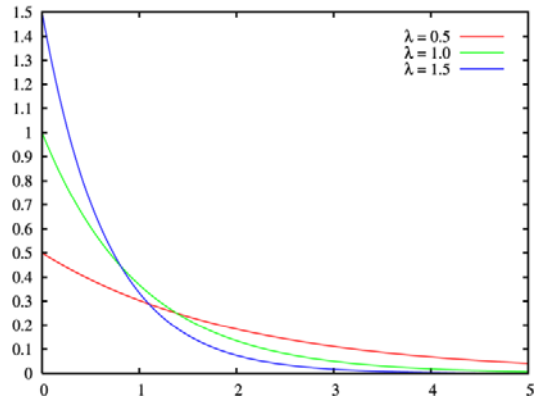

Figure 2. The CDF of R_{NF}

Therefore, the small value of these parameters which makes a large amount of the loss messages occur, however, we cannot persistently increase the value of these parameters, because it cannot detect the accident of node failure and decrease the reliability of system. There are some works have investigated how to set the appropriate value of these parameters to make the system more reliable, thus, for the limitation of space, in our work, the optimization problem only considers how to minimize the loss rate caused by NQ.

A. Process power enhancing method (PPE)

If the time of queue message stay on a node exceeds a time threshold value θ, then we say the node cannot satisfy the Qos, here, $\theta < T_{am} + T_{res}$. We can obtain the probability that a node can not satisfy the Qos is

$$p_{Qos} = P(W \geq \theta) = e^{-(\mu_i - \lambda_i)\theta} \quad (6)$$

Hence, we can obtain the probability of the system can not satisfy the Qos is :

$$p_{Qos} = \min_{i=1}^{n}(e^{-(\mu_i - \lambda_i)\theta}) \quad (7)$$

We consider a scenario that if the time of message queue stay on a node exceeds a unit time (1 second), the system has a deficit value c, if node deal with messages ahead of time, the system have a benefit value b, the cost of a node is proportional to process rate and in unit time is $h\mu$, if the income of a query process is completed is a, then in unit time, the income of λ numbers of queries process are completed is λa, then we can obtain that the average ahead of time of node i to deal with queries is

$$t_{ah} = \int_0^{\theta}(\theta - t)(\mu - \lambda)e^{-(\mu - \lambda)t}dt$$
$$= \theta - \frac{1}{\mu - \lambda}\{1 - e^{-(\mu-\lambda)\theta}\} \quad (8)$$

Average delay time is

$$t_{del} = \int_{\theta}^{\infty}(t - \theta)(\mu - \lambda)e^{-(\mu - \lambda)t}dt$$
$$= \frac{1}{\mu - \lambda}e^{-(\mu-\lambda)\theta} \quad (9)$$

The total income is

$$f(\mu) = \sum_{i=1}^{n}(-\mu h + \lambda a - \lambda c t_{del} + \lambda b t_{ah}) \ (10)$$

$$= \sum_{i=1}^{n}(-\mu h + \lambda a - \frac{c}{\mu-\lambda}e^{-(\mu-\lambda)\theta} + \lambda b\theta - \frac{\lambda b}{\mu-\lambda} + \frac{\lambda b}{\mu-\lambda}e^{-(\mu-\lambda)\theta})$$

$$= \sum_{i=1}^{n} -\mu h + \lambda(\frac{b-c}{\mu-\lambda})e^{-(\mu-\lambda)\theta} - \frac{\lambda b}{\mu-\lambda} + \lambda(a+b\theta) \ (11)$$

$$\frac{df(\mu)}{d\mu} = \sum_{i=1}^{n}\frac{(\mu-\lambda)\theta+1}{(\mu-\lambda)^2}[\frac{\lambda b - h(\mu-\lambda)^2}{(\mu-\lambda)\theta+1}) - \lambda(b-c)e^{-(\mu-\lambda)\theta}] \ (12)$$

We firstly consider the situation that there is only one node in the system, then

1. When b=c, we can get that

$$\mu^* = \lambda + \sqrt{\frac{\lambda b}{h}} \ (13)$$

Thus, we obtain the average value of $\overline{\mu^*}$ to make the system to achieve the optimal performance.

$$\overline{\mu^*} = \frac{1}{n}(\sum_{i=1}^{n}\lambda_i + \sqrt{\frac{\lambda_i b}{h}}) \ (14)$$
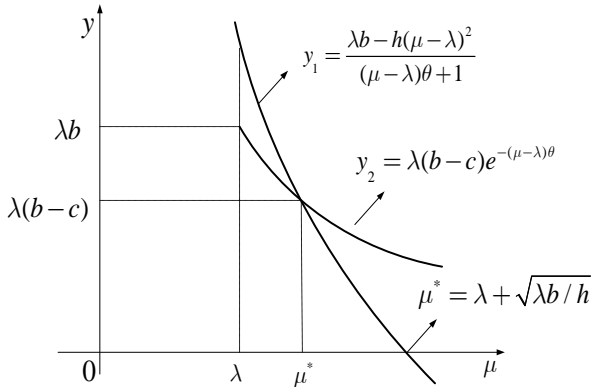
2. When $b > c$, we can get that

Figure 3. How to obtain the objective μ^* when $b > c$
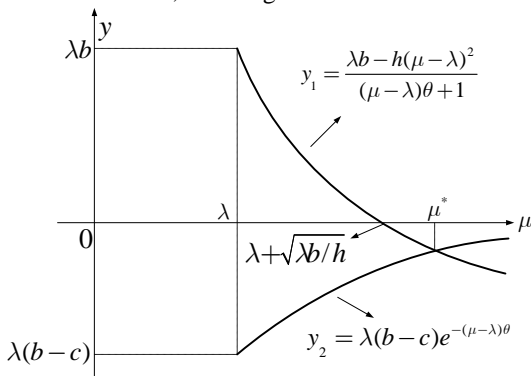
3. When $b < c$, we can get that

Figure 4. How to obtain the objective μ^* when $b < c$

B. Space Partition Method (SP)

So far, we have discussed the loss rate caused by NF and QF, optimization problem of performance under the heterogeneous p2p system, and then we discuss how to maximize performance of system under homogeneous p2p system. Homogeneous p2p system, namely each node of system has the same attributes such as capacity, process power, bandwidth etc,.

The main idea of space partition method is inspired by "virtual server method" [6][7]. Like the virtual server method, there are two spaces, one is the logical space of Chord, another space called partition space which consists of multiple nodes, and we say these nodes as agent nodes. Note that, in virtual server method, each node may hosting multiple virtual servers, however in space partition method, each node only match to one agent node, thus the number of agent nodes is equal to the number of system nodes.

We give the relationship between Chord space and partition space in the figure 2. As shown in the figure, Chord node N1 is matched to agent node agent_x, it means that the segment managed by agent_x is taken over by N1. From the figure, in addition to the segment {a}, N1 should manages the segment {b}, the reason is that segment {a} and {b} is managed by agent_x in the partition space.

The most important advantage of space partition method is that the partition space can be divided into arbitrary number of blocks because of the segment managed by each agent node consists of (key, pointers), here the pointer is address of the node which maintains the corresponding (key, values). Therefore, arbitrary partition of the space do not occur any update events.

As previously mentioned, the size of each node's zone is much different, thus in the homogeneous p2p system with each node has the same process power and bandwidth, in order to obtain the optimal performance, it should balance the query traffic load over the nodes.

R Cuevas [8] indicates that with the uniform (random) distribution, the query traffic each node received is proportional to the size of the segment the node managed.
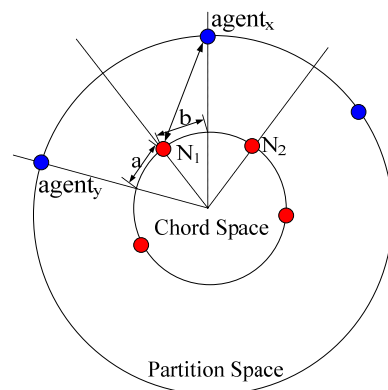

Figure 5. Relationship between Chord space and partition space

Therefore, we should divide the partition space into N equal parts.

We assume that, the size of the segment managed by node i is p_i, we use $\sigma_i = \dfrac{p_i}{\sum_{i=1}^{n} p_i}$ to represent the proportion of ID space node i owns. Note that, in our work, $\sigma_i = \dfrac{1}{n}$;

In order to solve the optimization problem, we use the function:

$$F_N = \frac{|\sum_{i=1}^{n} \xi_i|^2}{n \sum_{i=1}^{n} \zeta_i^2} \quad (15)$$

to evaluate whether each node process the equal number of messages, here ξ_i is the number of query messages node i received. From the function, we can get that, the function gives a value between 0 and 1, where 0 presents that some nodes may receive a great amount of messages but some node may receive little messages, this situation makes the loss rate of system much higher. When the value is 1, that means each node received the same number of messages which makes the loss rate of system is minimal.

The simulation results is given in the table 1, as shown in the figure, we consider 5 topologies of Chord with the number of nodes are 1000, 2000, 3000, 4000 and 5000 respectively. The entries of second column give the value of F_N with the standard Chord. The entries of third column give the value of F_N using "SP" method.

TABLE I.

Nodes	Chord	Chord (SP)
1000	0.623	0.897
2000	0.614	0.889
3000	0.602	0.846
4000	0.589	0.893
5000	0.578	0.884

IV. CONCLUSION

In the research field of reliability of structured p2p system, many works only consider the situation that the loss rate is caused by next hop node failure. As the matter of fact, the lost of queuing query messages on the failed node should be contribute to the real loss rate as well. Hence, in order to fill this gap and deeply investigate the reliability of structured p2p systems, this paper proposes an analytical model to exactly estimate the loss rate which considers the both situation. Based on this model, we give the optimization problem of system performance under heterogeneous and homogeneous p2p system respectively.

REFERENCES

[1] I. Stoica, R. Morris, D. Karger, M. F. Kaashoek, and H. Balakrishnan, "Chord: A scalable peer-to-peer lookup protocol for internet applications," in IEEE/ACM Transact- ions on Networking, vol. 11, no. 1, pp. 17–32, 2003.

[2] G Kunzmann, R Nagel, J Eberspacher, "Increasing the reliability of structured P2P networks," In proceeding of design of reliable communication networks, 2005.

[3] S Lam, H Y Liu, "Failure recovery of structured p2p networks: protocol design and performance under churn," in Computer Networks,vol 20, no. 16, pp 3084-3104.

[4] A Binzenhofer, D Staehle, R Henjes, "On the stability of chord-based P2P systems," in IEEE Global Telecommunications Conference, 2005.

[5] Z M Yao, D Loguinov, "Understanding Disconnection and Stabilization of Chord," in IEEE Infocom 2008, pp. 1049-1057.

[6] D R Karger, M Ruhl, "Simple Efficient Load Balancing Algorithm for Peer-to-Peer Systems," in ACM Symposium on Parallelism in Algorithms and Architectures (SPAA), 2004.

[7] D R Karger, M Ruhl, "New Algorithm for Load Balancing in Peer-to-Peer Systems," in the third International Workshop on Peer to Peer Systems (IPTPS), 2004

[8] R Cuevas, M Uruena, A Banchs, "Routing fairness in Chord: Analysis and Enhancement," in IEEE Infocom 2009.