Sensors **2014**, *14*, 16617-16629; doi:10.3390/s140916617

OPEN ACCESS

sensors

ISSN 1424-8220
www.mdpi.com/journal/sensors

Article

An Adaptive Failure Detector Based on Quality of Service in Peer-to-Peer Networks

Jian Dong *, Xiao Ren, Decheng Zuo and Hongwei Liu

School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China; E-Mails: renxiao@hit.edu.cn (X.R.); zuodc@hit.edu.cn (D.Z.); liuhw@hit.edu.cn (H.L.)

* Author to whom correspondence should be addressed; E-Mail: dan@hit.edu.cn; Tel./Fax: +86-451-8641-3754.

Received: 1 July 2014; in revised form: 4 August 2014 / Accepted: 5 August 2014 / Published: 5 September 2014

Abstract: The failure detector is one of the fundamental components that maintain high availability of Peer-to-Peer (P2P) networks. Under different network conditions, the adaptive failure detector based on quality of service (QoS) can achieve the detection time and accuracy required by upper applications with lower detection overhead. In P2P systems, complexity of network and high churn lead to high message loss rate. To reduce the impact on detection accuracy, baseline detection strategy based on retransmission mechanism has been employed widely in many P2P applications; however, Chen's classic adaptive model cannot describe this kind of detection strategy. In order to provide an efficient service of failure detection in P2P systems, this paper establishes a novel QoS evaluation model for the baseline detection strategy. The relationship between the detection period and the QoS is discussed and on this basis, an adaptive failure detector (B-AFD) is proposed, which can meet the quantitative QoS metrics under changing network environment. Meanwhile, it is observed from the experimental analysis that B-AFD achieves better detection accuracy and time with lower detection overhead compared to the traditional baseline strategy and the adaptive detectors based on Chen's model. Moreover, B-AFD has better adaptability to P2P network.

Keywords: distributed system; Peer-to-Peer networks; adaptive failure detector; quality of service

1. Introduction

Recently, Peer-to-Peer (P2P) network has rapidly become the major computing platform to share resources and services on Internet [1,2]. It can greatly improve network efficiency by taking full advantage of network bandwidth and the computing power of individual network nodes. An important guarantee to maintain the efficiency of P2P network is the high availability in the presence of node failures, as a basic building block for a distributed system of high availability [3], failure detector has always been a hot research topic in the field of P2P networks.

In P2P networks, failure detector provides support for routing recovery and update when failure occurs to maintain the validity of system topology by periodically probing the states of nodes in the system. Quality of Service (QoS) provided by failure detector is an important factor that affects the performance of P2P system, poor accuracy of detection will lead to plenty of unnecessary routing repair and data transfer, which will greatly increase the system overhead, especially in structured P2P networks. The reliability of routing imposes higher requirements on detection time, and the loss of normal packets (sent to failed nodes) has a great impact on the performance of upper applications such as task completion time, network throughput, video frame loss rate, *etc.* [4]. Meanwhile, the large scale and high churn of P2P networks [5,6] generate heavy detection overhead, which have become the major source of P2P traffic [7]. In the early version of Gnutella, the number of PING messages used for failure detection has exceeded 50% of the entire traffic [8], so far, P2P systems have occupied more than 60% of Internet traffic in China. We can see that detection time, accuracy and overhead all have significant effects on the improvement of P2P system performance. On the other hand, P2P systems may contain participants from any corner of the Internet, different capabilities of processing and network accessing generate complex network environment [9]. For example, comparing a task completed in open network with the one that is completed on a company's internal network, huge difference exists on their network conditions. Therefore, the study of adaptive failure detectors that can ensure the QoS of failure detection with lower overhead according to the changes in network conditions is very significant for building P2P applications.

Most of the current study on QoS-based adaptive failure detector is based on the classic adaptive model proposed by Chen [10], a series of adaptive detectors have been proposed to achieve the quantitative QoS metrics on the dynamic adjustment of detector parameters under different network conditions [10–12]. However, P2P applications covers the entire Internet, complex network conditions and the rapid joining and leaving by large number of nodes make the message loss rate very high, which may easily lead to an increase in false positive rate [13]. Therefore, baseline detection strategy based on retransmission mechanism [4,14] has been widely employed in P2P networks to reduce the impact of message loss on detection results and improve detection accuracy. However, the QoS of baseline strategy cannot be evaluated by Chen's classic model. To address the problem, this paper establishes a novel QoS evaluation model for the baseline detection strategy. The relationship between the detection period and the QoS is discussed and on this basis, a QoS-based adaptive failure detector (B-AFD) is proposed. Under the quantitative control by QoS basic metric (T_D^U, T_{MR}^L, T_M^U) [10], B-AFD can adapt to the changing network environment and achieves better accuracy and detection time with lower detection loads. Meanwhile, experimental analysis shows that B-AFD has better adaptability to P2P networks as compared to the traditional baseline strategy and the adaptive detector based on Chen's model.

2. Related Work

2.1. QoS Metrics for Failure Detector

It is well known that in the asynchronous distributed systems with crashed nodes, many important and fundamental problems (e.g., consensus) cannot be solved due to the fact that crashed nodes cannot be distinguished correctly from the nodes that response slowly [15]. Failure detector was first proposed and formally specified by Chandra [16] as an effective way to enhance the computational model for asynchronous system. Currently it has been applied widely in many related fields such as grid computing, cluster management, P2P networks, *etc.* [17,18]. To evaluate the ability of failure detector to solve consensus problems, Chandra categorized the detection capabilities according to completeness and accuracy. However, in actual systems, an application is much concerned about how fast a failure detector detects crashes and how well it avoids false detections, that is, the detection time and accuracy of failure detector. To evaluate these attributes accurately and quantitatively, Chen *et al.* proposed a set of quantitative QoS metrics [10,12]. Let T denote the output of failure detector when the monitored node is normal and S denotes the output when the node is failed. T-transition occurs when the output changes from S to T, while S-transition occurs when the output changes from T to S. The QoS provided by failure detector can be quantitatively described by the following basic metrics [10].

Detection time (T_D): T_D measures the time that elapses from the moment when a node crashes to the time when it starts being suspected (S-transition).

Mistake recurrence time (T_{MR}): T_{MR} measures the time between two consecutive mistakes, it is a random variable representing the time that elapses from an S-transition to the next one.

Mistake duration (T_M): T_M measures the time it takes the failure detector to correct a mistake, it is a random variable representing the time that elapses from an S-transition to the next T-transition. T_{MR} and T_M can specify the detection accuracy.

Based on these three primary metrics, other QoS metrics can be uniquely derived, such as average mistake rate (λ_M) and query accuracy probability (P_A), *etc*.

2.2. Adaptive Failure Detector

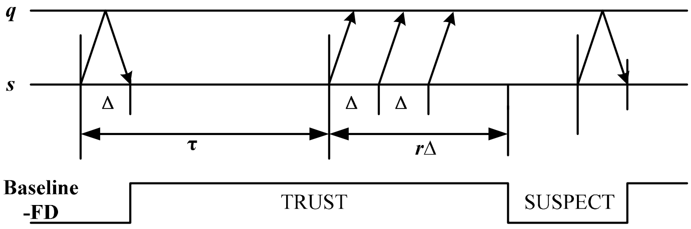
QoS requirement is the main basis for the design of failure detector. Most of failure detectors presented in the literature are implemented using the timeout-based mechanism [19], in which the probing messages are sent out periodically to detect the states of other nodes. Under this mechanism, a detector's behavior can be determined by the failure detection period η and the timeout value δ. As the network conditions (e.g., packets loss rate, transmission delay, *etc.*) are changing constantly, the QoS of failure detection cannot be guaranteed to meet the design requirements all the time. Therefore, adaptive failure detection algorithms have been proposed to adapt to the changing network conditions by adjusting the parameters η and δ automatically. Initially, such adjustments were achieved by modifying δ according to the prediction of the arrival time for detection messages and a simple trade-off was made between detection time and accuracy. Falai [20] presented the performance comparison of various commonly used prediction methods such as LAST, MEAN, linear sequence *etc.* After proposing the QoS metrics, Chen [9] presented a classic adaptive model based on the network probability model, and on the basis, many adaptive failure detectors [10–12] are proposed to achieve the quantitative QoS metrics required by upper application with lower detection overhead under different network conditions. However, to reduce the impact on detection accuracy by the high message loss rate resulted from complexity of network and high churn, baseline detection strategy based on retransmission mechanism has been employed widely in P2P systems [4,7,14]. These changes make Chen's classic adaptive model no longer applicable to P2P systems. In order to provide the service of QoS-based adaptive failure detection in P2P systems, this paper establishes a novel QoS evaluation model for the baseline detection strategy that is based on retransmission mechanism and PULL style. Furthermore, a QoS-based adaptive failure detector B-AFD is proposed for P2P networks.

3. QoS Evaluation Model

3.1. Baseline Failure Detection Strategy

We consider an asynchronous distributed systems which consists of a set of n nodes. The same failure model is employed here as crashing used by Chandra [16]. The channels between nodes are fair-lossy link [21], that is, only a finite number of messages are allowed to be lost, assume the message loss probability is p_l. To simply the description, we consider a failure detector module in node s, which detects the state of node q. PULL style [22] is used in the baseline detection strategy and retransmission mechanism is designed in each detection period as shown in Figure 1.

Figure 1. Baseline detection strategy.



In Figure 1, failure detector module of s sends a probing message "are you alive" to the monitored node q in each detection period (denoted as τ). Node q will send an "ack" message back to acknowledge the receipt of the probing message. If s receives an acknowledge message from q in interval Δ, then the detector output is T. Otherwise, probing packets are sent out periodically at interval Δ. If r consecutive packets cannot receive the "ack" message, then the output of detector is S.

3.2. QoS Analysis

The analysis of QoS evaluation metrics is an important basis for the study of QoS-based adaptive failure detector. Theorem 1 gives the QoS calculation model based on the baseline detection strategy shown in Figure 1.

Theorem 1. *In the baseline detection strategy shown in Figure 1, let p_l denote the message loss probability, $P(D < x)$ denote the probability distribution of detection message delay (D) and s denote the number of bytes in detection message. We have:*

(1) *Average mistake recurrence time* $E(T_{MR}) = \dfrac{\tau}{p^r(1-p^r)}$

(2) *Average mistake duration* $E(T_M) = \dfrac{\tau}{1-p^r} - \dfrac{r\Delta}{1-p^r} + \dfrac{\Delta}{1-p}$

(3) *Detection time* $T_D \le \tau + r\Delta$

(4) *Average detection overhead* $E(B) = \dfrac{1-p^r}{1-p}\cdot\dfrac{s}{\tau}$ *where* $p = p_l + (1-p_l)P(D > \Delta)$.

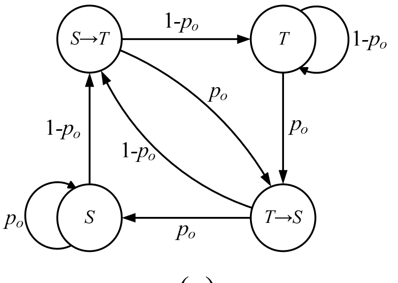
The detection message delay is the time that elapses from the sending of probing message to the receipt of "ack" message. We suppose the monitored node q always keeps accurate. For any detection message mq_{ij}*(* $i > 0, r \ge j > 0$ *), the probability that no "ack" is received at time* $(i\tau + j\Delta)$ *is* $p_l + (1-p_l)P(D > \Delta)$*, that is, the value of p in Theorem 1. To complete the proof of Theorem 1, two lemmas should be given first.*

Lemma 1. $E(T_{MR}) = \dfrac{\tau}{p^r(1-p^r)}$, *where* $p = p_l + (1-p_l)P(D > \Delta)$.

Let $\{X_i, i > 0\}$ *be a random sequence, where random variable* X_i *represents the output state of the detector at time* $(i\tau + r\Delta)$*. For the baseline strategy in Figure 1, the state space can be defined as* $G = (T \rightarrow S, S \rightarrow T, S, T)$*, e.g., S-transition, T-transition, maintaining S state and maintaining T state. The state transition diagram and the n-step transition probability matrix* $M^{(n)}$ *(n > 2) for* $\{X_i, i > 0\}$ *are shown in Figure 2, where* p_0 *is the probability that none of the detection messages is received during a detection period. Thus, we have* $p_0 = p^r$.

From the Figure 2, we can see that the value of variable $X_i (i > 0)$ *is determined jointly by* $\{mq_{ij}, 0 \le j \le r-1\}$ *and the value of* X_{i-1}*. Moreover, it is irrelevant to* $X_j (0 < j < i-1)$*, that is,* $P(X_i \mid X_{i-1}) = P(X_i \mid X_{i-1}, X_{i-2}, ..., X_0)$*. Hence, the random sequence* $\{X_i, i > 0\}$ *is a finite state Markov chain. Since* $0 < p_0 < 1$*, we can get that* $\{X_i, i > 0\}$ *is a recurrent Markov chain and the state* $T \rightarrow S$ *is ergodic from the transition matrix* $M^{(n)}$*. For state* $T \rightarrow S$ *, there exists* $\lim_{n\to\infty} M^{(n)}_{T\to S, T\to S} = p_o(1-p_o) > 0$*, the average recurrence time is* $1/p_0(1-p_0)$*. Consider the time needed for an occurrence of state transition in a detector system as* τ *, we have* $E(T_{MR}) = \tau / p^r(1-p^r)$.

Figure 2. (**a**) The state transition diagram; (**b**) n-step transition probability matrix (n > 2).



(**a**)

$$M^{(n)} = \begin{pmatrix} p_o(1-p_o) & (1-p_o)^2 & p_o^2 & p_o(1-p_o) \\ p_o(1-p_o) & (1-p_o)^2 & p_o^2 & p_o(1-p_o) \\ p_o(1-p_o) & (1-p_o)^2 & p_o^2 & p_o(1-p_o) \\ p_o(1-p_o) & (1-p_o)^2 & p_o^2 & p_o(1-p_o) \end{pmatrix}$$

(**b**)

Lemma 2. *In the baseline detection strategy shown in Figure 1, the query accuracy probability is* $P_A = (1-p^r) + \dfrac{r\Delta}{\tau}p^r - \dfrac{\Delta}{\tau}p^r\dfrac{1-p^r}{1-p}$, $p = p_l + (1-p_l)P(D > \Delta)$.

Proof of Lemma 2. Query accuracy probability P_A is the probability that the failure detector's output is *T* at a random time *t*. That is, the probability that users get an accurate output when query the failure detector at any time.

Consider any detection period *i* (*i* > 0) in baseline strategy, suppose Y_i is a random variable representing the time that the detector output is *T* during the period $[i\tau,(i+1)\tau]$. Let's discuss Y_i under two situations:

The final output is T during the detection period $[(i-1)\tau,i\tau)$

In this case, in detection period $[i\tau,(i+1)\tau)$, if at least one of the *r* detection messages have been acknowledged in interval Δ, then the output will keep *T* during the entire period τ. Otherwise, *S*-transition will occur at time $(i\tau+r\Delta)$. Hence, we have $E(Y_i)_T = r\Delta\cdot p^r + \tau\cdot(1-p^r)$.

The final output is S during the detection period $[(i-1)\tau,i\tau)$

If the detector's state is *S* in detection period $[(i-1)\tau,i\tau)$, then the moment *T*-transition occurs in period *i* depends on the time when the first detection message is acknowledged. Thus, we have

$$\begin{aligned} E(Y_i)_S &= \sum_{i=0}^{r-1}(1-p)p^i(\tau-(i+1)\Delta) \\ &= \tau(1-p^r)-\Delta\frac{1-p^r}{1-p}+r\Delta p^r \end{aligned}$$

In summary, the time that the detector's output is *T* during the period $[i\tau,(i+1)\tau)$ is

$$\begin{aligned} E(Y_i) &= (1-p^r)E(Y_i)_T + p^r E(Y_i)_S \\ &= \tau(1-p^r)+r\Delta p^r - \frac{1-p^r}{1-p}p^r\Delta \end{aligned}$$

In the above equation, the value of $E(Y_i)$ is irrelevant to the detection period *i*. Hence, the query accuracy probability is

$$P_A = (1-p^r)+\frac{r\Delta}{\tau}p^r - \frac{\Delta}{\tau}p^r\frac{1-p^r}{1-p}$$

Now we prove Theorem 1 using Lemma 1 and Lemma 2.

Proof of Theorem 1.

(1) It has been proved by Lemma 1.

(2) According to Chen's QoS metrics, we have $P_A = 1-E(T_M)/E(T_{MR})$. Then we get $E(T_M)=(1-P_A)E(T_{MR})$. According to Lemma 2, we have

$$E(T_M)=\frac{\tau}{1-p^r}-\frac{r\Delta}{1-p^r}+\frac{\Delta}{1-p}$$

(3) In the detection period *i*, if node *q* crashes during $[i\tau,(i+1)\tau)$, then the failure will be detected no later than $((i+1)\tau+r\Delta)$. Thus, the failure detection time satisfies $T_D \le \tau + r\Delta$.

(4) In every detection period of baseline strategy, at most *r* messages are sent out and no more messages will be sent after the first acknowledgement message is received. In detection period *i*, suppose there is a random variable Z_{ij},

$$Z_{ij} = \begin{cases} 1 & \textit{if } mq_{ij} \textit{ has been acknowledged by time } (i\tau + j\Delta) \\ 0 & \textit{else} \end{cases}, \ i > 0, r \geq j > 0$$

We have $P(Z_{ij} = 1) = 1 - p$ and Z_{ij} satisfies the Bernoulli distribution with success probability as $1 - p$. Hence, the number of messages N_m sent within a detection period satisfies the geometric distribution that is

$$E(N_m) = \sum_{j=1}^{r} p^{j-1}(1-p) \cdot j + rp^r$$

Now we can get the average detection loads of the algorithm as $E(B) = \frac{1-p^r}{1-p} \cdot \frac{s}{\tau}$.

In summary, Theorem 1 is proved.

4. QoS-Based Adaptive Failure Detector B-AFD

(T_D^U, T_{MR}^L, T_M^U) is used to describe the quantitative requirements of system designers on the detection accuracy and time of failure detectors, where T_D^U is an upper bound on the detection time, T_{MR}^L is a lower bound on the average mistake recurrence time, and T_M^U is an upper bound on the average mistake duration. B-AFD is based on baseline failure detection strategy, where parameters r and τ are adjusted automatically to adapt to the dynamic network environment in P2P system so that the requirements of QoS metrics (T_D^U, T_{MR}^L, T_M^U) are met with relatively low detection overhead. The B-AFD failure detector is described in Alogorithm 1.

Alogrithm 1: B-AFD failure detector

```
For node $s$:
detector_module:
    at time $\tau_i$: (the $i$th detection period)
        $(r_b, \tau_b)$ =get_opti_para( );
        if ( $(r_b = 0) \wedge (\tau_b = 0)$ ) then
            exit ("QoS cannot be achieved");
        $\tau_{i+1} = \tau_i + \tau_b$ ; $j \leftarrow 0$;
        do {
            send $mq_{ij}$ to node $q$ at time $\tau_i + j\Delta$;
            if ( receive $ma_{ij}$ before $\tau_i + (j+1)\Delta$ ) then
                result $\leftarrow$ $T$; break;
            else $j$++;
        }while( $j < r_b$ );
        if ( $j \geq r_b$ ) then result $\leftarrow$ $S$;
gatherer_module:
    upon receive $ma_{ij}$ from $q$ do
        add($t_{current}$- $(\tau_i + j\Delta)$ ) to $W_D$;

For node $q$:
    upon receive $mq_{ij}$ form $s$ do
        send $ma_{ij}$ to $s$;
```

B-AFD failure detector is composed of detector_module and gatherer_module. In detector_module, at most r_b probing messages mq_{ij} are sent out during each detection period. If none of the messages can receive the corresponding "ack" message ma_{ij} successfully, the detector outputs *S* for node *q*; otherwise, the detector believes that *q* maintains accurate state. From Alogorithm 1, we can see that function get opti para is the core mechanism in B-AFD for the automatic adaptation to network states, which is called to compute the parameters *r* and τ at the beginning of each detection period. As shown in Alogorithm 2, the algorithm get opti para uses (T_D^U, T_{MR}^L, T_M^U) as input parameters so that the detector can be configured to meet the QoS required by upper applications with the minimum overhead. The gatherer_module in B-AFD creates a sliding window W_D with fixed size (*w*). The detection message delay of the *w* probing messages that are recently acknowledged is saved to calculate the message loss probability p_l and establish samples for the estimation of the probability distribution of detection message delay $P(D < x)$.

Alogorithm 2: Adaptive configuration of parameters

```
get_opti_para( )
Input: $T_D^U, T_{MR}^L, T_M^U$
Output: $r, \tau$
Step 1: compute $p = p_l + (1 - p_l)P(D > \Delta)$;
Step 2: if $T_M^U < \Delta/(1 - p)$ then return (0,0);
Step 3: compute $r_{\max} = \lfloor T_D^U / 2\Delta \rfloor$, $R = [1, r_{\max}]$;
        compute $R_L = \{r \mid L(r) \ge 0, r \in R\}$ and $R_U = \{r \mid U(r) \ge 0, r \in R\}$;
        $L(r) = (1 - p^r)(T_M^U - \frac{\Delta}{1 - p} - T_{MR}^L p^r) + r\Delta$
        $U(r) = T_D^U - T_{MR}^L p^r (1 - p^r) - r\Delta$
Step 4: if $R_L \cap R_U == \varnothing$ then return (0,0);
Step 5: Compute $Output = \{(r, \tau) \mid r \in R_L \cap R_U \text{ and } \tau \text{ satisfies } (1)\}$;
Step 6: find $(r_b, \tau_b) \in Output$ such that
    $B(r_b, \tau_b) = \min\{B(r, \tau) \mid (r, \tau) \in Output\}$;
      $B(r, \tau) = \frac{1 - p^r}{1 - p} \cdot \frac{s}{\tau}$
Step 7: return $(r_b, \tau_b)$
```

Theorem 2. *If the return value of the algorithm get_opti_para in* Alogorithm 2 *satisfies* $(r > 0) \wedge (\tau > 0)$, *then the outputs of the failure detector B-AFD can meet the QoS requirements* (T_D^U, T_{MR}^L, T_M^U), *else B-AFD cannot achieve the given QoS requirements.*

Proof of Lemma 2. According to the results in Theorem 1, there are $E(T_{MR}) \ge T_{MR}^L$, $E(T_M) \le T_M^U$ and $T_D \le T_D^U$. The parameters *r* and τ satisfy the constraint relationship in Equation (1).

$$\begin{cases} \tau \le T_M^U (1 - p^r) + r\Delta - \dfrac{1 - p^r}{1 - p}\Delta \\ \tau \ge T_{MR}^L p^r (1 - p^r) \\ \tau \le T_D^U - r\Delta \\ \tau \ge r\Delta \end{cases} \tag{1}$$

According to Equation (1), if τ exists, then it need satisfy

$$\begin{cases} (1-p^r)(T_M^U - T_{MR}^L p^r - \Delta/(1-p)) + r\Delta \geq 0 & (a) \\ T_D^U - T_{MR}^L p^r (1-p^r) - r\Delta \geq 0 & (b) \\ T_M^U - \Delta/(1-p) \geq 0 & (c) \\ T_D^U - 2r\Delta \geq 0 & (d) \end{cases} \tag{2}$$

From Equation (2-*c*), we get $T_M^U < \Delta/(1-p)$ (step2), otherwise the detector is unable to meet the given QoS requirements. From Equation (2-*d*) we have $r \leq T_D^U/2\Delta$, then the range of *r* satisfies $1 \leq r \leq T_D^U/2\Delta$. Meanwhile, it can be proved that on interval $[1, T_D^U/2\Delta]$, *L*(*r*) is monotonically increasing and *U*(*r*)is monotonically decreasing. According to the constraints Equation (2-*a*,2-*b*), we can get the upper bound r_U and the lower bound r_L for parameter *r* respectively. If $r_L \leq r_U$, that is, the condition $R_L \cap R_U \neq \varnothing$ is satisfied in step 4, then *r* exists; otherwise the detector cannot meet the given QoS requirements under current network conditions. In step 5, using Equation (1), the collection *Output* contains all the combinations of values for (r, τ) which are capable of meeting the QoS requirements. By the screening procedure in step 6, we can get the parameter configuration (r_b, τ_b) which achieves the minimum detection overhead. Therefore, if the return value of get opti para is non-zero, then the outputs of the failure detector B-AFD configured according to that return value will be able to meet the QoS requirements (T_D^U, T_{MR}^L, T_M^U). Theorem 2 is proved.

5. Experimental Results and Analysis

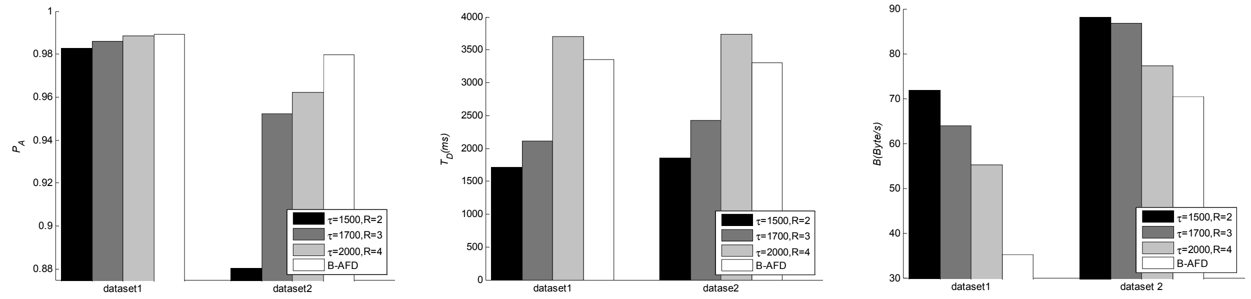
To evaluate the performance of B-AFD, we compare B-AFD with two typical failure detectors. One is traditional baseline detection strategy with fixed parameters (shown in Figure 1), which has been used the most commonly in current P2P environment. This experiment is to verify the improvement on detection accuracy and overhead by the adaptive mechanism in B-AFD. The other one is QoS-based adaptive detector NFD-E [10], which is compared to verify B-AFD's ability of adapting to the complex P2P network environment. To ensure the authenticity of the experiments, some nodes in currently prevalent P2P applications (emule and Bittorrent) are selected as detection objects, the failure detector node located in Harbin City (China).

In P2P systems, nodes may come from any corner of the Internet with huge difference in network conditions, thereby, the experiments are carried out on two sets of nodes which represent two typical network conditions in P2P networks. One group (dataset 1) contains monitored nodes located in China, which have good network connections with detector node, the message delay and message loss probability are low (p_l = 0.39%, *E*(*D*) = 125 ms, where *E*(*D*) is the expectation of detection message delay). The other group (dataset 2) mainly consists of monitored nodes located in the United States, which have relatively poor connections with detector node (in China), in which the message transmission delay and loss rates are high (p_l = 3.65%, *E*(*D*) = 412 ms). We choose exponential distribution for detection message delay D with reference to Lakshman's research [23] about failure detector in a P2P storage system, *i.e.*, $P(D \leq x) = 1 - e^{-x/E(D)}$, for all *x* > 0.

5.1. Comparisons with the Traditional Baseline Strategy

The comparison with baseline strategy is evaluated in three aspects: Detection accuracy, detection time and overhead. Since baseline uses fixed parameters, for fairness, three sets of different parameters are selected for baseline under different network environment. The comparison results are shown in Figure 3.

Figure 3. Comparison with baseline strategy. (**a**) detection accuracy; (**b**) detection time; (**c**) overhead.



(**a**) (**b**) (**c**)

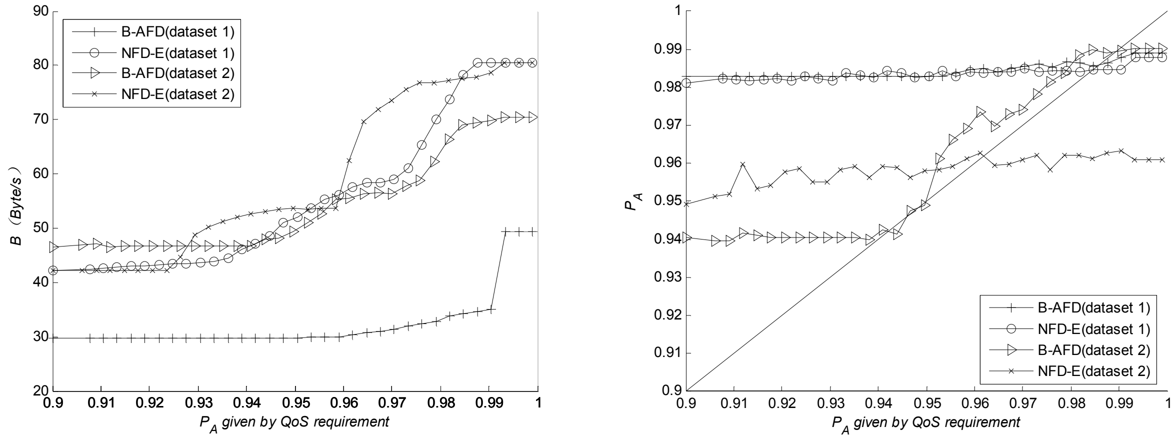
As we can see from Figure 3, under good network conditions (dataset 1), all the four sets of experiments have achieved high accuracy. Smaller detection period leads to lower detection time, but with higher detection overhead. In the experiments on dataset 2, where the distance of the nodes is far and network condition is poor, the detection accuracy drops significantly for the baseline with smaller detection period. The changes in network conditions make huge difference on QoS, therefore, baseline strategy with fixed period is not applicable to the kind of large-scale distributed systems, such as P2P. Experimental results have shown that B-AFD detectors have good adaptability to different network environment due to the detection parameters adjusted adaptively according to network changing, there is little change in the accuracy and detection time under different network environment and the specified QoS metrics are still satisfied while minimum overhead are maintained. Moreover, Figure 3-c shows that the decrease in overhead is more obvious when the network condition is better.

5.2. Comparisons with NFD-E Adaptive Detector

NFD-E is a classical adaptive failure detector proposed by Chen [8], which can meet the requirements of quantitative QoS metrics under different network environment as well without retransmission mechanism. Therefore, experiments are carried out to compare B-AFD and NFD-E in two aspects: The overhead needed to meet the same QoS metrics and the capability of adapting to complex network. The results are shown in Figure 4.

To obtain the same detection accuracy, Figure 4a shows that the detection overhead generated by B-AFD is significantly lower than NFD-E, especially when the requirement of detection accuracy is higher. As can be seen from Figure 4b, B-AFD demonstrates better adaptability under poor network conditions. Given the same QoS metrics, from the experimental results of dataset 2, NFD-E is no longer meet the requirement when the query accuracy exceeds 96%. However, nearly 99% of the query accuracy requirements are met by B-AFD under the same network environment. It is clear that the retransmission mechanism can significantly improve the accuracy under poor network conditions without adding overhead. Under good network conditions, B-AFD demonstrates a very close adaptability to NFD-E while keeping obvious advantage in detection overhead as shown in Figure 4. Therefore, B-AFD is more appropriate for the kind of large-scale distributed systems, such as P2P that have wide coverage and complex network conditions, especially for the structured P2P systems whose node routing tables may contain the nodes from internal LAN and the nodes from overseas at the same time.

Figure 4. Comparisons with NFD-E. (**a**) Detection overhead (**b**) adaptability.



(**a**) (**b**)

6. Conclusions

P2P networks have become the major source of Internet traffic. The study of adaptive failure detector with low detection overhead provides an important method to reduce the overhead of P2P networks. To address the challenges posed to failure detection by the complexity of network and high churn in P2P system, this paper establishes a novel QoS evaluation model for the baseline detection strategy based on retransmission mechanism. On this basis, an adaptive failure detector B-AFD is proposed based on the basic QoS metrics (T_D^U, T_{MR}^L, T_M^U). It can adapt to the changing network environment and achieve better detection accuracy and time with lower overhead. Meanwhile, experimental analysis shows that compared to Chen's adaptive detectors, B-AFD achieves better adaptability to the complex network conditions in P2P systems.

Acknowledgments

This project was supported by National Natural Science Foundation of China under (No. 61100029).

Author Contributions

Jian Dong proposed the original idea, took the leadership on this research work. Xiao Ren contributed to implementation of failure detection and design of experiments. Decheng Zuo and Hongwei Liu contributed to data analyzing. All authors contributed to the discussion, to experiment improvements, to the analysis of results, and to the manuscript writing.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Eng, K.L.; Crowcroft, J.; Pias, M.; Sharma, R.; Lim, S. A survey and comparison of peer-to-peer overlay network schemes. *IEEE Commun. Surv. Tutor.* **2005**, *7*, 72–93.
2. Kurian, J.; Sarac, K. A survey on the design, applications, and enhancements of application-layer overlay networks. *Acm Comput. Surv.* **2012**, *43*, 1–34.
3. Xiong, N.; Vasilakos, A.V.; Wu, J.; Yang, Y.R.; Rindos, A.; Zhou, Y.Z.; Song, W.Z.; Pan, Y. A Self-tuning Failure Detection Scheme for Cloud Computing Service. In Proceedings of the 26th International Parallel and Distributed Processing Symposium, Shanghai, China, 21–25 May 2012; pp. 668–679.
4. Zhuang, S.Q.; Geels, D.; Stoica, I.; Katz, R.H. On failure detection algorithms in overlay networks. In Proceedings of the 24th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM 2005), Miami, FL, USA, 13–17 March 2005; pp. 2112–2123.
5. Ohzahata, S.; Kawashima, K. An experimental study of peer behavior in a pure P2P network. *J. Syst. Softw.* **2011**, *84*, 21–28.
6. Benhamida, F.Z.; Challal, Y.; Koudil, M. ALLONE: A New Adaptive Failure Detector Model for Low-Power Lossy Networks. In Proceedings of the IEEE Global Communications Conference, Atlanta, GA, USA, 9–13 December 2013; pp. 91–96.
7. Castro, M.; Costa, M.; Rowstron, A. Performance and dependability of structured peer-to-peer overlays. In Proceedings of the International Conference on Dependable Systems and Networks, Florence, Italy, 28 June–1 July 2004; pp. 9–18.
8. Dedinski, I.; Hofmann, A.; Sick, B. Cooperative keep-alives: An efficient outage detection algorithm for p2p overlay networks. In Proceedings of the Seventh IEEE International Conference on Peer-to-Peer Computing, Galway, Ireland, 2–5 September 2007; pp. 140–150.
9. Tian, J.; Dai, Y. Understanding the dynamic of peer-to-peer systems. In Proceedings of the Seventh IEEE International Conference on Peer-to-Peer Computing, Galway, Ireland, 2–5 September 2007; pp. 151–158.
10. Chen, W.; Toueg, S.; Aguilera, M.K. On the quality of service of failure detectors. *IEEE Trans. Comput.* **2000**, *51*, 13–32.
11. Bertier, M.; Marin, O.; Sens, P. Implementation and performance evaluation of an adaptable failure detector. In Proceedings of the International Conference on Dependable Systems and Networks, Washington, DC, USA, 23–26 June 2002; pp. 354–363.
12. Tiejun, M.; Hillston, J.; Anderson, S. On the quality of service of crash-recovery failure detectors. *IEEE Trans. Dependable Secur. Comput.* **2010**, *7*, 271–283.
13. Stutzbach, D.; Rejaie, R. Understanding churn in peer-to-peer networks. In Proceedings of the Internet Measurement Conference, Rio de Janeiro, Brazil, 25 October 2006; pp. 189–202.
14. Price, R.; Tino, P. Still alive: Extending keep-alive intervals in P2P overlay networks. In Proceedings of the 5th International Conference on Collaborative Computing: Networking, Applications and Worksharing, Washington, DC, USA, 11–14 November 2009; pp. 1–10.
15. Fischer, M.J.; Lynch, N.A.; Paterson, M.S. Impossibility of distributed consensus with one faulty process. *J. ACM* **1985**, *32*, 374–382.

16. Chandra, T.D.; Toueg, S. Unreliable failure detectors for reliable distributed systems. *J. ACM* **1996**, *43*, 225–267.
17. Lavinia, A.; Dobre, C.; Pop, F.; Cristea, V. A failure detection system for large scale distributed systems. In Proceedings of the 2010 International Conference on Complex, Intelligent and Software Intensive Systems (CISIS), Krakow, Poland, 15–18 February 2010; pp. 482–489.
18. Costache, S.; Ropars, T.; Morin, C. Towards highly available and self-healing grid services. Le Chesnay, France: Institut National des Sciences Appliquées de Rennes (INRIA), 2010. inria-00476276.
19. Pasin, M.; Fontaine, S.; Bouchenak, S. Failure detection in large scale systems: A survey. In Proceedings of the IEEE Network Operations and Management Symposium Workshops, Bahia, Brazil, 7–11 April 2008; pp. 165–168.
20. Falai, L.; Bondavalli, A. Experimental evaluation of the QoS of failure detectors on wide area network. In Proceedings of the International Conference on Dependable Systems and Networks, Yokohama, Japan, 28 June–1 July 2005; pp. 624–633.
21. Zhang, J.; Chen, W. Implementing uniform reliable broadcast with binary consensus in systems with fair-lossy links. *Inf. Process. Lett.* **2009**, *110*, 13–19.
22. Felber, P.; Defago, X.; Guerraoui, R.; Oser, P. Failure detectors as first class objects. In Proceedings of the International Symposium on Distributed Objects and Applications, Edinburgh, UK, 5–6 September 1999; pp. 132–141.
23. Lakshman, A.; Malik, P. Cassandra: A decentralized structured storage system. *ACM SIGOPS Oper. Syst. Rev.* **2010**, *44*, 35–40.