

一种面向容灾自适应的 PULL & PUSH 心跳检测机制

王浩铭 穆道生 高丽娟
(装备学院 北京 101416)

摘要 心跳检测技术作为容灾应急技术的重要环节,是整个容灾系统发挥作用的基础,文章研究的一种面向容灾自适应 PULL & PUSH 心跳检测机制,将两种心跳模式进行优势互补,并结合网络环境和应用 QoS 需求作为评判失效的依据,从而来提高检测机制的可靠性与实时性。

关键词 心跳检测, 自适应, QoS, PULL & PUSH

中图法分类号 TP306 **文献标识码** A

Adaptive PULL & PUSH Heartbeat Detection Machine-made for Disaster Recovery

WANG Hao-ming MU Dao-sheng GAO Li-juan
(The Equipment Academy, Beijing 101416, China)

Abstract As an important part of the capacity of disaster emergency technology, heartbeat detection technology is the foundation of the disaster recover system. This paper studies a adaptive PULL & PUSH heartbeat detection machine-made for disaster recovery, made the two heartbeat model complementary advantages, and combined the network environment with application QoS requirements as the basis of judging the failure, so as to improve the reliability and real-time performance of detection mechanism.

Keywords Heartbeat detection, Adaptive, QoS, PULL & PUSH

1 引言

心跳检测技术按照其实现方式的不同,分为 PUSH 和 PULL 两种模式^[1],二者都是通过节点之间周期性地发送检测信息来检测对方的状态。所不同的是,PUSH 模式是分节点主动向它的主节点周期性地发送 PUSH 心跳信息,而 PULL 模式则是由主节点向分节点发送查询信息,并等待分节点反馈消息来判断是否失效。PULL 方式与 PUSH 方式相比,得到同样的结果需要消耗两倍甚至更多的通信开销和时间周期,但 PULL 方式是一种主动检测方式,可以只在需要的时候才发起检测,更加符合复杂系统环境下失效检测的需求,拥有更好的针对性和高可靠性。

传统判断心跳检测的依据主要有 3 种:1)根据对系统心跳信号传输时间的估计与测量来确定一个大致通信失效的时间阈值,通过比较心跳信号的收到时间是否在阈值范围内来判断失效性。2)在心跳检测中采用一个算法输出一个布尔值,来直接判断失效与否。这两种方法直观,但是结果过于单一,没有考虑到复杂网络环境下系统的调整,也没有加入系统的动态分析。3)在过程中将检测与判断分离,输出一个变化量来代表此节点的信任程度,并根据具体网络环境判断失效与否。这种方法可定量地得出节点的有效性,但是在定性此节点是否失效问题上还没有得到很好解决。

因此本文提出的检测机制主要考虑可靠性和实时性两方面,在方式上融合 PULL 模式和 PUSH 模式,在评判依据上

结合上述 3 种方法进行改进,即加入定量 QoS^[2]评价体系确定阈值,并与检测结果的信任度值进行对比,从而衡量失效检测的有效性。

2 检测机制描述

本文提出的面向容灾自适应的 PULL & PUSH 心跳检测机制(见图 1)在执行过程中主要分为 3 个阶段,分别是 PUSH 检测阶段、疑似失效 PULL 检测阶段和失效确认阶段。

PUSH 检测阶段:由于 PUSH 检测过程较短的时间周期和少量的通信开销,因此在系统常规检测中从节点全部利用 PUSH 检测模式向主节点发送心跳信号;由于主从两节点具有时间统一性,因此主节点可以根据从节点发送心跳信息时间的延迟计算出各个节点的超时期望,当主节点超时时间达到超时期望的 Δt_i 时,即认为该点疑似失效并迅速启动 PULL 检测机制。

疑似失效 PULL 检测阶段:PULL 机制包括主、从节点上运行的两个检测进程 a 和 b ,用以检测对方运行状态。进程 a 周期性地向进程 b 发送带有标志的心跳信号,当后者接到前者的消息时,立即恢复带有相同标志的心跳信号。通过统计进程 a 和 b 传输信息的超时时间计算出对应的期望和方差,并根据实时超时量计算一个 ρ 值来代表被测节点可信度。

失效确认阶段:在此过程主要考虑输出 ρ 和失效阈值 P ,不同的输出值 ρ 表示了不同网络状态下被检测方的可用度,

王浩铭(1989—),男,硕士生,主要研究方向为容灾备份应急服务;穆道生(1965—),男,硕士,教授,主要研究方向为军事通信技术;高丽娟(1978—),女,博士,讲师,主要研究方向为空间通信技术。

而失效阈值 P 代表了用户对检测质量的要求,由用户给出的参数计算得出。 ρ 值和 P 值体现了对系统的自适应性^[3],二者的比较过程是被检测节点当前状态是否满足用户检测要求的判断过程,也是判断是否失效的依据。具体算法如下:

$$\rho = \begin{cases} 1 - \frac{\delta}{(t_a - t_r - \mu)^2 + \delta}, & \text{if } (t_a - t_r) > \mu \\ 0, & \text{else} \end{cases} \quad (1)$$

式中, t_a 代表信息到达时刻, t_r 代表最新收到的信息期望到达时间, μ 和 δ 为超时时间 D 的期望和方差。然后,需要计算代表检测准确性范围的阈值,根据文献[4]的阈值标准设计阈值为:

$$P = \text{MAX} \left\{ \frac{1 + \sqrt{1 - 4\Delta t_i / T_{MR}^U}}{2(1 - P_L)}, \frac{\Delta t_i}{T_M^U(1 - P_L)} \right\} \quad (2)$$

式中, P_L 为传输过程中消息的丢失率, Δt_i 为 a 向 b 发送信息的预定时间间隔, T_{MR}^U 为错误时间间隔下限, T_M^U 为错误时间间隔上限。 T_{MR}^U 、 T_M^U 和 Δt_i 均根据具体容灾系统的要求来确定值的大小。考虑到失效检测的及时性问题,设定参数 T_0 来表示超时时间上限值,一旦监测到的超时时间大于 T_0 ,则判定被监测对象失效。 T_0 的大小同样根据容灾系统的要求设定。由此,判定节点失效的阈值为 $(\rho > P) \vee (t_a - t_r > T_0)$ 。

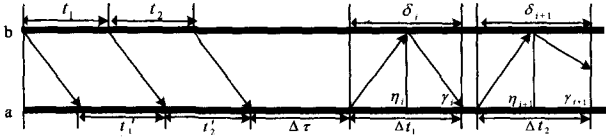


图1 PULL&PUSH 检测机制示意图

3 检测机制分析

(1) 通信开销与可靠性

假设两种检测方式单向传输信息的开销相同,即通信数据包占用总线通信时间相同,定义为 κ 。当采用 PULL 模式检测一次就需要总线控制器向总线上 n 个模块发送查询信息,收到后回应用户信息,双方交互一次需要有发送和反馈两个环节,所用通信开销为:

$$K_{PULL} = n\kappa \quad (3)$$

而 PUSH 检测模式只需单向发送信号,通信开销约为 PULL 模式的一半:

$$K_{PUSH} = \frac{1}{2}n\kappa \quad (4)$$

设每个模块检测失效概率为 p ,当发生失效时的概率 $P_f = 1 - (1 - p)^n$,如果没有收到 PUSH 心跳信号,则总线控制器会发起 PULL 检测。所以在一次完整的检测过程中, PULL 检测的通信开销为:

$$K_{\&} = K_{PUSH} + K_{PULL}P_f = \frac{1}{2}n\kappa + n\kappa(1 - (1 - p)^n) \quad (5)$$

因为 $0 < p < 1$,故 $K_{PULL} - K_{\&} > 0$,从而 PUSH 结合 PULL 的模式比只使用 PULL 的模式降低了通信开销。

然而单纯使用 PUSH 检测机制要求当发生主节点超过一定时间收到从节点 PUSH 信号时,即确认此节点失效,使得判断依据过于简单,极易产生误判。实际上,由于网络状况的变化和系统延迟往往会造成超时现象,但实际节点并没有失效,因此单纯从 PUSH 机制的时间角度确认失效具有一定武断性。而本文提出的 PULL&PUSH 检测机制是由 PULL 配合 PUSH 机制对疑似失效节点做确认检测并输出

一个关于可信度值 ρ ,结合对应 QoS 要求量化处理并与自适应的阈值 P 进行对比,以此来判断节点失效,不仅利用了 PUSH 算法的简单有效、节省通信开销的优势,也进一步增加了检测机制的可靠性。

(2) 失效阈值 P 的计算

本文提出的检测机制中,阈值 $P = \text{MAX} \left\{ \frac{1 + \sqrt{1 - 4\Delta t_i / T_{MR}^U}}{2(1 - P_L)}, \frac{\Delta t_i}{T_M^U(1 - P_L)} \right\}$ 的计算过程借鉴了文献[4]的思路,此阈值能满足新算法的需要,证明如下:

检测机制采用 PULL 模式的策略,假设 b 状态正常,在传统超时认证机制, a 在 τ_{i+1} 时刻没有收到应答消息的概率 $P_0 = (1 - P_L)P(D > \Delta t_0) + P_L$,此种情况存在两个引理:

引理 1 $E(T_{MR}) = \Delta t_i / P_s$,其中 $P_s = (1 - P_L)P(D < \Delta t_0) \cdot [(1 - P_L)P(D > \Delta t_0) + P_L]$

引理 2

$$\begin{cases} \frac{\Delta t_i}{(1 - P_L)P(D < \Delta t_0)[(1 - P_L)P(D > \Delta t_0) + P_L]} \geq T_{MR}^U \\ \frac{\Delta t_i}{(1 - P_L)P(D < \Delta t_0)} \leq T_M^U \end{cases}$$

根据条件 $E(T_{MR}) \geq T_{MR}^U, E(T_M) \leq T_M^U$,可得:

$$\begin{cases} \frac{\Delta t_i}{(1 - P_L)P(D < \Delta t_0)[(1 - P_L)P(D > \Delta t_0) + P_L]} \geq T_{MR}^U \\ \frac{\Delta t_i}{(1 - P_L)P(D < \Delta t_0)} \leq T_M^U \end{cases}$$

求解可得:

$$\begin{cases} P(D < \Delta t_0) \geq \frac{1 + \sqrt{1 - 4\Delta t_i / T_{MR}^U}}{2(1 - P_L)} \\ P(D < \Delta t_0) \geq \frac{\Delta t_i}{T_M^U(1 - P_L)} \end{cases}$$

如果 $1 - 4\Delta t_i / T_{MR}^U \geq 0$ 不成立,则证明平均每四个检测周期会检测出一次错误,得到的错误发生率过高,因此在参数设置时要选择合适的 Δt_i 满足不等式和算法要求。

由单边切比雪夫不等式,在 $\forall t_{iqu} > t$ 时得出:

$$\text{在查询时刻 } \tau_{i+1}, P(D < t) \geq 1 - \frac{V(D)}{(t - E(D))^2 + V(D)}, \text{ 令}$$

$t_{iqu} = t_a - t_r$,由算法得: $\rho \leq P(D < t_{iqu})$ 且 $t_{iqu} \leq \Delta t_0$,随着超时时间的设定值增大,得到不超时的概率增大,即 $P(D < X)$ 随 X 增大而增大,当 $X = \Delta t_0$ 时, $P(D < \Delta t_0)$ 具有下限: $\text{Max} \left[\frac{1 + \sqrt{1 - 4\Delta t_i / T_{MR}^U}}{2(1 - P_L)}, \frac{\Delta t_i}{T_M^U(1 - P_L)} \right]$,即为阈值 P 。

当 ρ 大于阈值时: $P(D < t_{iqu}) > \text{Max} \left[\frac{1 + \sqrt{1 - 4\Delta t_i / T_{MR}^U}}{2(1 - P_L)}, \frac{\Delta t_i}{T_M^U(1 - P_L)} \right]$ 。

由于 $P(D < t_0) > \text{Max} \left[\frac{1 + \sqrt{1 - 4\Delta t_i / T_{MR}^U}}{2(1 - P_L)}, \frac{\Delta t_i}{T_M^U(1 - P_L)} \right]$,而且函数 $P(D < X)$ 单调递增,因此此时 $t_{iqu} \geq \Delta t_0$,判断被检测对象失效。

(3) 算法有效性

进程 a 检测进程 b 时,如果 b 失效,则存在 $t_{iqu} \geq t$,此机制能够判定 b 失效。一旦 b 失效则不会应答 a 发出的任何消息,所以 b 统计的信息超时时间的 $E(D)$ 和 $V(D)$ 将不会变化。根据式(1),当 $t_a - t_r > \mu$ 时, ρ 是呈递增趋势的。设 m_i 是 a 在 t_{last} 时刻从 b 收到的最后一个应答消息,则在 $t > t_{last}$ 的任何时刻 t_r 都不会发生变化。随着 t_r 的增加, ρ 值单调递增,则一定存在一个时刻 $t > t_{last}$,使得 $\rho = P$,且 $\forall t_{iqu} > t, \rho > P$ 恒成

立。由此,在此机制运行过程中任意大于 t 的时刻,都会判定 b 为失效进程。相反 b 检测 a 失效的证明过程同上。

4 心跳机制测试

(1) 通信开销测试

根据式(3)、式(5),只使用 PULL 模式和 PULL&PUSH 结合模式通信开销分别为 K_{PUSH} 和 $K_{\&}$, 设 p 为 0.02, n 为 15 通过 Matlab 仿真得出开销结果如图 2 所示。

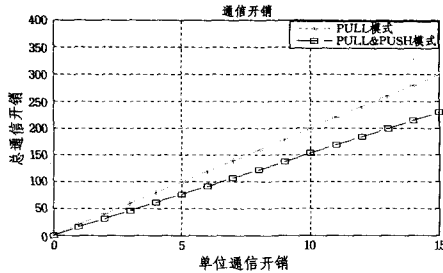


图2 两种模式开销仿真结果

在相同的单位通信开销下,结合 PULL 和 PUSH 模式是对单纯 PUSH 心跳检测的改进,大大降低了通信开销。

(2) 自适应性能测试

利用搭建好的服务器集群作为仿真实验平台,选择主服务器和其中一台从服务器作为主从节点,互相发送检测信息。延迟数据的采集是在校园局域网上完成的,具体过程是利用 ping 程序对学校主页网站和新闻网站进行响应测试,从周一早 7 点到周三早 7 点,共 48 小时,统计近万条响应数据。从数据可以得出,最小的响应时间为 18ms,最大的响应时间在 5000ms 左右,信息的丢失率控制在 1% 以内,满足容灾系统中网络延迟和丢包的数据范围,在实验中可用。

由算法有效性,将 $\rho=P$ 时的 t_{iqu} 定义为超时时间。并根据 $1-\delta/((t_{iqu}-\mu)^2+\delta)=P$ 得到 $t_{iqu}=\mu+\delta\sqrt{P/(1-P)}$, 其中 μ 代表超时时间的期望,而 δ 为超时时间方差的开方,此时 P 为阈值。由此可见,超时时间为一个随机变化值,值的大小只与信息超时时间和阈值 P 有关。又因为超时时间 T_0 的存在,算法超时时间为 $\text{Min}[T_0, \mu+\delta\sqrt{P/(1-P)}]$ 。

我们从校园网测得的数据中顺序截取 1000 条,得到样本的响应时间期望值 0.4221, 方差值 0.00914; 对应参数设置如下: 令平均错误间隔 T_{MR} 为 100s, 信息发送时间间隔 Δt_i 为 1s, 网络信息丢失率为 0.001, 错误平均持续时间 T_M 为 2s。考虑到该机制的自适应性,将 T_0 的值设置为 3s。

测试过程中,启动到 PULL 模式时,机制自动调整超时时间如图 3 所示。图中,纵坐标表示对应编号消到达时刻系统计算出的超时时间,横坐标表示信息编号。根据公式 $t_{iqu}=\mu+\delta P/(1-P)$ 和以上数据得出,该机制输出的超时时间为传输中信息的偶尔延迟,因此可以适应网络变化,具有自适应性。

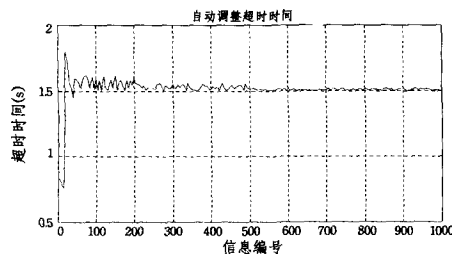


图3 自动调整超时时间间隔仿真

(2) 检测错误率测试

令 λ_f 表示失效检测中失败的概率。在面向容灾任务的检测机制中,一旦检测节点失效,就会进行资源切换,在此过程将付出巨大的资源转换代价,因此检测的可靠性指标是衡量整个机制的重要指标。由文献[2]的定义,测试过程的平均错误与 $\lambda_f=1/E(T_{MR})$ 直接相关。

图 4 显示了测试过程中 T_D 对于 λ_f 的影响。实验中从校园主页网测得的数据中顺序取 3000 条,响应时间的样本信息丢失率约为 0.1%, 期望值为 0.1232, 方差值为 0.04333。对应参数设置如下: 网络信息丢失率为 0.001, 错误平均持续时间 T_M 为 2s, 信息发送时间间隔 Δt_i 等于 1s, 平均错误间隔 T_{MR} 为 100s。实验通过设置 100ms, 200ms, 300ms, 400ms, 500ms, 600ms 和 700ms 7 个值来检测平均错误率。图中横坐标代表设置的超时时间 T_0 , 纵坐标表示该值下 1000 条数据的平均错误率。从图 4 中可以看出,当 T_0 增大到某一个值后,平均错误率降低到预设标准。

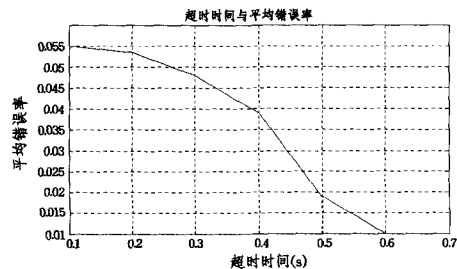


图4 超时时间 T_0 对 λ_f 的影响

结束语 本文提出一种面向容灾自适应的 PULL & PUSH 心跳检测机制, PUSH 机制用于日常检测, 当发现疑似失效情况时切换 PULL 机制, 此时量化输出和特定阈值根据对应 QoS 需求而定。这种方法主要利用 PUSH 机制, 简化全盘使用 PULL 机制的系统开销, 并在 PULL 机制自适应性克服了传统心跳机制中判断失效、方式简单、QoS 无法量化、无法解决复杂环境下网络检测等问题。在后续工作中将针对如何满足多节点复杂网络环境下降低通讯开销、提高检测准确率方面进行深入研究。

参考文献

- [1] de C P, Mo D, REILLY R B. Automatic classification of heartbeats using ECG morphology and heartbeat interval features [J]. IEEE Transactions on Biomedical Engineering, 2004, 51 (7): 1196-1206
- [2] Hu Jia, Min Ge-yong, Woodward M E. Performance analysis of a threshold-based dynamic TXOP scheme for intra-AC QoS in wireless LANs [J]. Future Generation Computer Systems, 2013
- [3] 王明, 张春熹, 伊小素. 基于自适应心跳算法的分布式系统故障检测器 [J]. 北京航空航天大学学报, 2013, (7): 952-956
- [4] Chen W, Toueg S. On the Quality of Service of Failure Detectors [J]. IEEE Trans on Computers, 2002, 51(1): 561-580
- [5] 陈宁江, 魏峻, 杨波, 等. Web 应用服务器的适应性失效检测 [J]. 软件学报, 2005, 16(11): 1929-1938
- [6] Chen Ning-jiang, Wei Jun, Yang Bo. Adaptive failure detection of Web application server [J]. Journal Software, 2005, 16(11): 1929-1938
- [7] 侯宗浩, 董小社, 郑守淇, 等. 一种支持负载均衡的多机心跳模型 [J]. 小型微型计算机系统, 2005, 26(1): 11-14
- [8] De Chazal P, O'Dwyer M, Reilly R B. Automatic classification of heartbeats using ECG morphology and heartbeat interval features. [J]. IEEE Transactions on Biomedical Engineering, 2004, 51