# Optimization of Failure Detection Retry Times

Aad P. A. van Moorsel
Hewlett Packard Laboratories
1501 Page Mill Rd., Palo Alto, CA
94034, USA
aad@hpl.hp.com.com

and

Katinka Wolter
Humboldt-Universität Berlin
Institut für Informatik
Unter den Linden 6, 10099 Berlin,
Germany
wolter@informatik.hu-berlin.de

## Abstract

We present an on-line algorithm for optimizing and adapting the retry times for failure detection mechanisms. The optimization is carried out with respect to both the expectation and the variance of total detection time. We discuss in this paper the situation with transient failures.

## 1 Model

Assume one uses a 'ping' mechanism to diagnose if a service is up or down. That is, the 'source' sends a message to the diagnosed service and expects a reply stating the service is up. If the reply does not come, either the service is indeed down, or some failure happened in the transmission or computation needed to process pings. In the latter case, one may retry the ping, possibly with better luck. The question that arises is how quick and how often one should retry.

This paper provides an on-line scalable algorithm that optimizes and adapts the retry time for situations where only transient failures need to be considered. That is, we assume the service is not experiencing a hard or permanent failure but that other phenomena caused the source not to receive a reply. This assumption allows us to model consecutive diagnosis attempts as independent of each other. Our analysis also assumes that retries preempt earlier attempts, thus leading to a worst case analysis. The resulting analysis and algorithms are rather elegant, and provide the base for future work on systems with mixtures of transient and permanent failures.

## 2 Basic Results

Let $f(t)$ be the probability density function of the return time without retries. Assume $\tau$ is a retry time, and the overhead associated with retrying is $c$ time units for each retry. Then we can introduce the random variable $T_\tau$ corresponding to the total diagnosis time when an unlimited number of retries is allowed. That is, every time no response is received at the source after $\tau$ time units, one retries. Assuming that a retry preempts the previous attempt, the expected total diagnosis time $E_\tau$ is given in [1] as:

$$E_\tau = \frac{M_\tau}{p_\tau} + \frac{1-p_\tau}{p_\tau}(\tau + c), \qquad (1)$$

where

$$p_\tau = \int_0^\tau f(t)dt, \qquad (2)$$

and

$$M_\tau = \int_0^\tau t f(t)dt. \qquad (3)$$

Maurer and Huberman [1] also remark that both the expectation and variance of $T_\tau$ have their minimum at the same time value $\tau^*$. Moreover, one can derive that for this optimal retry time $\tau^*$ it holds that:

$$\frac{1-p_{\tau^*}}{f(\tau^*)} - c = E_{\tau^*}. \qquad (4)$$

Under certain regularity conditions, this implies that a retry time $\tau$ is not optimal and should be increased as long as

$$\frac{1-p_\tau}{f(\tau)} - c < E_\tau. \qquad (5)$$

## 3 The Algorithm

In our scalable on-line algorithm, we assume we collect data for a system with some retry time $\tau$ set beforehand, out of our control. Based on the collected data, we adapt $\tau$ to improve the expectation and variance of the total diagnosis time $T_\tau$. Of course, if one continues collecting data, the amount of data eventually gets prohibitely large. We therefore keep track of results per 'bucket,' that is, we divide the observations over $H$ buckets, each of size $h = \tau/H$, and only keep track of the average return time $M_i$ and number of samples $N_i$ within each interval ($i = 1, 2, \ldots, H$). In the $i$-th bucket, we thus consider the observations with values in the interval $[(i-1)\cdot h, i \cdot h]$, and we label the observations $t_{i,1} \ldots t_{i,N_i}$. Then $M_i$ is estimated by:

$$\hat{M}_i = \frac{1}{N_i}\sum_{j=1}^{N_i} t_{i,j}.$$

We also keep count of $N_\tau$, the total number of observations that take at least $\tau$ time units. (Note that for these observations a retry is initiated, and that this number thus also includes possible failures.) For candidate retry time $\tau_i = i \cdot h$, we then obtain the following estimators for (2) and (3):

$$\hat{p}_{\tau_i} = \frac{\sum_{k=1}^{i} N_k}{\sum_{k=1}^{H} N_k + N_\tau},$$

$$\hat{M}_{\tau_i} = \frac{\sum_{k=1}^{i} N_k \cdot \hat{M}_k}{\sum_{k=1}^{i} N_k}.$$

We estimate the expected diagnosis time $E_{\tau_i}$ by the asymptotically unbiased estimator

$$\hat{E}_{\tau_i} = \frac{\hat{M}_{\tau_i}}{\hat{p}_{\tau_i}} + \frac{1-\hat{p}_{\tau_i}}{\hat{p}_{\tau_i}} \cdot (\tau_i + c).$$

We can also estimate the probability density at each interval, to get for (5):

$$\frac{1-\hat{p}_{\tau_j}}{\frac{N_j}{(\sum_{k=1}^{H} N_k + N_\tau)\cdot h}} - c \quad < \quad \hat{E}_{\tau_j}. \tag{6}$$

This latest result we use to judge if the used retry time should be increased beyond its current value.

# 4 Simulation

We test our algorithm by sampling return times from what we term the 'failed exponential' distribution. The failed exponential is a degenerate distribution: it is exponentially distributed with probability $1-f$, and takes 'infinite' value with probability $f$. The latter case represents the situation in which no reply is received at the source, thus indicating a transient failure. For the failed exponential, (2) and (3) become:

$$p_{f,\tau} = (1-f)(1-e^{-\lambda\tau}), \tag{7}$$

$$M_{f,\tau} = (1-f)(\tau + \frac{1}{\lambda}(1-e^{-\lambda\tau})). \tag{8}$$

Provided $f > 0$ and $c > 0$, we derive for the optimal retry time $\tau^*$ that

$$e^{\lambda\tau^*} = 1 + \lambda\tau^* + \frac{\lambda c}{f}. \tag{9}$$

These analytic expressions and equation we compare against our estimators. It is worthwhile to note that if $f = 0$ and $c = 0$ all retry times are equally good (and have expected diagnosis time $\frac{1}{\lambda}$), as it should be for the memoryless exponential distribution. If $f = 0$ and $c > 0$ it never pays off to retry. If $f > 0$ and $c = 0$ (that is, retrying consumes no time), one should always retry immediately again ($\tau^* = 0$)! The non-pathological case arises for $f > 0$ and $c > 0$, in which case there is an optimum retry time $\tau^* > 0$, as follows from (9). Also realize that retrying *guarantees* that the expected diagnosis time is finite, even though the expected return time for a single ping has non-defined ('infinite') expectation. This important result was also observed in [2] in general setting.

In our simulation the failed exponential has parameters $\lambda = 1$, $f = 0.5$, and the cost of retrying is $c = 1.0$. It turns out that the optimal retry time is $t^* = 1.505$, with expected total diagnosis time equal to $E_{\tau^*} = 4.505$. The following tables then show the estimate of $\tau^*$ and $E_{\tau^*}$ as a function of the number of samples and of the size of the buckets. Since we only consider the multiples of bucket size as candidates retry times, we present for each bucket size the theoretical best $\tau^*$ and associated $E_{\tau^*}$.

As expected, with smaller bucket size, one tends to get better results, but note that both $\tau^*$ and $E_{\tau^*}$ do not converge particulary fast to the analytic results, and that smaller buckets do not necessarily lead to retry times closer to $\tau^*$. More rigorous analysis involving confidence intervals is needed to identify what bucket size works best.

| samples/ bucket size | 1000 | 10000 | 100000 | 1000000 | best |
|---|---|---|---|---|---|
| 0.64 | 1.28 | 1.92 | 1.28 | 1.28 | 1.28 |
| 0.32 | 1.6 | 1.6 | 1.6 | 1.6 | 1.6 |
| 0.16 | 1.44 | 1.44 | 1.44 | 1.44 | 1.44 |
| 0.08 | 1.44 | 1.44 | 1.52 | 1.52 | 1.52 |
| 0.04 | 1.44 | 1.44 | 1.48 | 1.48 | 1.52 |
| 0.02 | 1.44 | 1.46 | 1.48 | 1.48 | 1.5 |
| 0.01 | 1.44 | 1.45 | 1.48 | 1.48 | 1.51 |

Table 1: Optimal retry time $\tau^*$ as function of sample and bucket size

| samples/ bucket | 1000 | 10000 | 100000 | 1000000 | best |
|---|---|---|---|---|---|
| 0.64 | 4.495 | 4.509 | 4.517 | 4.548 | 4.543 |
| 0.32 | 4.423 | 4.448 | 4.475 | 4.517 | 4.511 |
| 0.16 | 4.356 | 4.440 | 4.468 | 4.511 | 4.508 |
| 0.08 | 4.356 | 4.440 | 4.465 | 4.509 | 4.505 |
| 0.04 | 4.356 | 4.440 | 4.464 | 4.509 | 4.505 |
| 0.02 | 4.356 | 4.439 | 4.464 | 4.509 | 4.503 |
| 0.01 | 4.356 | 4.436 | 4.464 | 4.509 | 4.503 |

Table 2: Expected total diagnosis time $E_{\tau^*}$ as function of sample and bucket size

# References


[1] S. M. Maurer and B. A. Huberman, "Restart strategies and Internet congestion," in Journal of Economic Dynamics and Control, vol. 25, pp. 641–654, 2001

[2] W. Chen, S. Toueg and M. K. Aguilera, "On the Quality of Service of Failure Detectors," in Proceedings of the International Conference on Dependable Systems and Networks (DSN 2000), pp. 191–200, 2000.