

Virtual Machine Dynamic Migration Strategy Based on Load Prediction

Wei Liu^{a*}, Chaojun Fu^b, Yu Zhang^c and Hongxin Wang^d

Network Center of Shenyang Jianzhu University, Shenyang, China

^awind@sjzu.edu.cn, ^bcos@sjzu.edu.cn, ^azy@sjzu.edu.cn, ^bchw@sjzu.edu.cn

Keywords: Virtual machine; Dynamic migration; Time sequence; Load prediction; Historical data

Abstract. This article studies the trigger strategy in dynamic migration process of virtual machine under cloud computing environment. For the problem of resource wasting caused by instantaneous peak in simple double threshold trigger strategy and single threshold trigger strategy, we propose ARMA time series prediction model based on historical load value data of virtual machine. The experiments indicate our migration trigger strategy can reduce the times of virtual machine migration effectively and corresponding energy consumption.

Introduction

Virtual machine dynamic migration means when virtual machine can provide services in operation, it will migrate information on a virtual platform server to another virtual platform server. When the CPU workload is unbalanced, a part of the CPU overload occurs and the other part of the CPU get idle status[1,2]. In this case, the resources can not be utilized very well, which is a waste of resources. Then the energy consumption is increased and the performance of the system is reduced. The virtual machine migration technologies emerge for such problems. During the migration process, choosing appropriate virtual machine migration triggering strategy can reduce the number of migration and save energy consumption. This article first discusses the basic concept and process of virtual machine dynamic migration, and expounds the energy consumption of virtual machine and the energy consumption of virtual machine dynamic migration. Then it analyzes the timing and strategy of virtual machine dynamic migration triggering, and provides basic requirements to design the strategy. Finally, a dual threshold trigger strategy based on ARMA model is designed in detail. The simulation experiments demonstrate the advantages of the dual threshold migration and trigger strategy based on timing prediction in two aspects, namely, the virtual machine energy consumption and the times of virtual machines migration.

Analysis on the Trigger Time of Dynamic Migration

The popular understanding of virtual machine migration triggering strategy aims at the time of the virtual machine migration to achieve maximum utilization of resources. It is mainly used to judge the necessity of the virtual machine migration on the host computer based on the utilization rate of the host CPU. Since the utilization rate of CPU is not stable, the waste cost brought by the pointless migration of instantaneous load peak or valley value is minimized as far as possible. Only when CPU utilization exceeds the set of double threshold time can it be migrated[3]. The purpose of this paper is studying the dual threshold triggering mode by using a certain historical data, with time series prediction method. The CPU utilization rate is set corresponding to a high threshold. Assuming the threshold value is less than that in the host load, then in the next N_{th} load value after time t , if there is M load value beyond to set threshold, the virtual machine migration will occur only when the average weight of N load values also exceed the set threshold. Because the utilization rate of CPU has a momentary concussion, when a CPU utilization rate is at instantaneous peak value or valley value, there will be unnecessary virtual machine migration. For such case, the prediction technology can be used.

Migration Trigger Strategy Based on Time Series Prediction

Stability Test and Pattern Recognition. In this paper, the resource state value of the virtual machine and its stability remain to be tested. We can use difference sequence method to determine whether the sequence of the state value of the virtual machine is stable.

Assuming $\{Y_t\}$ is nonstationary sequence, and it is smoothed by difference. $\Delta Y_t = Y_t - Y_{t-1}$ denotes the first order difference and $\{\Delta Y_t\}$ denotes the first order difference sequence. Similarly, $\Delta^2 Y_t = \Delta Y_t - \Delta Y_{t-1}$ denotes the second order difference and $\{\Delta^2 Y_t\}$ denotes the second order difference sequence. In actual case, nonstationary first order difference is stationary sequence and the second order difference from the highest to nonstationary ones are almost stationary sequences.

Whatever the time sequence of virtual machine load is stationary sequence, it can be assured to be stationary one. Then we need to confirm the order numbers of the model after the test of stationary sequence.

First, we compute the estimation of autocorrelation coefficient as

$$\hat{\rho}_k = \frac{\sum_{t=1}^{n-k} [Y_t - E(Y)][Y_{t+k} - E(Y)]}{\sum_{t=1}^n [Y_t - E(Y)]^2} \quad (1)$$

where $E(Y)$ denotes the expected value and it can be thought to be the mean value of Y .

We define $\hat{\phi}_{kk}$ as

$$\hat{\phi}_{kk} = \begin{cases} \hat{\rho}_1, & k = 1 \\ \hat{\rho}_k - \frac{\sum_{j=1}^{k-1} \hat{\phi}_{k-1,j} \hat{\rho}_{k-j}}{1 - \sum_{j=1}^{k-1} \hat{\phi}_{k-1,j} \hat{\rho}_j}, & k = 2, 3, \dots \end{cases} \quad (2)$$

For $AR(P)$, $\hat{\phi}_{kk} = 0, k > p$, and $\rho - k$ of $MA(q)$ model and $AR(q)$ is similar. Therefore, we get the following determinant:

MA model determinant: Supposing $\hat{\rho}_k$ lagged q-order truncation, while $\hat{\phi}_{kk}$ is trailing, its order is q ;

AR model determinant: Supposing $\hat{\phi}_{kk}$ lagged p-order truncation, while $\hat{\rho}_k$ is trailing, its order is p ;

Thus, it can be seen that such discriminant methods accurate and effective for both *MA* model and *AR* model. However, if partial autocorrelation coefficient of $\hat{\phi}_{kk}$ and autocorrelation coefficient cannot locate the truncation, extended autocorrelation algorithm EACF can be used to determined the order of sequene model.

When the sliding average model *AR* of hybrid model ARMA is determined[4], *AR* will be removed from the sequence. Since the coefficient of autoregressive model can be achieved by multiple regression estimations and the computattion quantity is small, we can acquire autoregressive model *MA*, and the order q can be determined by $\hat{\rho}_k$.

Virtual Machine Load Prediction

Time sequence based historical data, that is, $Y_1, Y_2, \dots, Y_{t-1}, Y_t$, denotes the virtual machine load at time 1, 2, ..., t. For prediction value Y_{t+l} , t is prediction starting point and l is the length of preset time. By statistical principle we know when variable X is known, the best prediction for Y is the conditional

expectation of Y to X , according to the rule of the minimum MSN. In known conditions, the optimal predictive expression is

$$\hat{Y}_t(l) = E(Y_{t+l} | Y_1, Y_2, \dots, Y_t) \quad (3)$$

The recursive expression for prediction using $ARMA(p, q)$ model is:

$$\begin{aligned} \hat{Y}_t(l) = & \hat{\phi}_1 \hat{Y}_t(l-1) + \hat{\phi}_2 \hat{Y}_t(l-2) + \dots + \hat{\phi}_p \hat{Y}_t(l-p) + \theta_0 \\ & + \theta_1 E(e_{t+l-1} | Y_1, Y_2, \dots, Y_t) + \theta_2 E(e_{t+l-2} | Y_1, Y_2, \dots, Y_t) \\ & + \dots + \theta_q E(e_{t+l-q} | Y_1, Y_2, \dots, Y_t) \end{aligned} \quad (4)$$

Above process depicts the prediction of future load values detail through the way of prediction of time series. The load time is calculated before the N_{th} time points. The result is calculated by ARMA (p, c) model which is the first load value in the future, and the predicted value is added to the observed value. The obtained value is the second time to predict historical data and to calculate the second load prediction values, so we can deduce the predicted value of the N_{th} load. Then, we compute the load value according to weight a_1, a_2, \dots, a_n orderly, whose weights obey the distribution of function $y = 1/x$.

The prediction model is used to calculate and deduce the load value, and the residual sequence of the model is tested at the same time. ARMA(p, q) model will be a very good time series autocorrelation display[5], and stochastic time series or factors are related to periodic time series. Since the number of the samples is very limited, if the prediction point is estimated further, the prediction result and the actual use of the deviation between the values will be greater. In actual use, we need to get recent load data and modify the models used to get the most accurate results, so the predicted value and development trend of virtual load can be predicted most accurately.

Simulations

Select a certain period average slices, and each time slice can be set for a second or two seconds. The value of the monitoring each time are recorded. The monitoring items are virtual machine CPU usage, the average load of the use of memory and the virtual machine system virtual machine. In the simulation environment of Cloud Sim, the migration time series prediction algorithm based on double threshold trigger strategy test the migration of energy consumption. The test results will be recorded as shown in table 1, which also tested the single threshold energy consumption data migration strategy algorithm ST algorithm and existing simple double threshold strategy of DT algorithm[6,7].

Table 1 Comparison of energy consumption of virtual machine migration

Energy	Threshold				
	0.5	0.6	0.7	0.8	0.9
ST algorithm	14.45	14.06	13.92	13.81	13.74
DT algorithm	12.19	12.05	11.79	11.88	11.02
Our strategy	11.18	10.91	10.88	10.78	10.64

To verify the difference of the energy consumption among three migration strategies, we calculate the migration energy gap of three algorithms under the same threshold condition, and set the migration energy consumption value of the our strategy as 0. Then the relative energy consumption values of ST algorithm and the DT algorithm are calculated as shown in Table 2.

Table 2 Comparison of energy consumption of virtual machine migration

Energy	Threshold				
	0.5	0.6	0.7	0.8	0.9
ST algorithm	3.21	3.17	3.1	3.03	3.08
DT algorithm	1.7	1.18	1.16	1.05	0.37
Our strategy	0	0	0	0	0

The values in this table is relative energy values, so the we can clearly see the performance of Sthree algorithms . Whichever in any threshold, the energy consumption value relative to our strategy are positive, so it indicates that the migration strategy of weighted prediction algorithm based on dual threshold migration energy consumption of improved algorithm is better.

The above analysis shows that for the migration of energy consumption, the prediction of dual threshold strategy has the minimum energy consumption. The reason is: dual threshold strategy used in this paper get the use of the time series prediction model, and the prediction value is acquired according to the needs of all observation value before the prediction. Considering the load the virtual machine value and prediction value of the more backward prediction, it will become increasingly inaccurate. Therefore, the weight with the nature of declining is analyzed instead of the average value of simple dual threshold decision. In addition, it avoid some special cases such as migration according to the instantaneous peak and consumes more energy, to achieve the purpose and effect of energy saving.

Conclusion

This paper studies the trigger strategy of virtual resource monitoring and migration in dynamic migration process under cloud computing environment. By the studies of existing single threshold strategy, and other simple double threshold trigger transient peak strategy, it is found to cause the problem such as unreasonable waste of resources of migration or trigger. Thus, we propose the double threshold trigger strategy based on weighted historical data using time series prediction technology to achieve the purpose of saving resource to reduce the migration times. Finally, through the comparison and analysis of experimental results, it is verified that the trigger migration strategy is superior to traditional migration strategy in terms of migration times and energy consumption.

Acknowledgements

This study was supported by the scientific research project of colleges and universities of Liaoning province (LJZ2017036)

References

- [1] Salfner F, Tröger P, Richly M. Dependable Estimation of Downtime for Virtual Machine Live Migration. *International Journal on Advances in Systems & Measurements*, Vol. 5 (2012) No. 2, p.70-88.
- [2] Sun Z H, Hu X L. Live Migration for Virtual Machine Based on Kalman Prediction of Dirty Pages. *Applied Mechanics & Materials*, Vol. 668-669 (2014), p.1363-1367.
- [3] An, Chun Ling, Li, Chun Lin, Luo, You Long, et al. Research on Trigger Strategy of Virtual Machine Dynamic Migration in Cloud Computing. *Advanced Materials Research*, Vol. 926 (2016) No.4, p.2084-2087.
- [4] Raghunath B R, Annappa B. Virtual Machine Migration Triggering using Application Workload Prediction. *Procedia Computer Science*, Vol.54 (2015) 167-176.
- [6] Yang Z, Dong G, Li C. Algorithm Study of Load Balance Based On Live Migration of VM in Cloud Computing. *International Journal of Advancements in Computing Technology*, Vol. 5 (2013) No. 3, p.481-489.

- [7] Yong C, Lin Y, Yi G, et al. Optimizing Live Migration of Virtual Machines with Context Based Prediction Algorithm. Proceedings of International Workshop on Cloud Computing & Information Security, Vol.52 (2013) No.1391, p:441-444.