

Evaluation and Analysis of Web Service Based on Heavy Tailed Distribution

Xun Yuan^{1, a}, Liping Huang^{2, b} Bin Zhang^{3, c}

¹ Institute of Information Science & Engineering, Shenyang Ligong University, Shenyang, China

² Software College, Northeastern University, Shenyang, China

³ College of Information Science & Engineering, Northeastern University, Shenyang, China

^ayuanxun117@163.com, ^bhuanglp@swc.neu.edu.cn, ^czhangbin@ise.neu.edu.cn

Keywords: Web service, Qos, Heavy tailed distribution.

Abstract. More and more web services are composited together to accomplish more complicated tasks. Replaceable web services with same function need to be evaluated effectively, so as to select the best one. This paper selects better service by response time, which is easy to collect on client side machine. According to collected data, this paper points out that response time jumps periodically in a long time period, and in a short time period the response time shows heavy tailed distribution which contains normal distribution. This paper puts forward that if we use tail weight of response time distribution, together with average value of response time, we will evaluate more precisely the web service reflecting uncertainty of web networks. This paper gives method of computing tail weight and comprehensive evaluation metric. Experiments show the validity of the method. The method can also be used for evaluation of web services, web services composition, and cloud service.

Introduction

Service-Oriented Architecture(SOA) is a new generation of software architecture. Services developed based on SOAP, WSDL, BPEL have better loose coupling and higher reusability. By service composition, business process can be build flexibly, efficiently and dynamically. web services providing same or similar functions become more and more. Recently appeared cloud computing also provides cloud services to user. So non functional attribute of service, especially Quality of Service(QoS), become key of service selection and service composition. Quality of service of web service is a series of non-functional attributes[1], such as performance, reliability, availability, security.

Services of service nodes of service composition are selected and replaced based on non-functional attributes mentioned above. Among all these attributes, most important and most intuitive, and most related with user experience, non-functional attribute is response time. This paper assume functional attributes and non-functional attributes including reliability, availability, security, are all satisfied before service evaluation and selection. Evaluation and selection of service in this paper, is based on client experience of response time.

The response time obeys the heavy tailed distribution. This paper from theoretically reasoning to experiment measurement reveals the characteristics of the heavy tailed distribution of the Internet response time. Through the establishment of log file, from historical data collected, the heavy tail index can be found, and weight of tail can be calculated. Weight of tail can measure uncertainty of response time. This paper will give a calculation method of tail weight. The selection of the service is based on the dual criteria of the expected value of the response time and the tail weight. The service evaluation calculates the weighted sum of response time expectation and tail weight as the evaluation score.

The same method can be applied to the QoS metric of composite service. With an instance of combination of services as key, log file of response time can be established, and response time expectation and heavy tail index of the combination of service can be calculated and used to evaluate

the combination of service. This method can evaluate the combination of service without establish mathematical model for various combination modes.

Cloud service of cloud computing, can be evaluated in the same way as above.

On a group of data collected, we observed response time in Fig.1 jumps up and down in a long period, so it cannot be modeled by heavy tailed distribution. But in a short period we can figure out busy degree of internet and the response time obeys heavy tailed distribution.

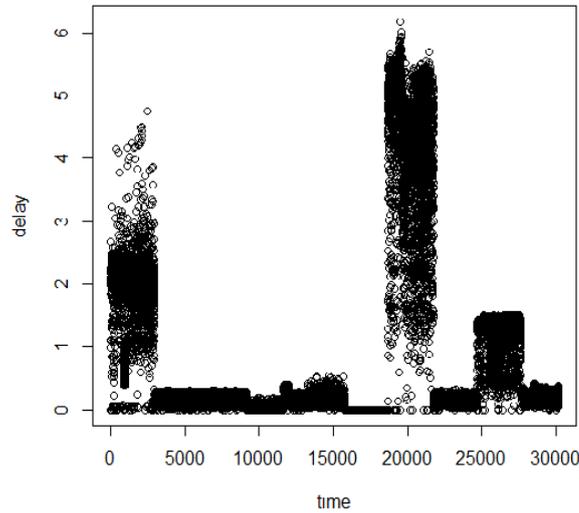


Figure 1 stage jump in long term observation

Related Work

Network delay is main part of response time. Many researches is done on network delay. In [2], the authors present a heavy tailed distribution research on point to point delay of internet nodes. But this paper will do the research from the perspective of the service web client. In [3] it is pointed out that internet file sizes obey heavy tailed distribution. This paper further points out internet service response time obeys heavy tailed distribution and tries to use the conclusion on service evaluation and selection.

Many Researches are also done on service evaluation and selection. Most researches evaluate service by data on server side. For example [4] use execute time on server side to compute, not response time on client side. So the result must be narrow and not comprise uncertainty of internet and intuitive customer experience. In [5], the authors establish contract between server and client according to client side response time. But because client is in different place in the internet, a client will get different response time. So establishing contract using same response time is not reasonable. [6]gives model on server side, and provide predictions for service performance.

Research of [7] points out response time and other Qos attributes is metrics changing over time. It does real-time intelligent forecasting by recent history data. But the method can't reflect the uncertainty of internet.

In paper[8], T Location-scale distribution is used to describe web service response time. The parameter of the distribution is calculated by recorded data. But the T Location-scale distribution can't reflect heavy tailed attribution of response time.

Modeling Based on Heavy-tailed Distribution

Building up Heavy-tailed Distribution model

Component of Response Time

$$T = \sum_{i=1}^M t_i + t_r \quad (1)$$

In Eq.1, t_i is the delay time for the i th time to transfer packet, M is transmission times. t_r is the time for executing service on server. t_i is made up of sum of waiting time in the queues of client side, and

routing nodes. According to queuing analysis theory, delay time of each time waiting in queue obeys Poisson distribution. If routing node number is big enough, the one time transmission time obeys normal distribution according to law of large numbers. Retransmission number M obeys heavy-tailed distribution. t_r is mainly decided by waiting time in queue of server, and obeys Poisson distribution.

Number of Retransmission Satisfies Simon Model and Obeys Power Law Distribution.

Definition of Power Law Distribution: $p(k)$ is Probability distribution of discrete random variables, and $p(k) = ck^{-(\alpha+1)}$, $\alpha > 0$, $c > 0$. $p(k)$ is called Power Law Distribution.

According to Simon Model [9], service data transfer is described as below:

$N(S)$ is number of package retransmitted when transmitted package number is S . $1 \leq N(s) \leq S$. $f(k,S)$ is number of k times retransmitted package when transmitted package number is S . $\theta(S)$ is probability of new package coming in when transmitted package number is S . Suppose that:

(1) Probability of New package coming in is a constant.

(2) Probability of Newly entered package, which is used k times, is in proportional to $kf(k,S)$. We can have equations below:

$$f(k,S+1) - f(k,S) = K(S) \{ (k-1)f(k-1,S) - kf(k,S) \} \quad k=2, \dots, S \tag{2}$$

$$f(1,S+1) - f(1,S) = \theta - K(S)f(1,S) \tag{3}$$

$K(S)$ is a proportional factor.

(3) Assumption of steady state

$$f(k,S)/S = f(k,S+1)/S+1 \quad k=1, \dots, S \tag{4}$$

Eq.4 is applied to Eq.2 and Eq.3, and make $\alpha = 1/(1-\theta)$, we can have

$$f(k,s) = (k-1)! \theta s \prod_{i=1}^k \frac{1}{\alpha+i} = \alpha \frac{\Gamma(k)\Gamma(\alpha+1)}{\Gamma(k+\alpha+1)} \theta s \tag{5}$$

Among Eq.5

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt \quad (\alpha > 0)$$

It is Gamma function. We use Sterling approximation

$$\frac{\Gamma(k+\alpha)}{\Gamma(k)} \sim k^\alpha \tag{6}$$

Then we can have :

$$f(k,s) \sim \alpha \Gamma(\alpha+1) \theta s k^{-(\alpha+1)} \tag{7}$$

So we know number of k times transmitted package obeys Power law distribution as Eq.8:

$$f(k) = ck^{-(\alpha+1)} \tag{8}$$

Power Law Distribution proves Pareto Distribution. Definition of Pareto Distribution :

Probability of that Random variable X is not less than x exists simple inverse relationship with constant power of x : $P[X \geq x] \sim x^{-k}$.

Suppose $f(k)$ is number of package that is k times transmitted. It obeys power law distribution and Eq.8. Note k_m as the number of the max package transmission times. Transmission times between k and k_m is

$$\sum_{i=k}^{k_m} f(i) = c \sum_{i=k}^{k_m} i^{-(\alpha+1)} \tag{9}$$

When k_m is big enough,

$$P(X \geq k) = \sum_{i=k}^{k_m} f(i) \approx \int_k^\infty cx^{-(\alpha+1)} dx = \frac{c}{\alpha} k^{-\alpha} \tag{10}$$

This equation meets definition of Pareto distribution .

Calculation of heavy tail index

In his paper the method of Hill estimation is adopted for the tail index. Hill estimation is a classical estimation algorithm for Parreto distribution. Suppose that $y(i), i=1,2,\dots,r+1,\dots,k$, are k response time sampling data. The y obeys Pareto distribution

$$P(y \leq x) = 1 - Cy^{-\alpha} \quad (11)$$

In k sample data, select maximum $r+1$ sample, and calculate as below

$$\hat{\alpha} = (r + 1)/[\sum_{i=1}^r \ln y(i) - r \ln y(r + 1)] \quad (12)$$

$$\hat{C} = [y(r + 1)]^{\hat{\alpha}}(r + 1)/k \quad (13)$$

They are maximum likelihood estimation of α and C , respectively.

Comprehensive evaluation of web service

Comprehensive evaluation of web need to calculate average response time and weight of tail of response time distribution.

First we calculate 1/2 quantile from

$$P\{X \leq x\} = 1 - Cx^{-\alpha} = 1/2 \quad (14)$$

and we get the quantile

$$x_{0.5} = (2C)^{\frac{1}{\alpha}} \quad (15)$$

Definition 3.1: **Weight of Tail** is defined by the equation below:

$$G = \int_{x_{0.5}}^{+\infty} xf(x)dx \quad (16)$$

Among Eq.16

$$f(x) = (1 - Cx^{-\alpha})' = C\alpha x^{-\alpha-1} \quad (17)$$

Bring Eq.15 and Eq.17 into Eq.16, and we get

$$G = \int_{(2C)^{1/\alpha}}^{+\infty} C\alpha x^{-\alpha} dx, \quad C>0, \alpha > 1. \quad (18)$$

So we can have:

$$G = \frac{-C\alpha}{1-\alpha} (2C)^{\frac{1-\alpha}{\alpha}}, \quad C>0, \alpha > 1. \quad (19)$$

Evaluation of service is done according to Eq.20:

$$E = W1*T + W2*G, \text{ while } W1+W2=1. \quad (20)$$

where

$$T = \sum_{i=0}^I t_i * v_i \quad (21)$$

where I is number of response time sampled. The values are in seconds for rounding. v_i means frequency of t_i .

Experiment

Experiment depicted below uses data provided by [10].

Verify web service obeys heavy tailed distribution. We use a group of data from [10], whose service provider is planet1.pittsburgh.intel-research.net, and it has 10258 records. Unit of response

time is millisecond and we change unit to second, and result is rounded down. Probability of each response time value is calculated and we have Fig.2.

Calculate according to Eq.22

$$P(X \geq x) = \frac{\sum_{i=x}^{+\infty} f(i)}{\sum_{i=1}^{+\infty} f(i)}. \tag{22}$$

we have Fig 3.

From Eq.11 we can get

$$\text{Log}(P(y \geq x)) = \text{Log}(C) - \alpha \text{Log}(x) \tag{23}$$

So Loglog plot of Fig3 will be a line. This is shown in Fig4. From the fifth point the graphics approximation is a line. This proves the response times obey Pareto distribution.

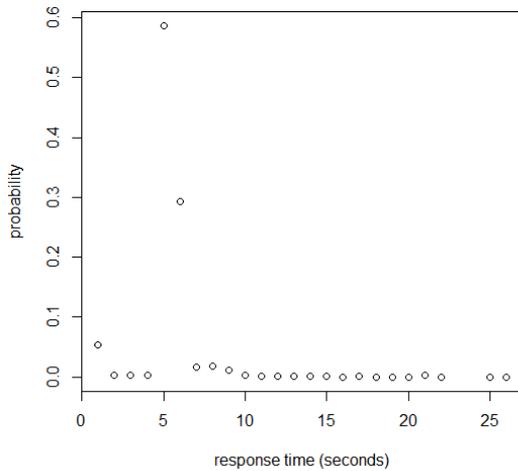


Figure 2. Probability of response time

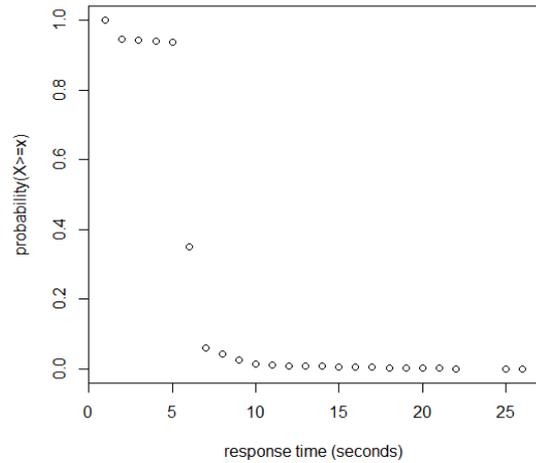


Figure 3 Cumulative probability of response time

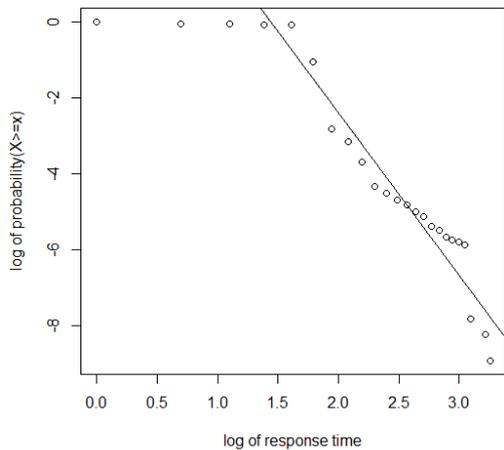


Figure 4 Loglog plot of response time and its cumulative probability

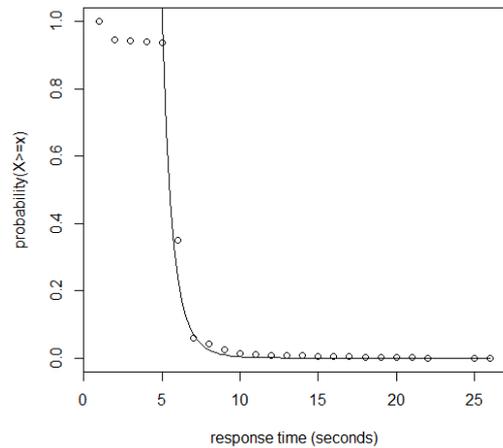


Figure 5 curve by Hill Estimation and cumulative probability of response time

Calculation of indexes of heavy tailed distribution. Without loss of generality, we select 200 records from 10258 records to do Hill Estimation. Unit of Response time is changed from millisecond to second and round down the results. Make $r = 199$, $k = r + 1 = 200$. Response times are sorted from large to small and are brought into Eq.12 and Eq.13. And we get $\alpha = 7.881885$, $C = 322999$. Thus we get a curve that well fits the graph of Fig 3. It's shown in Fig. 5.

Calculation of Weight of Tail. Take value of C and α into Eq.19, then we get weight of tail G is 3.126477.

Comprehensive evaluation of service. According Eq.21, we get average response time $T=5.320917$. According to Eq.20, weight may wish to be $W1=0.5$ and $W2=0.5$ and we get comprehensive evaluation $E=0.5T+0.5G= 4.223697$. Weight in Eq.20 is decided by degree of sensitiveness of client.

Conclusion

The uncertainty of Internet is one of the main challenges for service and service composition. In this paper, the probability of heavy tailed distribution is introduced to analyze the response time, and the degree of uncertainty of network service access is measured. The correctness of the method is proved by experiments.

References

- [1] Shuping Ran: A Model for Web Services Discovery with QoS. SIGecom Exch,4(1):1-10,2003.
- [2] Zhao Pengyu: Measurement-Based Analysis and Modeling of Internet Delay. Dissertation for the Masteral Degree in Engineering. Harbin Institute of Technology. June,2008.
- [3] Crovella, Mark E., et al: Heavy-tailed probability distributions in the World Wide Web. A Practical Guide to Heavy Tails. Birkhauser Boston Inc, 1998, 3-25.
- [4] Florian Wagner, Adrian Klein, Benjamin Klopper, Fuyuki Ishikawa, Shinichi Honiden: Multi-objective Service Composition with Time- and Input-Dependent QoS. 2012 IEEE 19th International Conference on Web Services, p.234-241.
- [5] Huiyuan Zheng, Jian Yang, Weiliang Zhao: QoS Probability Distribution Estimation for Web Services and Service Compositions, 2010 IEEE International Conference on Service-Oriented Computing and Applications (SOCA).
- [6] Chen Leilei: QoS Modeling, Prediction and Assurance in SOA. Dissertation for the Doctoral Degree in Engineering. Fudan University. April, 2012.
- [7] Yilei Zhang, Zibin Zheng and Michael R. Lyu. WSPred: A Time-Aware Personalized QoS Prediction Framework for Web Services. 2011 22nd IEEE International Symposium on Software Reliability Engineering.
- [8] Sidney Rosario, Albert Benveniste, Stefan Haar, and Claude Jard: Probabilistic QoS and Soft Contracts for Transaction-Based Web Services Orchestrations. IEEE TRANSACTIONS ON SERVICES COMPUTING, VOL. 1, NO. 4, OCTOBER-DECEMBER 2008.
- [9] Simon, H.A: On a class of skew distribution functions. Biometrika[J]. 1955,42:425-440.
- [10] Zibin Zheng and Michael R. Lyu: Collaborative Reliability Prediction for Service-Oriented Systems, in Proceedings of the ACM/IEEE 32nd International Conference on Software Engineering (ICSE2010), Cape Town, South Africa, May 2-8, 2010, pp. 35-44.