Journal of Applied Sciences

ISSN 1812-5654

science alert

ANSInet
an open access publisher
http://ansinet.com

Journal of Applied Sciences 13 (9): 1551-1557, 2013
ISSN 1812-5654 / DOI: 10.3923/jas.2013.1551.1557

Networked Intelligent Sensor System Load Balance Based on Pp-gmcp Algorithm

[1]Yuebin Zhou, [2]Guixiong Liu and [2]Haibing Zhu
[1]School of Mechanical and Automotive Engineering, Hubei University of Arts and Science, Xiangyang, 441053, China
[2]School of Mechanical and Automotive Engineering, South China University of Technology, Guangzhou, 510640, China

Abstract: In order to solve the networked intelligent sensor system load balance problem and improve service response speed, a load balance realization method based on Probabilistic Preferred Grey Markov Chain Prediction (PP-GMCP) algorithm is proposed. This method real-time monitors Network Capable Application Processor (NCAP) load status, combines residual correction and grey Markov chain prediction, effectively predicts NCAP load capacity. A load balance simulation platform based on OPNET is constructed to validate algorithm performance. The test shows that the PP-GMCP algorithm effectively improves the service request processing speed, compared to weighted round robin and least connection scheduling its average service response delay reduces 11.1 and 25.1%, respectively, the NCAP load fluctuation range is the smallest and obtains better load balance effect.

Key words: Intelligent sensor, load balance, grey prediction, OPNET simulation

INTRODUCTION

IEEE 1451 standard makes intelligent sensor interface more standardized, compatibility and interoperability, is an ideal plan to build networked intelligent sensor system (Wobschall, 2008). The service requests processed by network intelligent sensor system become increasingly complex with user traffic and data flow growth. Load balance technology needs to be adopted to solve the network congestion problem, improve the server response speed and avoid network key position in failure. In networked intelligent sensor system, common load balance technologies are based on: Domain Name System (DNS), client, reverse proxy, IP layer and high-level protocol content, etc. (Wang and Fu, 2011; Wan and Dong, 2011). As user requests and server load state change constantly it is necessary to introduce dynamic balance strategies into existing load balancing technologies, which are capable of load prediction and consider the global status. This paper study the load balance realization of networked intelligent sensor system based on Probabilistic Preferred Grey Markov Chain Prediction (PP-GMCP) algorithm, analyzes NCAP service response load performance with the aid of queuing models, compares and verifies the PP-GMCP algorithm performance based on OPNET load balancing simulation platform.

LOAD BALANCE FUNCTION MODULES

The load balance function modules are shown in Fig. 1. Load balancer establishes an internal service access probability table and set up the corresponding service queue for different service type. After receiving user service requests, the service classification module resolves service information, determines service type and stores it in the appropriate queue waiting for distribution. Based on service queue length, access priority, service scheduling module uses Probability Priority (PP) strategy to select a service request for distribution. Load balancer regularly collects NCAPs load monitoring parameters, calculates each NCAP load capacity based on Grey Markov Chain Prediction (GMCP) algorithm and residual correction. According to the load capacity, service distribution module uses PP strategy forward service request to the least load NCAP, completes service request processing.

Service classification is based on service request access probability tables and service requests are classified and stored in service queues (Sharifian *et al*., 2009). Each service request type S_{Tj} owns a service queue $C_{Sj}(1 \le j \le 8)$ and the queue length is L_{max}. Then service queue C_{Sj} access priority $P_{Rj}(t)$ depends on current queue residual length $L_{sj}(t)(0 \le L_{Sj}(t) \le L_{max})$, service request arrive rate $\Delta_{sj}(t)$, service request access probability D_{Aj}. Normally, $P_{Rj}(t)$ can be calculated as Eq. (1), while $L_{sj}(t) = 0$, $P_{rj}(t) = 1$ and if $L_{sj}(t) = L_{max}$, $P_{rj}(t) = 0$:

Corresponding Author: Yuebin Zhou, School of Mechanical and Automotive Engineering, Hubei Univeristy of Arts and Science, Xiangyang, 441053, China

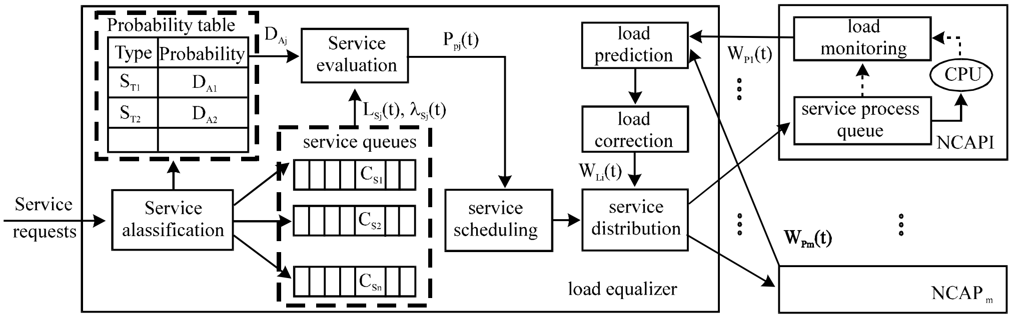

Fig. 1: Load balance function modules diagram

$$P_{Rj}(t) = \frac{(D_{Aj} \times \lambda_{Sj}(t)) / L_{Sj}(t)}{\sum_{j=1}^{8} [(D_{Aj} \times \lambda_{Sj}(t)) / L_{Sj}(t)]}, \quad 0 < L_{Sj}(t) < L_{max} \tag{1}$$

LOAD PREDICTION

Current load capacity calculation: The change of NCAP current load capacity $W_{Pi}(t)$ has a direct impact on load prediction results and service distribution adjustment. NCAP load is related to network bandwidth utilization rate $w_{ri}(t)$, CPU occupancy rate $w_{ci}(t)$, memory usage rate $w_{mi}(t)$, IO utilization rate $w_{ai}(t)$ and process queue occupancy rate $w_{qi}(t)$ (Eludiora *et al*., 2010). Hypothesis R is the aforementioned parameters influence weights on $W_{Pi}(t)$, can be set up according to the actual test, or be dynamically adjusted in real time and $\Sigma R_i = 1$, $W_{Pi}(t)$ can be calculated as Eq. 2. R_i can be changed according to actual circumstances in order to adjust each parameter influence in $W_{Pi}(t)$ (Zheng, 2009):

$$W_{Pi}(t) = R_1 W_{ri}(t) + R_2 w_{ci}(t) + R_3 w_{mi}(t) + R_4 w_{ai}(t) + R_5 w_{qi}(t) \tag{2}$$

Load capacity prediction: The commonly used prediction method includes time series, artificial neural network and grey prediction etc. The time series method takes the load as stationary time series, adopts a linear prediction model, the method is simple but the prediction accuracy is not high. The artificial neural network method does not need to establish an accurate mathematical model and has a good nonlinear approximation ability but needs to determine network structure and requires of the large training sample. The grey prediction method has higher precision, does not need a large number of samples and suitable for the prediction of short time, less data and small fluctuation (Samsudin *et al*., 2010; Devarasiddappa *et al*., 2012; Tan *et al*., 2011). Considering the actual NCAP load is a non-stationary random process, always deviates and swings around a variable trend, the GM(1,1) grey prediction model is used to predict NCAP load.

Get a set of raw data of current load capacity $W_{Pi}(t)$, constitute sequence $W_{Pi}^{(0)}$:

$$W_{Pi}^{(0)} = \{W_{pi}(1), W_{pi}(2), \ldots W_{Pi}(n) \tag{3}$$

Accumulate sequence $W_{Pi}^{(0)}$ to generate data sequence $W_{pi}^{(1)}$:

$$\begin{aligned} W_{Pi}^{(1)} &= \{W_{Pi}^{(1)}, W_{Pi}^{(1)}(2), \ldots, W_{Pi}^{(1)}(n)\} \\ W_{Pi}^{(1)}(t) &= \sum_{d=1}^{t} W_{Pi}(d), \quad t = 1, 2, \ldots, n \end{aligned} \tag{4}$$

Establish differential equation based on sequence $W_{Pi}^{(1)}$:

$$\frac{W_{Pi}^{(1)}}{dt} + aW_{Pi}^{(1)} = \mu \tag{5}$$

Estimate parameters a, u:

$$\hat{\alpha} = [a \quad \mu]T = (B^T B)^{-1} B^T Y \tag{6}$$

$$Y = \begin{bmatrix} W_{Pi}(2) \\ W_{Pi}(3) \\ \vdots \\ . \\ W_{Pi}(n) \end{bmatrix} B = \begin{bmatrix} -\frac{1}{2}[W_{Pi}(1) + W_{Pi}^{(1)}(2)] & 1 \\ -\frac{1}{2}[W_{Pi}^{(1)}(2) + W_{Pi}^{(1)}(3)] & 1 \\ \vdots & \vdots \\ -\frac{1}{2}[W_{Pi}^{(1)}(n-1) + W_{Pi}^{(1)}(n)] & 1 \end{bmatrix} \tag{7}$$

Set up GM(1,1) prediction model:

$$W_{Pi}^{(1)}(t) = (W_{pi}(1) - \frac{\mu}{a})e^{-a(t-1)} + \frac{\mu}{a} \quad (t = 1, 2, \ldots) \tag{8}$$

Predict load capacity $W_{fi}(t)$:

$$W_{Fi}(t) = (\hat{W}_{pi}(t) = \hat{W}_{pi}(t)\text{-}\hat{W}_{pi}^{(1)}(t-1) \\ = (1-e^{a})(W_{Pi}(1)-\frac{\mu}{a})e^{-a(t-1)} \quad (t \geq n) \tag{9}$$

The prediction performance indicators include: Normalized Mean Square Error (NMSE), Mean Absolute Percentage Error (MAPE) and correlation coefficient R_{co} which are defined as follows (Huang *et al*., 2011):

$$\text{NMSE} = \frac{n-1}{n} * \frac{\sum_{t=1}^{n}[W_{Pi}(t)-W_{Pi}(t)^2]}{\sum_{t=1}^{n}[W_{Pi}(t)-\overline{W}_{Pi}(t)^2]} \tag{10}$$

$$\text{MAPE} = \frac{1}{n}\sum_{t=1}^{n}\frac{|[W_{Pi}(t)-W_{Fi}(t)|}{W_{Pi}(t)} \times 100\% \tag{11}$$

$$\text{ROC} = \sum_{t=1}^{n}[W_{Pi}(t)-W_{Fi}(t)] / \sqrt{\sum_{t=1}^{n}W^2{}_{Pi}(t)\sum_{t=1}^{n}W^2{}_{Pi}(t)} \tag{12}$$

Kalman filtering and parameter correction prediction method are adopted to predict a distributed system host load (Ma, 2005). With the aid of GM(1,1) grey prediction model, this paper performs load prediction test using the same data as a test set. Prediction curves of Kalman filtering, parameter correction and GM(1,1) grey prediction three methods are shown in Fig. 2 and prediction performance comparison is shown in Table 1. It can be seen that GM(1,1) grey prediction curve is most approximate to the measured curve. The NMSE, MAPE and R_{co} of GM(1,1) grey prediction method are 0.1074, 0.6265% and 0.99998, respectively it shows that the GM(1,1) grey prediction result is superior to Kalman filtering and parameter correction prediction method.

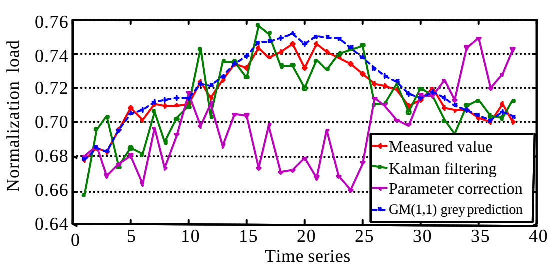

Fig. 2: Prediction curves of different methods

Table 1: Prediction performance comparison

Methods	NMSE	MAPE (%)	Rco
Kalman filtering	0.4866	1.50	0.99985
Parameter correction	4.7669	4.25	0.99895
GM(1,1)	0.1074	0.63	0.99998

Load capacity residual correction: GM(1,1) grey prediction is suitable for relatively smooth curves, once the random fluctuation of NCAP load capacity $W_{Pi}(t)$ increases, the accuracy of load capacity prediction value $W_{Fi}(t)$ can be affected. Markov chain prediction is adopted by the load correction module to correct residual error which can make up for the limited of GM(1,1) grey prediction. The specific mathematical description is as follows:

Calculate the mean value and mean square error S_{pi}:

$$\overline{W}_{Pi} = \frac{1}{n}\sum_{t=1}^{n}(W_{Pi}(t), S_{Pi} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(W_{Pi}(t)-\overline{W}_{Pi})^2} \tag{13}$$

Calculate the current load capacity residuals error $E_i(t)$:

$$E_i(t) = W_{Fi}(t) - W_{pi}(t) \quad (t = 1, 2,\ldots n) \tag{14}$$

Divide state space range of $E_i(t)$:

The state space range of $E_i(t)$ is divided into seven state interval (Q_1~Q_7) according to S_{Pi} value:

$$\begin{aligned} Q_1 &= [0.3S_{Pi}, +\infty) \\ Q_2 &= [0.2S_{Pi}, 0.3S_{Pi}) \\ Q_3 &= [0.1S_{Pi}, 0.2S_{Pi}) \\ Q_4 &= [-0.1S_{Pi}, 0.1S_{Pi}) \\ Q_5 &= [-0.2S_{Pi}, -0.1S_{Pi}) \\ Q_6 &= [-0.3S_{Pi}, -0.2S_{Pi}) \\ Q_7 &= (-\infty, -0.3S_{Pi}) \end{aligned} \tag{15}$$

Due to the sequence $W_{Pi}^{(0)}$ consists of a group of continuous data of time function $W_{Pi}(t)$ its mean square error S_{Pi} also changes with time, thus the state space Q_z (z =1, 2,..., 7) is dynamically changing and state range corresponding central value (θ_1~θ_7) can be set at $0.35S_{Pi}$, $0.25S_{Pi}$, $0.15S_{Pi}$, 0, $-0.15S_{Pi}$, $-0.25S_{Pi}$, $-0.35S_{Pi}$.

Construct $E_i(t)$ one step state probability transfer matrix P_{ei}:

$$P_{Ei} = \begin{bmatrix} \rho_{11} & \rho_{12}\cdots & \rho_{17} \\ \rho_{21} & \rho_{22}\cdots & \rho_{27} \\ \vdots & \vdots & \vdots \\ \rho_{71} & \rho_{72}\cdots & \rho_{77} \end{bmatrix} \tag{16}$$

The p_{xy}(x,y = 1, 2,..,7) is the transition probabilities of $E_i(t)$ (t = 1, 2,...,n-1) from state interval Q_x to the state of Q_y through one step. If M_x is the number of $E_i(t)$ be in the state interval Q_x and M_{xy} is the transition number of $E_i(t)$ from Q_x to Q_y through one step, then:

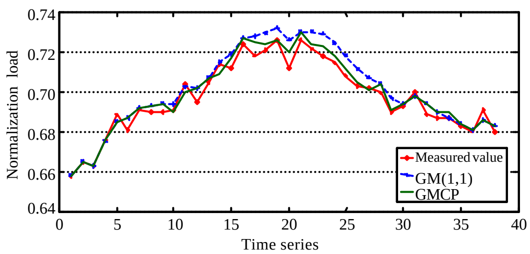

Fig. 3: GMCP) and GM(1,1) prediction curve

Table 2: GM(1,1) and GMCP prediction performance comparison

Methods	NMSE	MAPE (%)	Rco
GM(1,1)	0.1074	0.63	0.99998
GMCP	0.0436	0.42	0.99999

$$\rho_{xy} = \begin{cases} \dfrac{M_{xy}}{M_0^{x}} & (x, y = 1, 2, \ldots, 7 \\ & M_x = 0) \end{cases} \tag{17}$$

Determine the predicted value of $E_i(t)$.

If $E_i(t) \in Q_x(t = n)$ and $p_{xm} = \max\{p_{xy}\}(m \in y)$, take the central value θ_m of state interval Q_m as the $E_i(t)$ predicted value. If p_{xy} has multiple equal maximum value, then choose the minimum central value θ_y as $E_i(t)$ predicted value θ_m.

Calculate the corrected load capacity $W_{Li}(t)$:

$$\begin{aligned} W_{Li}(t) &= W_{Fi}(t) - \theta_m \\ &= (1 - e^{a}(W_{Pi}(1) - \frac{\mu}{a})e^{-a(t-1)} - \theta_m \qquad (t \le n) \end{aligned} \tag{18}$$

Residual correction grey Markov chain prediction (GMCP) and GM(1,1) grey prediction comparison prediction curves are shown in Fig. 3 and the performance comparison is shown in Table 2. It can be seen that GMCP prediction curve is most approximate to the measured curve. The NMSE, MAPE and R_{co} of GMCP method are 0.0436, 0.99999 and 0.4183%, respectively which are superior to GM(1, 1) grey prediction method.

SERVICE SCHEDULING AND DISTRIBUTION

The priority queue algorithm usually ensures high priority queue service with the price of sacrificing low priority queue service. Based on the priority queue algorithm, a probability priority algorithm is adopted in this paper to solve the problem that low priority queue service is not guaranteed. The Probability priority algorithm specific mathematical description is as follows:

- Obtain the queue evaluation index E_{sl}. Each queue provides an evaluation index $E_{sl} \in [0,1]$ $(l = 1, 2, \ldots, h)$, the smaller the numbers l, the higher the queue priority
- Calculate the queue relative weight r_{Sl}. Suppose B is non-empty queues set and g, f is the first and last queue numbers respectively ($g \le f$, g, $f \in B$):

$$r_{Sl} = \begin{cases} E_{Sg} & , l = g \\ E_{Sl}\prod_{v=g}^{l-1}(1 - E_{Sv}), & g < l < f \\ \prod_{v=g}^{l-1}(1 - E_{Sv}), & l = f \end{cases} \tag{19}$$

Calculate the queue standardized relative weight r_{Si}:

$$\hat{r}_{Sl} = \begin{cases} \dfrac{r_{Sl}}{\sum_{b \in B} r_{Sb}} & , l \in B, \sum_{b \in B} r_{Sb} \ne 0 \\ 1 & , l = f, \sum_{b \in B} r_{Sb} = 0 \\ 0 & , \text{else} \end{cases} \tag{20}$$

- Utilizing random number generator, a random number RN ($RN \in [0,1]$), obeying uniform distribution, can be obtained
- Accumulate the standardized relative weight from queue 1, until the accumulative value sum_l bigger or equal to RN:

$$\begin{aligned} & sum_0 = 0 \\ & sum_l = sum_{l-1} + \hat{r}_{Sl} \qquad , sum_{l-1} < RN \end{aligned} \tag{21}$$

The service queue *l* that meets the conditions of (5) is the optimization queue in the current probability interval which determined by current queue evaluation index E_{Sl}, the queue scheduling end and next scheduling will restart from (1).

Probability priority algorithm scheduling model can be applied to the service scheduling module and service distribution module of load balancer. Using the Highest Priority Probabilistic Preferred (HPPP) strategy, service scheduling module choices a service request to implement based on service queue access priority, completes the service scheduling. Based on the NCAP load monitoring data acquired by load balancer timing and the revised NCAP load capacity $W_{Li}(t)$, using the Minimum Load Probabilistic Preferred (MLPP) strategy, service distribution module forwards service request to NCAP and complete service distribution function.

The HPPP service scheduling strategy uses queue access priority $P_{Rj}(t)$ as a queue evaluation index E_{Sl}. The MLPP service scheduling strategy uses $1\text{-}W_{Li}(t)$ as a queue evaluation index E_{Sl}. A simulation test has implemented to compare it with the commonly used Strict Priority (SP) algorithm and Weighted Round Robin (WRR) algorithm. There are 4 data queues, each queue length L_{max} is 100, queue evaluation index E_{Si} are 0.4, 0.3, 0.2, 0.1 respectively, scheduler processing speed is 200 data packets sec, simulation time is 1000 sec, each data packet has the same size of 128 bytes and arrival rate is 0~50 data packets sec. The average queue delay and data loss rate of aforementioned three algorithms is severally shown in Fig. 4 and 5.

It can be seen, when the normalized scheduler load varies between 0~1, the average queue delay of PP algorithm is the minimum, less than 0.38 sec and compared with SP and WRR algorithm the average queue delay reduces 46.89 and 22.99%, respectively. The data loss rate of PP algorithm is slightly higher than WRR algorithm while scheduler load is lower than 0.3. While scheduler load is higher than 0.3, it is less than 8.56%, compared with SP and WRR algorithm, the data loss rate reduces 65.07 and 42.41%, respectively. PP algorithms adjusts probability partition through dynamic queue evaluation index adjustment, reduces the interaction effect between queues and ensure reasonable scheduling of different priority queues which can effectively solve data average queue delay and data loss rate increase problem while data traffic increase.

Access traffics of network intelligent sensor system users are always dynamically changing, their service requests are various and resource needs are different.The service scheduling and distribution of load balancer directly affect the whole system load balance performance. Based on the probability priority service scheduling and distribution algorithm, considering the different service request influence on NCAP resources, according to NCAP current load condition, evenly distributing service requests to different NCAP to realize parallel processing, NCAP can achieve the best utilization.

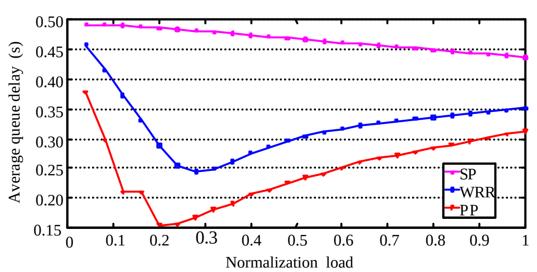

Fig. 4: Average queue delay

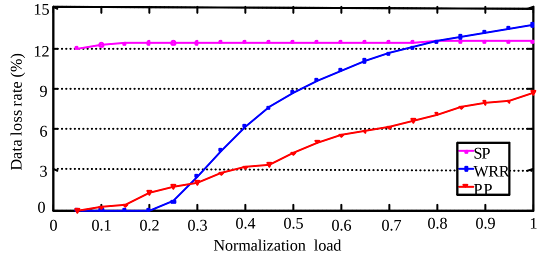

Fig. 5: Data loss rate

SIMULATION AND ANALYSIS

OPNET is a mainstream network simulation and modeling tool software, supports object-oriented modeling method and provides a graphical editing interface, can simulate complex network equipment and the environment, supports all kinds of network technology (Zhang and Hu, 2011). To validate PP-GMCP load balance algorithm comprehensive performance, a load balance simulation platform based on OPNET is constructed.

The load balancing network topology structure is shown in Fig. 6. It adopts 10/100 Mbps star Ethernet as the backbone. Eight end users connect together through a switch which connects a load balancer with 100M twisted-pair cable. The load balancer also connects to 12 NCAPs through a 100M hub; each NCAP has the same performance. The application module defines eight types of business requests and the profile module defines an application service specification. Based on the load balancer node mode of OPNET own, the finite state machine is used to implement PP-GMCP algorithm function. Load balancer node contains 8 queues, each of them has 50 packets queue length and the packet rate is 9600 bps. For comparison, the load balancer node model also embeds two classical load balancing algorithms: Weight Round Robin (WRR) and Least Connection

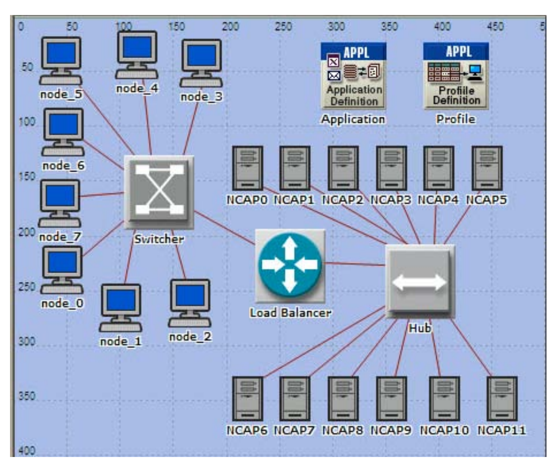

Fig. 6: Load balancing network topology structure

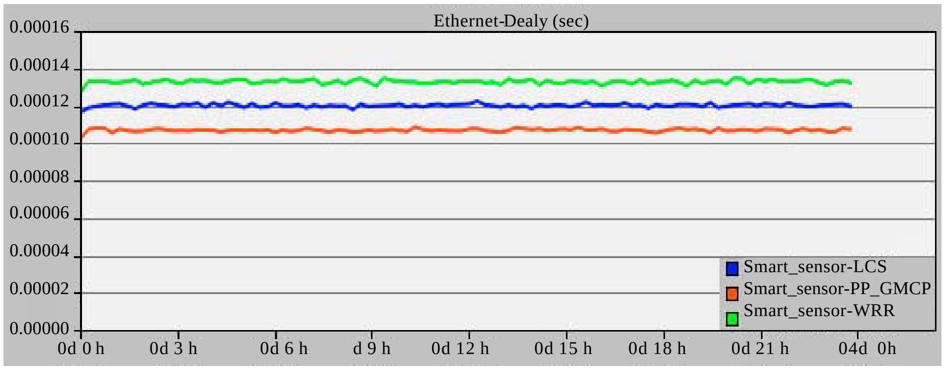

Fig. 7: Different scheduling algorithms average service response delay time

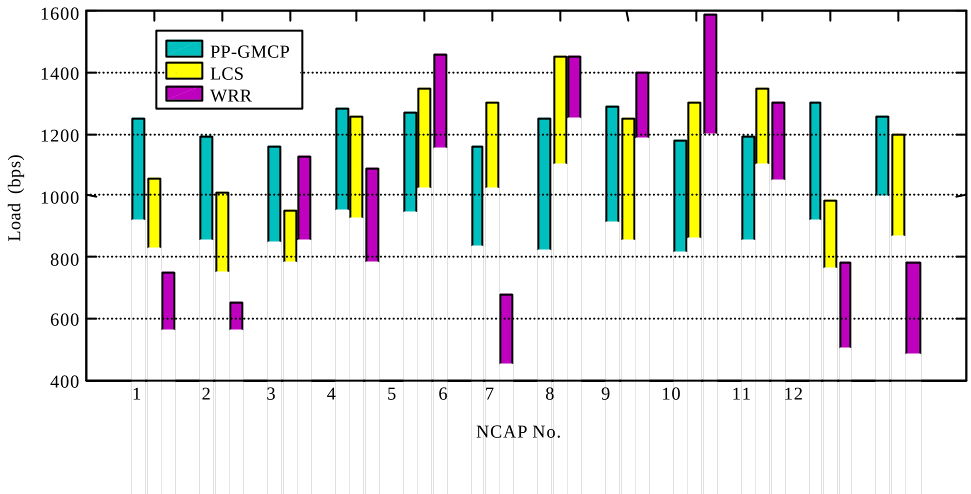

Fig. 8: NCAPs work load variation ranges in 24 h

Scheduling (LCS). Simulation region is configured to 500×500 m campus level, simulation time is 24 h, the simulation seed number is 128 and service request packet size and time interval obey to the exponential distribution.

Different scheduling algorithms average service response delay time curves are shown in Fig. 7. It can be seen, the average service response delay time of PP-GMCP algorithm is the minimum and reduces 11.1 and 25.1% respectively compared with LCS and WRR algorithm.

All NCAPs work load variation ranges in 24 h are shown in Fig. 8. It shows that NCAP work load variation range of PP-GMCP algorithm is the minimum and WRR algorithm is the maximum. NCAP actual work load range: PP-GMCP is 815~1300bps, WRR is 450~1585bps, LCS is 750~1450bps. According to NCAP current load condition, considering the resources influence of different service request, PP-GMCP distributes service requests to different NCAP to realize parallel processing which enabling the network load can be evenly distributed to each NCAP and achieves smaller load fluctuation range and better utilization.

CONCLUSION

Based on Markov chain residual correction prediction method, this paper effectively predicts the NCAP load capacity by real-time monitoring NCAP load status and the prediction performance indicators are superior to GM(1,1) grey prediction, Kalman filtering and parameter correction prediction method. HPPP strategy is adopted to realize service scheduling based on service queue access priority and MLPP strategy is used to complete service distribution based on the predicted NCAP load capacity. Using OPNET simulation tools, a network intelligent sensor system load balance simulation platform is constructed. The simulation test shows, synthetically considering different service request influence on NCAP resources, the PP-GMCP algorithm can evenly distribute service requests to different NCAP.

Compared with LCS and WRR algorithm, the PP-GMCP algorithm acquires smaller load fluctuation range and better load balance effect.

ACKNOWLEDGMENT

This work is supported by the Science and Technology Support Project of Guangzhou city (2009Z2-D531) and the Research Fund for Doctoral Program of Hubei University of Arts and Science (2013B005).

REFERENCES

Devarasiddappa, D., M. Chandrasekaran and A. Mandal, 2012. Artificial neural network modeling for predicting surface roughness in end milling of Al-SiCp metal matrix composites and its evaluation. J. Applied Sci., 12: 955-962.

Eludiora, S., O. Abiona, G. Aderounmu, A. Oluwatope, C. Onime and L. Kehinde, 2010. A load balancing policy for distributed web service. Int. J. Commun. Networks Syst. Sci., 3: 645-654.

Huang, Y.H., J.C. Peng and C.C. Li, 2011. Application of Markov theory in mid-long term load forecasting. Proc. CSU-EPSA, 23: 131-136.

Ma, X.M., 2005. Loading prediction of host computer in distributed system. Master's Thesis, Jilin University, Changchun, China.

Samsudin, R., A. Shabri and P. Saad, 2010. A comparison of time series forecasting using support vector machine and artificial neural network model. J. Applied Sci., 10: 950-958.

Sharifian, S., S.A. Motamedi and M.K. Akbari, 2009. Estimation-based load-balancing with admission control for cluster web servers. ETRI J., 31: 173-181.

Tan, M., H. Xu, L. Zeng and S. Xia, 2011. Research on fuzzy self-adaptive variable-weight combination prediction model for IP network traffic. Inform. Technol. J., 10: 2322-2328.

Wan, X.G. and L.L. Dong, 2011. Designing reverse proxy-based single sign-on system. Comput. Appl. Software, 28: 156-158.

Wang, X. and W. Fu, 2011. Research and implementation of the multiple exit load balancing of campus network based on DNS service. Proceedings of the IEEE International Conference on Computer Science and Automation Engineering, June 10-12, 2011, Shanghai China, pp: 471-474.

Wobschall, D., 2008. Networked sensor monitoring using the universal IEEE 1451 standard. IEEE Instrum. Meas. Mag., 11: 18-22.

Zhang, M.J. and Y.Q. Hu, 2011. OPNET-based switching node simulation model performance analysis. Comput. Appl. Software, 28: 225-227.

Zheng, Q., 2009. Research on load balancing algorithm of web server based on cluster. J. Zhejiang Univ. Sci. Technol., 21: 15-18.