Proactive Fault Detection Schema for Enterprise Information System Using Statistical Process Control

ChiHoon Lee[1], Doohyung Lee[1], Jahwan Koo[2], and Jinwook Chung[1]

[1] School of Information and Communication Engineering, Sungkyunkwan University, Chunchun-dong 300, Jangan-gu, Suwon 440-746, South Korea
chlee10@naver.com, drlee07@skku.edu, jwchung@songgang.skku.ac.kr
[2] Computer Sciences Department, University of Wisconsin-Madison, 1210 W. Dayton Street, Madison, WI 53706, USA
jhkoo@cs.wisc.edu

Abstract. This paper proposes a proactive fault detection schema using adaptive statistical approaches in order to enhance system availability and reliability in the heterogeneous & complicated information system environment. The proposed system applies Six Sigma SPC (Statistical Process Control) techniques already validated in industries in order to monitor the application system in the information system. This makes it possible to reduce false alarm rates for system faults and accurately detect faults by creating a control chart based on past performance data and controlling the distribution of performance based on the chart. The early detection of faults is also enabled through a fault prediction model. Therefore, the aforementioned system not only detect unknown or unseen faults but also resolve potential problems for system administrator by detecting abnormal behaviors before faults occur. In the experiment we show the superiority of our proposed model and the possibility to early detect system faults.

Keywords: System management, Proactive Fault detection, Statistical Process Control, Early Detection, EWMA.

1 Introduction

An explosive rise in Internet penetration has caused a massive traffic overload and required computer performance and capacity to be upgraded. Under these circumstances, the key challenge in service continuity is to ensure the availability and reliability of the information system. Such an issue can be resolved to a certain degree by preventing single-point-failures through system dualization and clustering, but which is accompanied by heavy costs. Therefore, the accurate and early detection of system faults must be guaranteed in order to achieve such a goal with limited physical and financial resources and to avoid potential serious issues.[1][2][3]

Fault management, a key area in system administration, is designed to early identify and detect faults, which has continuously attracted more attention. The information system, which most organizations heavily depend on, has become core infrastructure in an enterprise environment. The downtime caused from the system

M.J. Smith and G. Salvendy (Eds.): Human Interface, Part I, HCII 2009, LNCS 5617, pp. 113–122, 2009.

faults force companies to pay significant costs. Furthermore, a diversity of challenges should be handled in order to overcome such fault issues. First of all, the hardware or software of various vendors should be organically linked and operated in an open environment. At the same time it should meet user needs and up-to-date technology requirements and then it makes increasing fault risks [4]. Furthermore, continuous application changes and improvements to satisfy user requirements make it harder to control faults.

The biggest issue in fault detection is the fact that a fixed threshold method is used, based on the experiences of system administrators and experts. This makes it difficult to flexibly respond to changes and the extension of the information system. If a threshold value is set too high, it is not easy to detect faults. If it is too low, the system may suffer from frequent false alarms. Even though the threshold is properly set, lack of prior information makes it hard to early detect the symptoms of faults. [1].

The existing research on fault management focused on a rule-based expert system [5], finite state machines [6], a statistical model [7], and a data mining model [8]. The rule-based expert system requires the specifications of potential faults [1]. This is exposed to performance restrictions because all possible faults can not be handled, which is also vulnerable to unknown or novel types of faults and changes in the system environment. The statistical model can resolve part of the aforementioned issues but hardly forecast faults in advance. Lastly, the data-mining model can detect faults in advance to a certain degree but can not be easily applied in real time. This paper proposes a new proactive fault model that can accurately detect and forecast information system faults. This model is designed to 1) minimize FAR, 2) detect changes in distribution/average to forecast information system faults, 3) and detect in real time under the online environment by applying an adaptive threshold method based on the SPC validated in manufacturing processes. The purpose of this paper is to substantially apply the model to the management of enterprise information system faults.

In this paper Section 2 deals with system management overview and the background of SPC and EWMA. Section 3 describes proposed schema. Section 4 gives experimental results. Finally section 5 mentions conclusion.

2 Background

2.1 System Management

These days the reliance of the information system in the enterprises has increased and also been reinforced to ensure business continuity. The information system can stably support business activities only when it is harmonically activated, based on servers, a network, a database, and applications. However, the sudden deactivation of all or parts of such components can do great damage to the enterprise, with direct effects on the survival thereof. Up to now, staffs by area have supported stable operation by monitoring their partial field from time to time. However, in the current information system environment characterized by a rapidly rising number of servers & database and more complicated networks & applications, such monitoring by area has to be exposed to its own limitations. The System Management System was introduced to overcome such an obstacle. It is designed to ensure service continuity by immediately

notifying the system administrator of the occurrence of faults in real-time through a sophisticated agent and helping them quickly take corrective actions. A general SMS consists of managers and agents. An agent installed on a monitoring target gathers data (faults or performance items) in real-time, transferring them to Manager. Then, Manager gathers and analyzes data before reporting faults to system, network, DB, and application administrators.

The detected faults represent the state of real faults, the notified operator should promptly take corrective actions to fix faults. However, because such signals are detected after the fault occurs, the system is already in the state of fault at the point when the operator recognizes it, causing the business to stop. Therefore, the system operator is in dire need of a system that can monitor abnormal symptoms and prevent faults in advance.

2.2 Statistical Method-Based Fault Detection

SPC (Statistical Process Control). The SPC is designed to identify, interpret, and resolve problems based on accumulated statistic data, rather than intuition or guessing. In other words, it scientifically analyzes basic data to manufacture that meet quality requirements. The key to the SPC is to produce uniform quality goods featuring little quality dispersion. The main purpose of the SPC is to reduce process variability and enhance processes in order to manufacture higher-quality products. The SPC is currently applied to a wide range of industries including manufacturing, to control manufacturing facilities, curtail logistics costs, and improve software quality. [9][10] The control chart devised by Dr. Walter Shewhart [13] is used to detect process changes from special causes through the graph that can monitor data variability over time. It can be classified into variable control charts for continuous data and attribute control charts for discrete data. Specific examples can be illustrated as follows [15]:

Variable Control Chart: $\overline{X}$ - R, $\overline{X}$ - S, I&MR chart

Attribute Control Chart: P, NP, C, and U chart

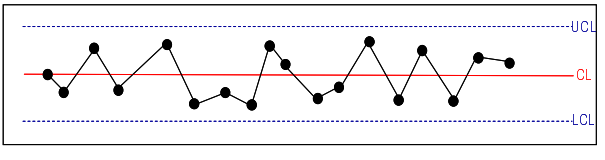

Fig. 1. General form of the control chart

Fig. 1 represents the average of quality characteristics around the Center Line. The UCL (Upper Control Limit) and the LCL (Lower Control Limit) are calculated by adding or deducting "each quality characteristic SD(Standard Deviation) multiplied by k" to or from the mean (average) of the quality characteristics. If the values calculated through data are scattered on the graph, all the points are placed between the control limits. Unless special factors are identified, the process is judged to be in a control or steady state. This paper applied the SPC in order to resolve fixed threshold issues.

EWMA (Exponentially Weighted Moving Average). The Exponentially Weighted Moving Average (EWMA) [12] is a statistic for monitoring the process that averages

the data in a way that gives less and less weight to data as they are further removed in time. For the Shewhart chart control technique [13], the decision regarding the state of control of the process at any time, t, depends solely on the most recent measurement from the process and, of course, the degree of 'trueness' of the estimates of the control limits from historical data. For the EWMA control technique, the decision depends on the EWMA statistic, which is an exponentially weighted average of all prior data, including the most recent measurement. By the choice of weighting factor, λ , the EWMA control procedure can be made sensitive to a small or gradual drift in the process, whereas the Shewhart control procedure can only react when the last data point is outside a control limit.

As with all control procedures, the EWMA procedure depends on a database of measurements that are truly representative of the process. Once the mean value and standard deviation have been calculated from this database, the process can enter the monitoring stage, provided the process was in control when the data were collected. This paper used the EWMA to forecast faults in advance.

3 Proposed Scheme

3.1 Limitations of the Existing SMS

The biggest issue facing the SMS is to properly set the threshold. This requires expert knowledge of system operation, accompanied by trials and errors. For example, if a threshold value is set not to exceed in any environment, faults can not be detected. In case it is set too low, system administrators will be flooded with false alarms. It will also be impossible to flexibly respond to changes in the system environment such as system introduction & enlargement and application addition. Even if a performance value exists within the threshold range, abrupt variances in the range can be thought of as the symptoms of the faults. However, the current SMS regards them as normal.

3.2 Proposed Fault Detection System Model

This paper proposes two types of fault detection models in order to overcome the limitations and errors of the existing fault detection method described in section 3.1. One is an adaptive threshold identification model to which the SPC applies. The other is a fault forecasting model to which the EWMA applies.

SPC-based Adaptive Threshold Identification. After analyzing performance data for a certain period of time by applying the SPC methodology, an optimized threshold is automatically identified, based on which the state of performance data gathered in real-time can be assessed. To this end, the following procedures are implemented:

(1) Select a proper control chart after choosing performance items to monitor. Most of the performance data gathered in management are considered to be variable ones. A proper control chart (Xbar-R chart, Xbar-S chart, I&MR chart, etc.) should be selected, depending on the quantity of collected data and the collection interval. In case the collection cycle is short and a large quantity of data is gathered, the following choice of the control chart is made by subgroup size. If the subgroup size is set to be 10 or larger, the Xbar-S chart is applied. If the size is smaller than 10, the Xbar-R

chart is used. If the data collection cycle is long and a small quantity of data are gathered, the I&MR chart will be a proper choice.

(2) In case a control chart and a subgroup size are chosen, an automatic control limit renewal cycle is set. The cycle is set, considering changes in system capacity and the quantity of data.

(3) Based on the selected control chart and the automatic control limit renewal cycle, a Daemon program which automatically generates a control limit at a specific time every day is implemented in order to generate the control chart, which was set after reading gathered performance data by item. As a result, an optimized threshold (control limit value) for the performance item can be obtained.

(4) The SMS transmits performance date collected in real-time to a SPC processing module in real time. The SPC processing module conducts calculations in accordance with the formula of the control chart designated in (1), after the size of real-time data reaches the level that was set in (1). In case the result is placed outside the control limit identified in (3), an event is generated. The resulting event is not influenced by performance data from instant rapid variances (too large or small value), reducing FAR and enabling accurate & reliable fault detection. In addition, the value is automatically calculated and renewed, depending on the control limit setting cycle, making it possible to respond to changes in the system environment and produce an adaptive threshold.

Early Fault Detection by EWMA. The EWMA model is applied in order to analyze data that are not in control or show big variances over time. Precisely this model can forecast subtle variances in the control limit to detect faults in advance. At the current point (t), variances (r_t) in performance data are calculated as follows:

$$r_t = \ln(p_t / p_{t-1}) \tag{1}$$

where P_t is the performance data at the point of t and ln() is the natural logarithm. Variability in the change rate can be calculated by using the sample standard deviation. The standard deviation (δ) for the sample size 'n' is calculated as follows:

$$\delta = \sqrt{\sum_{t=1}^{n} (r_t - \bar{r})^2 / n} \tag{2}$$

Above formula shows that the sum of the square of deviations from the mean, (r), of performance data change rates is divided by the sample size. It can also be regarded as averaging 'the square of deviations from the mean' multiplied by a certain weight (1/n). The variability can also be estimated by using the EWMA. This is is represented as follows:

$$\delta = \sqrt{(1-\lambda)\sum_{t=1}^{n} \lambda^{t-1} (r_t - \bar{r})^2} \tag{3}$$

Where λ is a decay factor. Because the variability estimated by the EWMA has the following recursive characteristic, it is easier to predict variability:

Let $\delta^2_{t+1|t}$ be $(1-\lambda)\sum_{i=0}^{\infty} \lambda^i r^2_{t-i}$

Then we get equation (4)

$$\begin{aligned}\delta^2_{t+1|t} &= (1-\lambda)\sum_{i=0}^{\infty}\lambda^i r^2_{t-i} \\ &= (1-\lambda)(r_t^2 + \lambda r^2_{t-1} + \lambda^2 r^2_{t-2} + \cdots) \\ &= (1-\lambda)r_t^2 + \lambda(1-\lambda)(r^2_{t-1} + \lambda r^2_{t-2} + \lambda^2 r^2_{t-3} + \cdots) \\ &= \lambda\delta^2_{t|t-1} + (1-\lambda)r_t^2\end{aligned} \tag{4}$$

Hence $t+1|t$ means predicting $t+1$ data by using t-point performance data.

In the above equation, the $t+1$ variance can be represented through the linear equation of the pervious change rate. Therefore, if the t point variance is identified, the $t+1$ point variability can also be repeatedly estimated.

If the sigma level (s; s= 1, ... , 6) is applied to the above formula, the UCL and LCL at the point t can be calculated. When t point data are obtained, we can check whether they are outside the t point UCL and LCL or not in order to measure the reliability of performance data.

When $t-1$ point performance data are obtained, we will make a judgment on whether they are within the range of UCL and LCL, based on obtained t point performance data(P_t). If they are outside the range, possible faults can be predicted in advance by raising an alarm. The concept of fault tolerance can be additionally applied in order to reduce FAR.

3.3 Implementation

The proposed system is mainly composed of an SPC server, an SPC generator, an SPC console, and API modules.

The SPC generator reads monitoring item once a day (the minimum cycle for automatic control limit renewal is one day). Based on this, both UCL and LCL are calculated and saved in the SPC DB. A user selects monitoring targets by using an SPC console before entering details for monitoring. He or she should make a decision on whether to use the SPC or EWMA. If the SPC model is selected, a control chart type, a data collection cycle, an automatic control limit renewal cycle and a minimum subgroup size are set and then saved in the SPC DB. If the EWMA is chosen, the SPC generator is not activated. The reason is that the EWMA calculates the next performance data estimate in real-time but the SPC needs calculated UCL and LCL to analyze current data.

The SPC server conducts a real-time analysis in accordance with the method (SPC or EWMA) set by monitoring target, immediately after receiving performance data through SPC API. In case of SPC analysis, the mean and standard deviation of the subgroup is calculated by referring to UCL and LCL pre-calculated by the SPC generator at the point when real-time performance data corresponding to the size of the subgroup are gathered. Then, a judgment is made on whether the value is outside UCL and LCL by control chart before generating fault events on a need-to-do basis. The subgroup is calculated in accordance with the sliding window method. For example, if the subgroup size is set to 7, the first subgroup is (p1, p2, p3, p4, p5, p6, p7). The subgroup at the point of collecting the 8th data is shifted to (p2, p3, p4, p5, p6, p7, p8,). As the subgroup is sequentially shifted in this way, the mean and standard deviation thereof are calculated in real-time in order to check whether they are within

the range of UCL and LCL. As for EWMA analysis, future point (Pt+1) performance data are forecast at the point when current point(Pt) performance data are collected. Therefore, UCL(t+1), LCL(t+1) values are calculated by using the change rate (rt) in order to check whether Pt+1 performance data are within the range of UCL and LCL. Only if they are outside the range, a fault event is generated.

A user selects monitoring targets and an analysis method (SPC or EWMA) through the SPC console. Also the SPC console makes it possible to query events generated from the SPC server's real-time calculation and analysis.

4 Experimental Results

The experiment was conducted for the following two purposes: One is to validate the effects of the fixed threshold and the SPC-based adaptive threshold. The other is to check whether to forecast faults in advance. A separate experimental environment was set up to generate faults, through which data were collected. However, the reliability of the deliverables may not be high enough, considering the fact that it is not easy to gather enough real world fault data. The web page response time was used for SPC validation while server CPU utilization data were applied to the EWMA analysis.

4.1 Results from SPC-Based Adaptive Threshold Model

The SPC model was applied to the web page response time. The Xbar-R chart was selected as a control chart, setting the subgroup size to 5 and designating the auto control limit renewal interval as one day. As a result, the control limit is automatically renewed on a daily basis. The Anderson-Darling test [14] was used as target data normality testing. The normality of the data was validated, as the p-value is larger than 0.05.

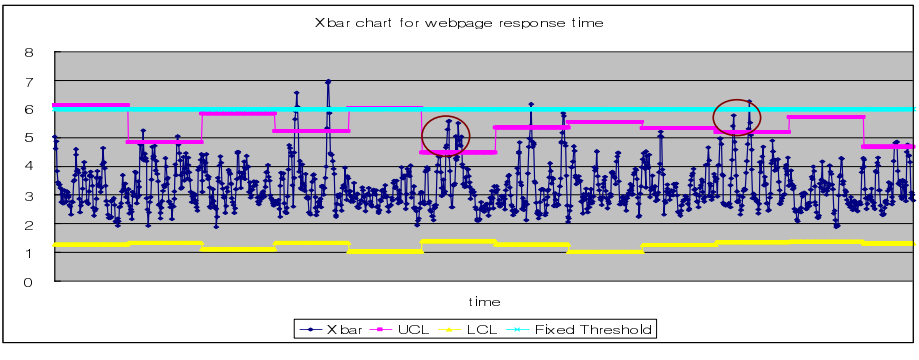

Fig. 2. Xbar chart for webpage response time

Fig. 2 shows that the control limit is renewed on a daily basis, resulting in the application of the adaptive threshold. The proposed model identified the points in the ellipse not detected by the fixed threshold as abnormal behaviors. This implies the feasibility of monitoring optimized to the system environment. Namely, in case envi-

ronmental changes occur such as set based on past data, flexibly responding to endless changes in the system environment.

4.2 Results from EWMA-Based Early Fault Detection

CPU utilization data for three days including the day faults occurred were collected by using experimental devices. During the first and third days, no fault was detected but on the second day, real faults were sensed. The data were analyzed by using the EWMA model. The λ (decay factor) and s (Sigma Level) were set 0.85 and 3, respectively.

Fig. 3(a) shows the results from EWMA analysis of the first day performance data. Around 08:30, a false alarm was raised, even though a real fault did not occur. Steep changes in performance data at the point of the false alarm clearly suggest server problems, even if it has yet to reach the level of a fault. Furthermore, taking into account that the false alarm was not frequently given, the system administrator needs to be pre-cautions against the situation.

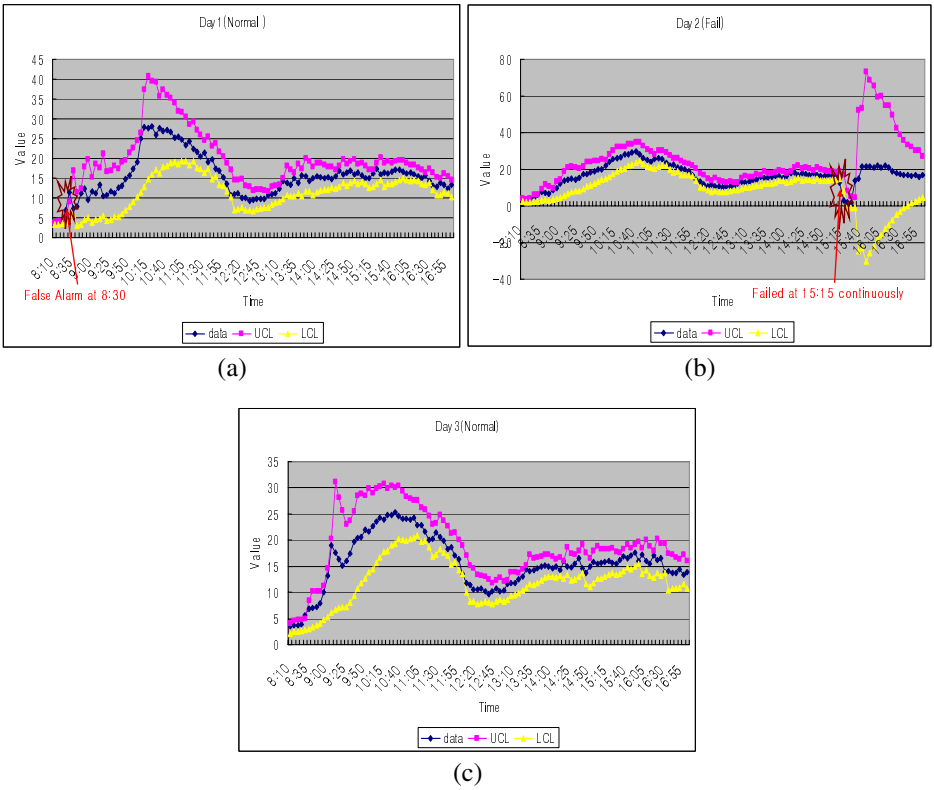

Fig. 3. CPU Utilization's performance data of specific server

As shown in Fig. 3(b) the Online Daemon Process hanged up around 15:15 on the second day. Therefore, if the current state had continued, system-down could not have

been avoided. We reasoned the Daemon Process Hang Up based on variability in CPU utilization value, rather than monitoring the process. As a result, the system administrator was successfully alerted to abnormal behaviors, solving serious potential issues by restarting the Daemon Process. As displayed in Fig. 3(c), it was normally processed without any false alarm on Day 3. This shows that the proposed model is stably activated.

5 Conclusion

Fault management plays a pivotal role in ensuring service continuity in the information system environment. This research presented two types of fault detection models. One is the SPC-based adaptive threshold model designed to resolve fixed threshold issues. This was very effective in accurately detecting potential faults and reducing FAR. The other is the EWMA-based early fault detection model. This enables system administrators to avoid serious potential issues by detecting abnormal behaviors before real faults occur.

This research can apply to the information system in a manufacturing or financial environment. The SPC model is suited to manufacturing processes featuring less variability in performance data while the EWMA paradigm is more effectively applied to the financial sector characterized by larger variability in performance data. The proposed models can accurately detect and estimate system/application faults and minimize downtime, enhancing the yield in manufacturing processes and system reliability in the financial arena. The follow-up research will focus on not only detecting faults but also statistically inferring their root causes in order to make it possible to both detect and fix faults

References

1. Hood, C.S., Ji, C.: Proactive Network-Fault Detection. IEEE Transactions on reliability 46(3), 333–341 (1997)
2. Hellerstein, J.L., Zhang, F., Shahabuddin, P.: An approach to predictive detection for service management. In: Sloman, M., Mazumdar, S., Lupu, E. (eds.) Proc. 6th IFIP/IEEE Int. Symp. Integrated Network Management (IM 1999), p. 309. IEEE Publishing, New York (1999)
3. Thottan, M., Ji, C.: Fault prediction at the network layer using intelligent agents. In: Sloman, M., Mazumdar, S., Lupu, E. (eds.) Proc. 6th IFIP/IEEE Int. Symp. Integrated Network Management (IM 1999), p. 745. IEEE Press, New York (1999)
4. Yemini, Y.: A critical survey of network management protocol standards. In: Aidarous, S., Plevyak, T. (eds.) Telecommunications Network Management into the 21st Century (1994)
5. Jakobson, G., Weissman, M.D.: Alarm correlation. IEEE Network 7, 52–59 (1993)
6. Rouvellou, I.: Graph identification techniques applied to network management faults, Ph. D Dissertation. Columbia University (1993)
7. Deng, R.H., Lazar, A.A., Wang, W.: A probabilistic approach to fault diagnosis in linear lightwave networks. IEEE J. Selected Areas in Communications 11, 1438–1448 (1993)

8. Garofalakis, M., Rastogi, R.: Data Mining Meets Network Management: The Nemesis Project. In: ACM SIGMOD Int'l Workshop on Research Issues in Data Mining and Knowledge Discovery (May 2001)
9. Florence, A.W.: The MITRE Corporation. CMM Level 4 Quantitative Analysis and Defect Prevention with Project Examples, 2000 Technical Papers (September 2000)
10. Radice, R.: Statistical Process Control for Software Projects (November 1997)
11. ERETEC INC., MINITAB Release 14 (November 2005)
12. NIST/SEMATECH e-Handbook of Statistical Methods: EWMA Control Charts
13. Shewhart, W.A.: Statistical Method from the Viewpoint of Quality Control (1939)
14. Anderson, T.W., Darling, D.A.: Asymptotic theory of certain "goodness-of-fit" criteria based on stochastic processes. Annals of Mathematical Statistics 23, 193–212 (1952)
15. Oakland, J.: Statistical Process Control (2002)