11-10-2011

Statistical Process Control for the Fairness of Network Resource Distribution

Qingyun Liu
Florida International University, qliu002@fiu.edu

FLORIDA INTERNATIONAL UNIVERSITY

Miami, Florida

STATISTICAL PROCESS CONTROL FOR THE FAIRNESS OF NETWORK RESOURCE DISTRIBUTION

A thesis submitted in partial fulfillment of the

requirements for the degree of

MASTER OF SCIENCE

in

STATISTICS

by

Qingyun Liu

2011

To: Dean Kenneth G. Furton
College of Arts and Sciences

This thesis, written by Qingyun Liu, and entitled Statistical Process Control for the Fairness of Network Resource Distribution, having been approved in respect to style and intellectual content, is referred to you for judgment.

We have read this thesis and recommend that it be approved.

____
Gauri Ghai

____
Tao Li

____
Zhenmin Chen, Major Professor

Date of Defense: November 10, 2011

The thesis of Qingyun Liu is approved.

____
Dean Kenneth G. Furton
College of Arts and Sciences

____
Dean Lakshmi N. Reddi
University Graduate School

Florida International University, 2011

DEDICATION

I dedicate this thesis to my family. Without their support, patience, understanding and most of all their love, the completion of this work would not have been possible.

ACKNOWLEDGMENTS

First of all, the author would like to express his sincere gratitude to his major professor, Dr. Zhenmin Chen, for his patient guidance, encouragement and friendship throughout this entire study. Special thanks also go to the members of the author's master committee, Dr. Gauri Ghai and Dr. Tao Li, for their valuable advice that has enhanced the quality of this thesis. Thanks also go to the Department of Mathematics and Statistics at Florida International University for providing the author with support over the entire study.

The author would like to express his great thanks to his parents, and his sincere love to his wife and daughters. He could not have finished this program without their love, encouragement, patience, and support.

ABSTRACT OF THE THESIS

STATISTICAL PROCESS CONTROL FOR THE FAIRNESS OF NETWORK RESOURCE DISTRIBUTION

by

Qingyun Liu

Florida International University, 2011

Miami, Florida

Professor Zhenmin Chen, Major Professor

The purpose of this research is to develop a statistical method to monitor the fairness of network resource distribution. The newly developed fairness score function allows users to have the same or different priority levels. Especially, this function possesses all the necessary properties required as a quality characteristic for the purpose of statistical process control.

The main objective is to find the critical values for the statistical test. Monte Carlo simulation is used to find the critical values. When the users have the same priority level, a table of the critical values is given for different sample sizes and different significance levels. When the users have different priority levels, it is difficult to generate a similar table since the users' priority levels vary. Therefore, the critical values are computed for given priority levels. In both cases, an example is given to demonstrate the approach developed in this study.

TABLE OF CONTENTS

1. INTRODUCTION

1.1 Problem Description

Generally, computer network is a system that links various devices together with hardware and software to support data communications across these devices which are called hosts or end systems in Internet jargon [1]. The end system devices include primarily traditional desktop PCs, Linux workstations, and servers that store and transmit information such as web pages and email messages, increasingly, nontraditional devices such as laptops, cell phones, TVs, gaming consoles, etc. The end systems are connected by a network of communication links and packet switches. The communication links are made up of different types of physical media such as fiber optics, copper wire, coaxial cable and radio spectrum. Different type of links can transmit data at different transmit rates measured in bits/second. The sending end system segments the data and attaches header bytes to each segment. The resulting packets of information are then sent through the network to the destination end system, where the packets are reassembled into the original data. The packet switch plays a role of taking a packet of data from the incoming communication link and forwarding that packet to the outgoing communication link. The end systems and packet switches follow the internet standards which are called protocols to control the sending and receiving of information within Internet network.

As described above, from a physical component view of point, a computer network is made of various hardware and software components. However, the computer network can be view as an infrastructure that provides services to applications. These applications include email, file transfer, web surfing, instant messaging, Voice-over-IP

(VoIP), internet radio, video streaming, teleconferencing, interactive games, television over the Internet, and much more. Out of all these applications, the data-orientated applications such as email, file transfer, web text/image browsing, are not sensitive to the time delay. However, the new multimedia networking applications, such as multimedia www websites, VoIP telephony, teleconferencing, interactive games, are highly sensitive to time delay and delay variation. There even exist some real-time online application scenarios, for example, online stock trading, large-scale distributed real-time games, real-time online auctions, etc. For these applications, the network is required not only to provide the communication but also to guarantee each user receives appropriate share of the resource as his/her competitors. Therefore, the fairness of the resource distribution is needed to provide necessary service quality to the applications. Monitoring the fairness of network resource distribution is one of the important issues of computer network management. The purpose of this research is to develop a statistical analysis method to monitor the fairness of network resource distribution.

1.2 Introduction to Statistical Process Control

Generally speaking, statistical process control (SPC) involves applying statistical methods to the monitoring of a process to keep the process under control as designed and intended through measuring, analyzing and reducing the process variability [2][3][4]. A typical tool in SPC is the control chart. A control chart is a graphical plot of a quality characteristic that has been measured or computed from a sample versus the sample number or time. A quality characteristic that can be measured in numerical scale is called a variable. The corresponding control chart is called a variable control chart. If the quality characteristic cannot be expressed as a numerical data but can only be classified as conforming or nonconforming, the control chart is called attribute control chart.

A typical control chart, as displayed in Figure 1, has a center line that represents the average value of the quality characteristic and two other lines, which are called upper control limit and lower control limit. The comparison between the quality characteristic and the control limits detects any unusual variation and, therefore, determines whether or not the process is under control. As long as the points plot randomly within the control limits, the process is said to be under control, and no action is required. Otherwise, the process is interpreted as out-of-control. Investigation and corrective action are required to find the causes responsible for this behavior.

It should be mentioned that the quality characteristics do not have to be controlled by both control limits. For example, the fairness of network resource distribution should be monitored so that the distribution is fair enough. In other words, actions are needed only when the distribution is significantly unfair. Nobody needs to worry about the case that the distribution is too fair. Therefore, only one-sided control is needed for this case.

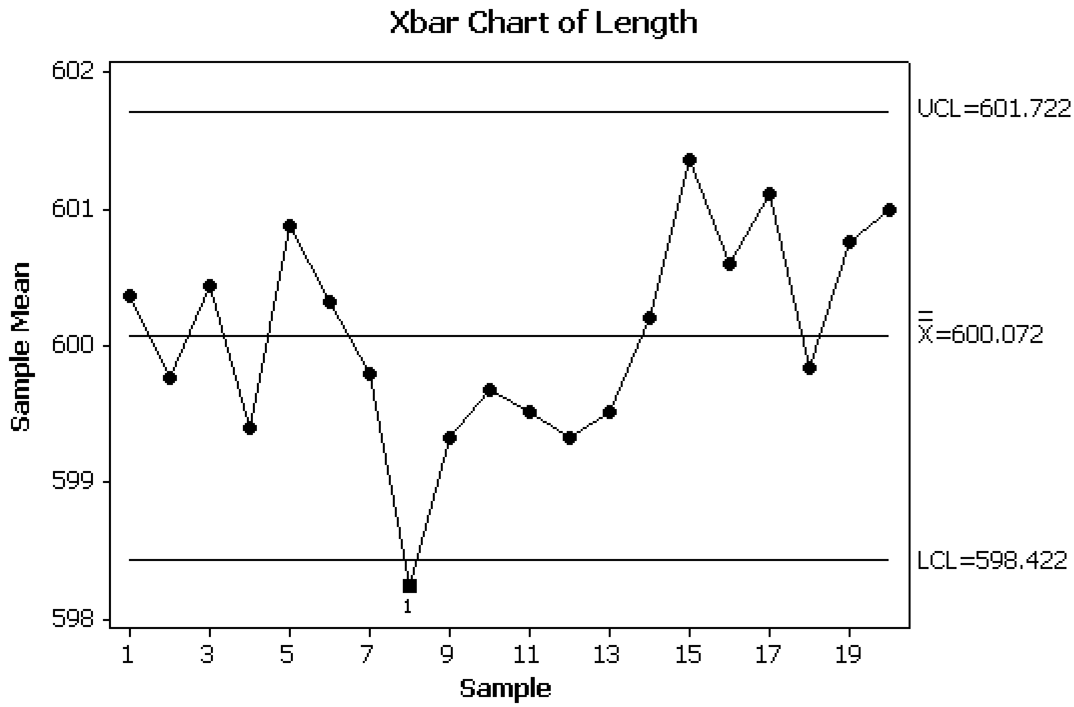

Figure 1 An example of control chart

The control chart, in some sense, tests the hypothesis that the process is under the statistical control, repeatedly at different points in time. The quality characteristic plays a role as the test statistic. Setting up control limits is equivalent to setting up critical regions. Thus, under-control is equivalent to failing to reject the hypothesis; and out-of-control is equivalent to rejecting the hypothesis.

The choice of control limits is critical in designing the control chart. By moving the control limits farther from the center line, the risk of type I error decreases and the risk of type II error increases. Type I error is the error of diagnosing the process as out-of-control while the process is actually under control. Type II error is the error of failing to diagnose the process out-of-control while the process is, in fact, out-of-control.

However, moving the control limits closer to the center line will have the opposite effect. That is, the risk of type I error increases, while the risk of type II error decreases. A general model for control chart was first proposed by Walter A. Shewhart in the early 1920's. For example, let x be a sample statistic that measures the quality characteristic of interest, and suppose that the mean of x is μ_x and the standard deviation of x is σ_x. Then the center line and the upper and the lower limits become $\text{center line} = \mu_x$, $\text{upper limit} = \mu_x + L\sigma_x$, and $\text{lower limit} = \mu_x - L\sigma_x$, where L is the distance of the control limits from the center line expressed in standard deviation units. In a typical Shewhart control chart, L is set to be three. If the quality characteristic is assumed normally distributed then the type I error is only 0.0026. A three-sigma control limit is commonly employed if the statistical distribution of the quality characteristic can be reasonably approximated by the normal distribution.

1.3 Challenge of this Research

To use SPC to monitor the fairness of network resource distribution, an appropriate quality characteristic is needed to evaluate the fairness of the network resource distribution. The control limits need to be computed appropriately. The selected quality characteristic must have some desirable properties. It must continuously reflect changes in the network resource allocation from the completely unfair case to the perfectly fair case. The value of the quality characteristic must monotonically increase or decrease when the network resource distribution becomes fairer or less fair. The value of the quality characteristic should not depend on scale. Furthermore, the control limits must be simulated when the statistical distribution of the selected quality characteristic is not known well enough.

1.4 Previous Research

The concept of fairness in computer networks was introduced by Jain, Chiu and Hawe [5]. A comprehensive review of the research of fairness in computer networks is given in [6]. The authors pointed out the research activities needed in the future investigations. Current research in the area of computer networking fairness mostly concerns the fairness of bandwidth sharing among competing users. Some published papers are Jain, Chiu and Hawe [5], Bertsekas and Gallager [7], Chiu and Jain [8], Kelly et al. [9], Mazumdar, Mason and Douligeris [10]. The quantitative fairness score function proposed by Jain, Chiu and Hawe [5] is widely adopted in network design and management. The fairness score function is defined as

$$F(x_1, x_2, \ldots, x_n) = \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n\left(\sum_{i=1}^{n} x_i^2\right)},$$

(1)

where n is the number of the users, and $x_1, x_2, \ldots, x_n$ are the amounts of network resource the users receive, respectively. This fairness score function $F(x_1, x_2, \ldots, x_n)$ possesses following prosperities.

(a) $0 \le F(x_1, x_2, \ldots, x_n) \le 1$ for any nonnegative $x_1, x_2, \ldots, x_n$.

(b) $F(x_1, x_2, \ldots, x_n) = 1/n$ when the network resource distribution is completely unfair, i.e., only one user occupies the entire network resource while the other users do not receive any.

(c) $F(x_1, x_2, \ldots, x_n) = k/n$ if only k out of n users share the entire network resource equally while the others do not receive any.

(d) $F(x_1, x_2, \ldots, x_n) = 1$ when the network resource distribution is perfectly fair, i.e., all the n users share the entire network resource equally.

(e) $F(x_1, x_2, \ldots, x_n)$ does not depend on scale.

(f) $F(x_1, x_2, \ldots, x_n)$ continuously reflects changes in network resource allocation.

Properties (a), (d), (e), and (f) are attractive to the researchers and users. However, the result that $F(x_1, x_2, \ldots, x_n) = 1/n$ for the completely unfair case does not fit the real situation well. In fact, if only one user occupies the entire network resource, the value of the function, $F(x, 0, \ldots, 0)$, should be zero, not $1/n$. The same thing happens to the k-out-of-n case. When $k = 1$, the same problem will occur because it is the completely unfair situation. The fairness score should be zero, not $1/n$.

In 2005, Chen and Zhang [11] proposed a fairness score function $G(x_1, x_2, \ldots, x_n)$ which keeps all the nice properties that Jain, Chiu and Hawse's fairness score function possesses. In addition, the proposed fairness score function has better performance in dealing with completely unfair cases. The fairness score function $G(x_1, x_2, \ldots, x_n)$ is defined as

$$G(x_1, x_2, \ldots, x_n) = 1 - \frac{n \sum_{i=1}^{n} \left(\frac{x_i}{\sum_{j=1}^{n} x_j} - \frac{1}{n} \right)^2}{n-1}. \tag{2}$$

For the case that the network resource distribution is completely unfair, it can be easily shown that $G(x_1, x_2, \ldots, x_n) = G(x, 0, \ldots, 0) = 0$.

For the case that only k out of n users share the entire network resource equally, it can also be shown that $G(x_1, x_2, \ldots, x_n) = \frac{n(k-1)}{k(n-1)}$. If $k = 1$, that is the case of completely unfair sharing. The fairness score for that case is zero.

The fairness score function $G(x_1, x_2, \ldots, x_n)$ assumes that all the users have the same priority level. However, equally distributing network resource to all the users is actually unfair if the users are at different priority levels. Instead, the system should distribute the network resource to the users proportionally according to their priority levels. For example, in a scenario where users pay different prices for their bandwidths, the weights in the fairness metric should be assigned in proportion to the bandwidth allocation. In 2010, Chen and Zhang proposed a modified version of their fairness score function $G(x_1, x_2, \ldots, x_n; w_1, w_2, \ldots, w_n)$ [12], which considers the case that different users may have different priority levels. It has been shown that the modified fairness score function keeps all the merits of the function $G(x_1, x_2, \ldots, x_n)$. Here $w_1, w_2, \ldots, w_n$ are the users' priority levels, respectively.

The current research is to adopt the modified fairness score function $G(x_1, x_2, \ldots, x_n; w_1, w_2, \ldots, w_n)$ as the test statistic to conduct statistical process control for the fairness of network resource distribution. The purpose is to check whether or not the network resource distribution is under fair control. It has been shown that the fairness score function $G(x_1, x_2, \ldots, x_n; w_1, w_2, \ldots, w_n)$ will monotonically increase or decrease when the network resource distribution becomes fairer or less fair. Furthermore, the function $G(x_1, x_2, \ldots, x_n; w_1, w_2, \ldots, w_n)$ does not depend on scale. Therefore, Monte Carlo simulation can be used to obtain the critical values of $G(x_1, x_2, \ldots, x_n; w_1, w_2, \ldots, w_n)$. When the observed value of $G(x_1, x_2, \ldots, x_n; w_1, w_2, \ldots, w_n)$ is lower than the critical value, it means that the network resource distribution is significantly unfair at certain level of significance.

1.5 Objective

The main objective of this research is to conduct statistical process control to monitor the fairness of network resource distribution. As mentioned above, a key step is to find the critical values for the statistical test. The following will be considered in this research:

1. For the case that users have the same priority level, a table of the critical values will be constructed for different sample sizes and different significance levels.

2. For the case that users have different priority levels, it is difficult to generate a similar table for the fairness score function $G(x_1, x_2, \ldots, x_n; w_1, w_2, \ldots, w_n)$ since the users' priority levels vary. Therefore the critical value will be computed dynamically for given priority levels.

3. For each of the above cases, an example will be presented to demonstrate this statistical analysis method.

1.6 Research Method

It is possible to find the critical values of a test statistic when the exact statistical distribution of the test statistic is known. However, the theoretical distribution of the fairness score function is not clear as of today. Therefore, Monte Carlo simulation is used in this study to find the critical values of the fairness score function at different significance levels. Monte Carlo simulation involves computer programming. The SAS system has a powerful variety of built-in statistical procedures (SAS/STAT). It also has versatile programming capability, especially, the interactive matrix language (SAS/IML) makes the SAS system ideal for conducting Monte Carlo simulation.

In the following sections, the fairness score function considering priority levels will be introduced in Section 2. General Monte Carlo simulation will be presented in Section 3 step by step. An example without considering the users' priority levels will be presented in Section 4. The effect of sample size will be discussed as well. An example considering the users' priority levels will be presented in Section 5. Finally, the discussion and conclusions will be presented in Section 6.

2. THE FAIRNESS SCORE FUNCTION CONSIDERING PRIORITY LEVELS

A new fairness score function $G(x_1, x_2, \ldots, x_n; w_1, w_2, \ldots, w_n)$ was developed by Chen and Zhang [12] as

$$G(x_1, x_2, \ldots, x_n; w_1, w_2, \ldots, w_n) = 1 - \frac{\left(\sum_{i=1}^{n} w_i\right)^2 \sum_{i=1}^{n} \left(\frac{x_i}{\sum_{j=1}^{n} x_j} - \frac{w_i}{\sum_{j=1}^{n} w_j} \right)^2}{\left(\sum_{i=2}^{n} w_{(i)}\right)^2 + \sum_{i=2}^{n} w_{(i)}^2},$$

(3)

where $x_1, x_2, \ldots, x_n$ are the amounts of network resource the users receive, $w_1, w_2, \ldots, w_n$ are the corresponding priority levels of the users, and $w_{(i)}$'s are the ordered priority levels, respectively.

In the case that all users are at the same priority level, the new fairness score function becomes the original one described in equation (2). Therefore, equation (3) keeps all the properties that equation (2) possesses when the users' priority levels are the same. The properties of the fairness score function defined in equation (3) are stated as follows.

Property 1. $0 \le G(x_1, x_2, \ldots, x_n; w_1, w_2, \ldots, w_n) \le 1$ for any $x_1 \ge 0, x_2 \ge 0, \ldots, x_n \ge 0$. It means that the fairness score is always between 0 and 1.

Property 2. $G(x_1, x_2, \ldots, x_n; w_1, w_2, \ldots, w_n) = 0$ if the resource distribution is completely unfair, which is the case that only the user with the lowest priority level occupies the entire network resource. From property 2, it implies that the future hypothesis test will be a lower tail test.

Property 3. $G(x_1, x_2, \ldots, x_n; w, w, \ldots, w) = \dfrac{n(k-1)}{k(n-1)} = \dfrac{1-1/k}{1-1/n}$ if all the n users are at the same priority level, and if only k out of the n users share the entire network resource equally while the other n-k users do not share any. The result is consistent with the fact that the network resource distribution will become fairer if more users share the entire resource.

Property 4. $G(x_1, x_2, \ldots, x_n; w_1, w_2, \ldots, w_n) = 1$ if the resource distribution is perfectly fair, which is the case that all the users share the entire network resource proportionally to their priority levels.

Property 5. For any $\eta > 0$, define

$$\begin{aligned} D(\eta) = {} & G(x_1, \ldots, x_s - \eta, \ldots, x_t + \eta, \ldots, x_n; w_1, w_2, \ldots, w_n) \\ & - G(x_1, x_2, \ldots, x_n; w_1, w_2, \ldots, w_n) \end{aligned}$$

(4)

and

$$\eta_0 = \left[\left(\frac{x_s}{\sum_{j=1}^{n} x_j} - \frac{w_s}{\sum_{j=1}^{n} w_j} \right) - \left(\frac{x_t}{\sum_{j=1}^{n} x_j} - \frac{w_t}{\sum_{j=1}^{n} w_j} \right) \right] \Big/ \sum_{j=1}^{n} x_j .$$

(5)

Then

$$D(\eta) \begin{cases} > 0 \;\; if \; \eta < \eta_0 \\ = 0 \;\; if \; \eta = \eta_0 \\ < 0 \;\; if \; \eta > \eta_0 . \end{cases}$$

(6)

It shows that if you take some amount of network resource from one user and give it to another user, the fairness score may remain the same, or become better or worse, depending on the amount you take. Therefore, the fairness score automatically adjusts accordingly.

Property 6. For any $\delta > 0$,

$$G(x_1 + \frac{\delta w_1}{\sum_{j=1}^{n} w_j}, x_2 + \frac{\delta w_2}{\sum_{j=1}^{n} w_j}, \ldots, x_n + \frac{\delta w_n}{\sum_{j=1}^{n} w_j}; w_1, w_2, \ldots, w_n)$$
$$\geq G(x_1, x_2, \ldots, x_n; w_1, w_2, \ldots, w_n).$$

(7)

It shows that if all the users are given extra amounts of network resource proportionally to their priority levels, then the fairness of the distribution will not decrease. Therefore, the fairness score has a monotonic property.

The properties described above show that $G(x_1, x_2, \ldots, x_n; w_1, w_2, \ldots, w_n)$ satisfies all the requirements mentioned in Section 1. The value of the new fairness score function continuously reflects changes in the network resource allocation from the completely unfair case to the perfectly fair case. It monotonically increases or decreases when the network resource distribution becomes fairer or less fair. Furthermore, the value of the new fairness score function does not depend on scale. The challenge faced right now is how to choose the control limits. From Property 2 and Property 4 described above, it is clear that only the lower control limit is needed. Also, because the true statistical distribution of the fairness score function is not clear as of today, the critical values are obtained using Monte Carlo simulation method as presented in the next section.

3. MONTE CARLO SIMULATION

Monte Carlo simulation is widely applied in a variety of disciplines. In statistics, typically, Monte Carlo simulation is applied in two situations [13]: assessing the consequence of assumption violation; and determining the sampling distribution of a statistic that has no theoretical distribution. In some situation, because of the complexity of a statistic, a theoretical sampling distribution of the statistic may not be available or difficult to obtain. Because a theoretical sampling distribution of the fairness score function is complex to develop, Monte Carlo simulation method is used for the current study. The sampling distribution of the fairness score function gives the variability of sample by sample, and the frequency of a sample occurrence by chance. Therefore, with knowing this, the observed fairness score can be judged whether or not it is an extreme value with some probability.

3.1 Sample Size and the Number of Replicates

The sample size is considered to be an important factor in Monte Carlo simulation. It is obvious that the variance of the sampling distribution is inversely proportional to the sample size. In general, larger samples will make it easier to detect small shifts in the process. In this study, the sample sizes will be set up from 5 to 50. Another factor which needs to be considered is the number of the pseudo samples drawn. Because the results obtained by Monte Carlo simulation is asymptotic to the real solution when the number of replicates goes to infinity. The accuracy can be improved with a large number of replicates. In this study, the number of replicates is set to be one million, which is considered to be large enough. The significance levels are set to be 0.1, 0.05, 0.025, and 0.01.

3.2 Implementation of Monte Carlo Simulation

Let $x_1, x_2, \ldots, x_n$ be the amounts of resource that the users receive respectively. Also let $w_1, w_2, \ldots, w_n$ be the corresponding priority levels of these users. Under the null hypothesis, which is the case that the network resource distribution process is under fair control, if the amount of resource which a basic user receives is x , then user i is supposed to receive $w_i x$ $(i = 1, 2, \cdots, n)$. A basic user is defined as a user with priority level of 1.

Define the amount of network resource that a basic user receives as a uniform random variable X . It means that a basic user will get amount of network resource between 0 and 1 with equal probability. Also define $Y_i = g(X) = w_i X \ (i = 1, 2, \cdots, n)$.

Because $X \sim U[0,1]$, $f_X(x) = 1,\ x \in [0,1]$. Then

$$f_{Y_i}(y) = f_X(g^{-1}(y)) \frac{d}{dy} g^{-1}(y) = \frac{1}{w_i},\ y \in [0, w_i]\ (i = 1, 2, \cdots, n) . \tag{8}$$

It means that Y_i is also a uniform random variable. $Y_i \sim U[0, w_i]\ (i = 1, 2, \cdots, n)$. Thus, a pseudorandom number representing the amount of network resource that a user with priority level w_i receives can be generated from a uniform distribution $U[0, w_i]$.

After the sample size and the number of replicates are determined, the Monte Carlo simulation can be accomplished in the following steps:

1. Input the users' priority levels $w_1, w_2, \ldots, w_n$.

2. Generate a pseudo random sample $x_1, x_2, \ldots, x_n$ from uniform distributions $U[0, w_1], U[0, w_2], \cdots, U[0, w_n]$, respectively.

3. Compute the fairness score $G(x_1, x_2, \ldots, x_n; w_1, w_2, \ldots, w_n)$ using formula (3).

4. Repeat steps 2 and 3 for one million times, and accumulate G across samples.

5. Sort G values in an ascending order.

6. Input significance level, 0.1, 0.05, 0.025, and 0.01.

7. Compute the $(100\alpha)^{th}$ percentile of G values as the critical value for different significance levels α . It can be calculated as following:

Define $n_1 = k\alpha$, where k is the number of samples drawn, then

$$G_\alpha = \begin{cases} G(n_1) & \text{if } n_1 \text{ is an integer} \\ [G([n_1]) + G([n_1]+1)]/2 & \text{if } n_1 \text{ is not an integer ,} \end{cases}$$

(9)

where $G(i)$ is the i^{th} element of the sorted fairness scores.

In the case that users have the same priority level, each user's priority level is specified as 1. Repeat step 1 to step 7 for different sample sizes. Finally, a table of the critical values can be constructed for different sample sizes at different significance levels. For the case that users have different priority levels, it is difficult to generate a similar table for the fairness score function $G(x_1, x_2, \ldots, x_n; w_1, w_2, \ldots, w_n)$ since the users' priority levels vary. Therefore, the critical value will be computed dynamically for given priority levels by following step 1 to step 7.

4 CASE 1: USERS' PRIORITY LEVELS ARE THE SAME

4.1 Construction of the Table of Critical Values

In the case that all the users have the same priority level, a priority level of 1 is assigned to each user. For example, for sample size of 5 case, $w_1 = w_2 = w_3 = w_4 = w_5 = 1$. By using the program developed in this study, the critical values will be calculated for sample size of 5 as following.

1. Input the users' priority levels $w_1 = w_2 = w_3 = w_4 = w_5 = 1$.

2. Generate a pseudo random sample x_1, x_2, x_3, x_4, x_5 from uniform distributions $U[0,1]$.

3. Compute the fairness score $G(x_1, x_2, x_3, x_4, x_5; 1,1,1,1,1)$ using formula (3).

4. Repeat steps 2 and 3 for one million times, and accumulate G across samples.

5. Sort G values in an ascending order.

6. Input significance level, 0.1, 0.05, 0.025, and 0.01.

7. Compute the 100,000th, 50,000th, 25,000th, and 10,000th percentile of G values as the critical values.

Similarly, a table of critical values is constructed for sample sizes from 5 to 50 as shown in Table 1.

In order to show the effect of sample size on the critical value, two simulated probability density functions are displayed in Figure 2 and Figure 3 with sample sizes 5 and 20, respectively. It can be seen that when the sample size is small, there is more variability in the sampling distribution than the case that the sample size is large. For a random sample of 5 observations, the fairness score can easily be as low as 0.90. However, when the sample size increases to 20, it becomes highly unlikely to obtain a fairness score as 0.96. As a result, the critical value increases with an increase in sample size for a fixed significance level.

Table 1 Critical value for different sample size n and significance level α

n \ α	0.1	0.05	0.025	0.01
5	0.8452	0.8086	0.7702	0.7165
6	0.8799	0.8536	0.8276	0.7924
7	0.9034	0.8843	0.8651	0.8389
8	0.9194	0.9044	0.8891	0.8686
9	0.9310	0.9187	0.9070	0.8907
10	0.9397	0.9296	0.9197	0.9071
11	0.9469	0.9386	0.9304	0.9192
12	0.9522	0.9452	0.9384	0.9296
13	0.9569	0.9508	0.9447	0.9372
14	0.9609	0.9554	0.9503	0.9438
15	0.9639	0.9591	0.9545	0.9487
16	0.9667	0.9625	0.9586	0.9532
17	0.9692	0.9654	0.9616	0.9569
18	0.9714	0.9679	0.9646	0.9604
19	0.9731	0.9699	0.9668	0.9630
20	0.9747	0.9718	0.9690	0.9654
21	0.9762	0.9736	0.9710	0.9679
22	0.9775	0.9751	0.9727	0.9696
23	0.9786	0.9763	0.9742	0.9715
24	0.9797	0.9776	0.9756	0.9730
25	0.9806	0.9786	0.9768	0.9744
26	0.9816	0.9797	0.9779	0.9756
27	0.9823	0.9806	0.9789	0.9768
28	0.9831	0.9815	0.9800	0.9780
29	0.9838	0.9822	0.9808	0.9789
30	0.9844	0.9829	0.9816	0.9798
31	0.9850	0.9836	0.9823	0.9807
32	0.9855	0.9842	0.9830	0.9814
33	0.9860	0.9848	0.9836	0.9821
34	0.9865	0.9853	0.9842	0.9828
35	0.9870	0.9858	0.9848	0.9835
36	0.9874	0.9863	0.9853	0.9840
37	0.9878	0.9867	0.9857	0.9846
38	0.9881	0.9871	0.9862	0.9851
39	0.9885	0.9876	0.9867	0.9857
40	0.9888	0.9879	0.9871	0.9860
41	0.9891	0.9883	0.9875	0.9865
42	0.9894	0.9886	0.9878	0.9869
43	0.9897	0.9889	0.9882	0.9872
44	0.9900	0.9892	0.9885	0.9876
45	0.9902	0.9895	0.9888	0.9880
46	0.9905	0.9897	0.9891	0.9883
47	0.9907	0.9900	0.9893	0.9885
48	0.9909	0.9902	0.9896	0.9889
49	0.9911	0.9905	0.9899	0.9891
50	0.9913	0.9907	0.9901	0.9894

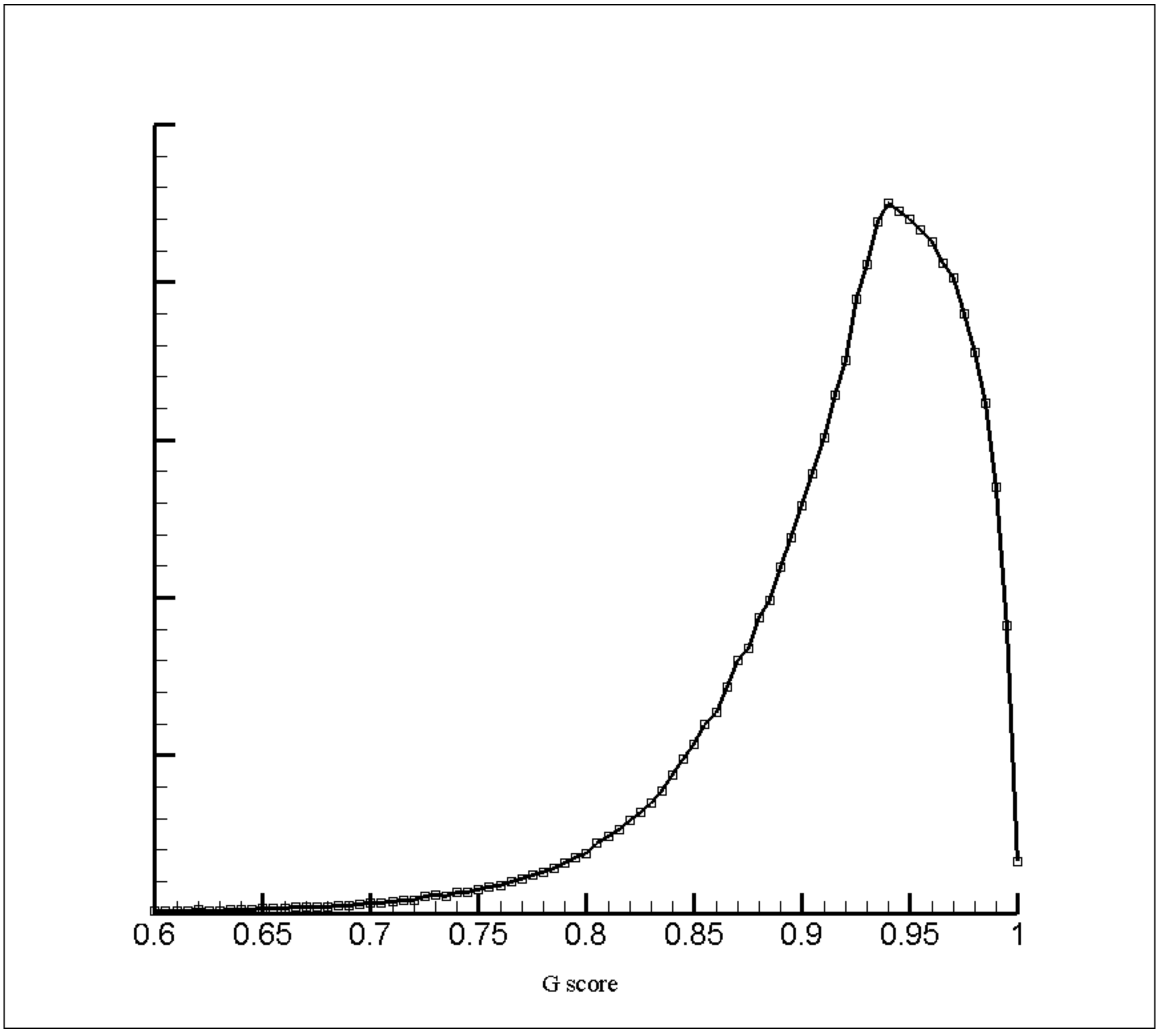

Figure 2 Shape of the probability density function of G with sample size 5

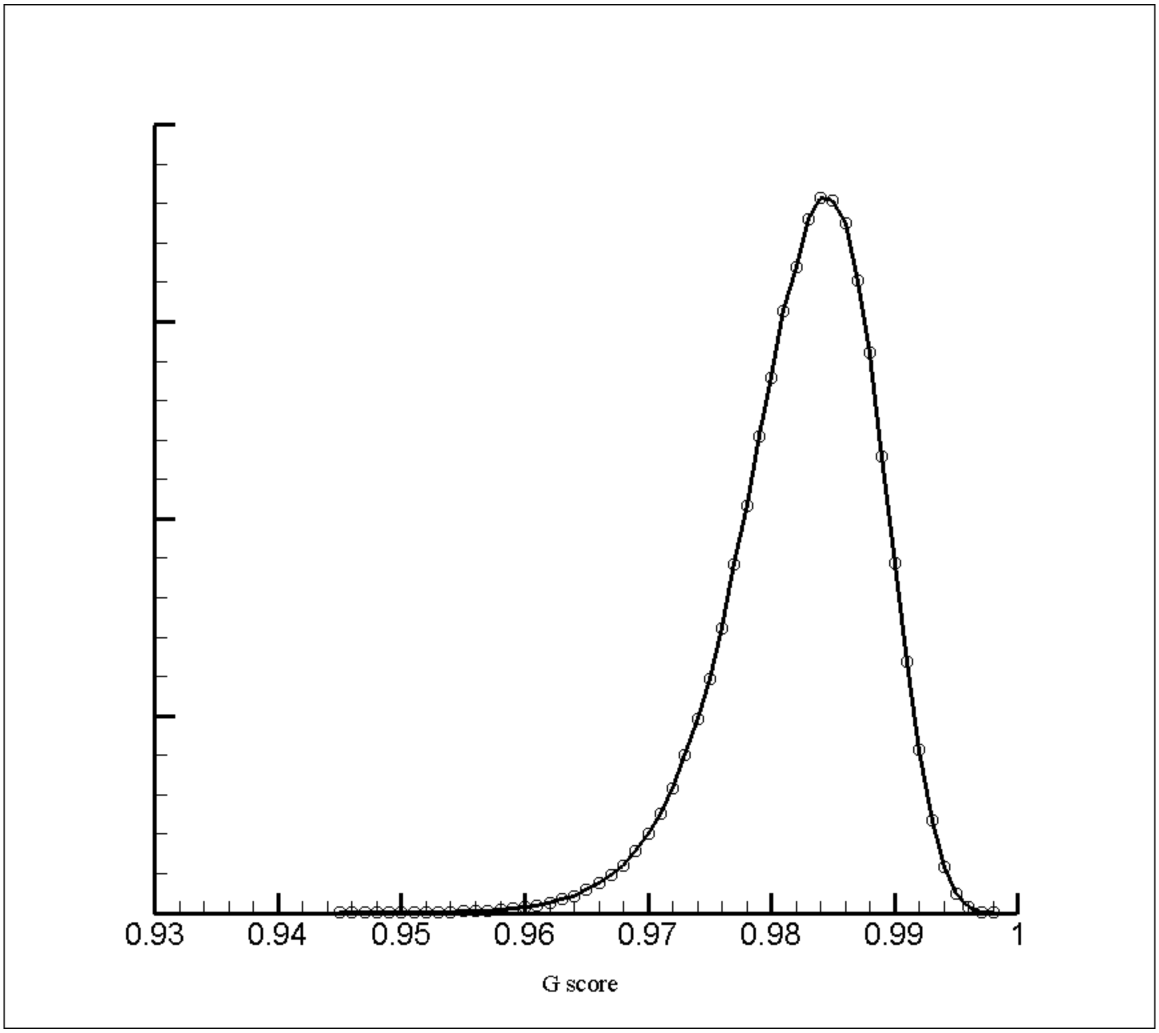

Figure 3 Shape of the probability density function of G with sample size 20

4.2 An Example

Suppose ten users with the same priority level are drawn from a population to check the fairness of the network resource distribution. The significance level for the statistical test is set to be 0.05.

The null and alternative hypotheses are:

H_0: The network resource distribution process is under fair control.

H_1: The network resource distribution process is not under fair control.

From Table 1, the critical value is 0.9296 for the significance level of 0.05.

In this example, a random sample of size 10 is drawn from a uniform distribution on [0, 1]. The sample data are listed in Table 2. The test statistic (fairness score) G calculated by formula (3) is 0.9454.

Table 2 Sample drawn from the uniform distribution

w_i	x_i
1	0.1849626
1	0.9700887
1	0.3998243
1	0.2593986
1	0.9216026
1	0.9692773
1	0.5429792
1	0.5316917
1	0.0497940
1	0.0665666

Because the value of the test statistic 0.9454 is larger than the critical value 0.9296, the null hypothesis cannot be rejected. It is concluded that there is insufficient evidence to say the network resource distribution is not under fair control.

4.3 Control Chart for Fairness Score

When the control limits are available, a control chart for fairness score can be constructed to monitor the fairness of network resource distribution. Because the fairness score is a numerical value calculated from the sample, the fairness score control chart is a variable control chart. Only the lower control limit is needed as mentioned early. For example, take sample size of 10, and significance level of 0.05, then the lower control limit will be 0.9296 from Table 1. In this case, the probability of committing type I error is 0.05, which means that a false out-of-control signal will be generated in 5 out of 100 samples even the process is under control.

After setting the control limits, one might collect sample data and compute the fairness score and plot into the control chart. The frequency of sampling depends on the sample size. Generally, when the sample size is small, draw samples more frequently; when the sample size is large, draw samples less frequently.

Some sensitizing rules for detecting the process of network resource distribution out-of-control can be adopted from Shewhart control chart [2].

(1) One or more points plot below the lower control limit. The probability for any single point plots below the lower control limit is 0.05.

(2) Two of three connective points plot below the lower control limit. The probability for this case is 0.0071.

(3) Four of five connective points plot below the lower control limit. The probability for this case is 2.97E-5.

(4) Six points in a row steadily decreasing.

The basic criterion is the first rule. Some other criteria may be applied simultaneously to increase the sensitivity of the control chart.

5 CASE 2: USERS' PRIORITY LEVELS ARE DIFFERENT

5.1 Procedure for Finding the Critical Values

Suppose n users with priority levels $w_1, w_2, \ldots, w_n$ are drawn from a population to check the fairness of the network resource distribution. By following the procedure described in Section 3, the critical value is calculated as following.

1. Input the users' priority levels $w_1, w_2, \ldots, w_n$.

2. Generate a pseudo random sample $x_1, x_2, \ldots, x_n$ from uniform distributions $U[0, w_1], U[0, w_2], \cdots, U[0, w_n]$, respectively.

3. Compute the fairness score $G(x_1, x_2, \ldots, x_n; w_1, w_2, \ldots, w_n)$ using formula (3).

4. Repeat steps 2 and 3 for one million times, and accumulate G across samples.

5. Sort G values in an ascending order.

6. Input significance level, 0.1, 0.05, 0.025, and 0.01.

7. Compute the 100,000th, 50,000th, 25,000th, and 10,000th percentile of G values as the critical values.

It can be assumed that $w_1, w_2, \ldots, w_n$ are all rational numbers. This assumption should not place any restriction on application problems.

For rational numbers $w_1, w_2, \ldots, w_n$, there exist integers $w_1', w_2', \ldots, w_n'$ such that $w_1 : w_2 : \ldots : w_n = w_1' : w_2' : \ldots : w_n'$. Therefore, the priority levels $w_1, w_2, \ldots, w_n$ can be considered to be integers without loss of generality. This is because $G(x_1, x_2, \ldots, x_n; w_1, w_2, \ldots, w_n) = G(x_1, x_2, \ldots, x_n; w_1', w_2', \ldots, w_n')$ for any $x_1, x_2, \ldots, x_n$.

5.2 An Example

Suppose 10 users with priority levels from 1 to 10 are drawn from a population to check the fairness of the network resource distribution. By following the procedure described above, the critical value is calculated as following.

1. Input the users' priority levels $w_1 = 1, w_2 = 2, \cdots, w_{10} = 10$.

2. Generate a pseudo random sample $x_1, x_2, \cdots, x_{10}$ from uniform distributions $U[0,1], U[0,2], \cdots, U[0,10]$, respectively.

3. Compute the fairness score $G(x_1, x_2, \cdots, x_{10}; 1, 2, \cdots, 10)$ using formula (3).

4. Repeat steps 2 and 3 for one million times, and accumulate G across samples.

5. Sort G values in an ascending order.

6. Input significance level, 0.1, 0.05, 0.025, and 0.01.

7. Compute the 100,000th, 50,000th, 25,000th, and 10,000th percentile of G values as the critical values. The critical values at different significance levels are listed in Table 3.

The simulated sampling histogram is displayed in Figure 4 that graphically illustrates the distribution of fairness score. The number on the top of each bar is the cumulative percentage. If the network resource distribution is under fair control, it can be expected that the sample fairness score should be close to 1.

Table 3 The critical values for the case $w_1 = 1, w_2 = 2, \cdots, w_{10} = 10$

α	0.1	0.05	0.025	0.01
G_α	0.9337	0.9219	0.9099	0.8936

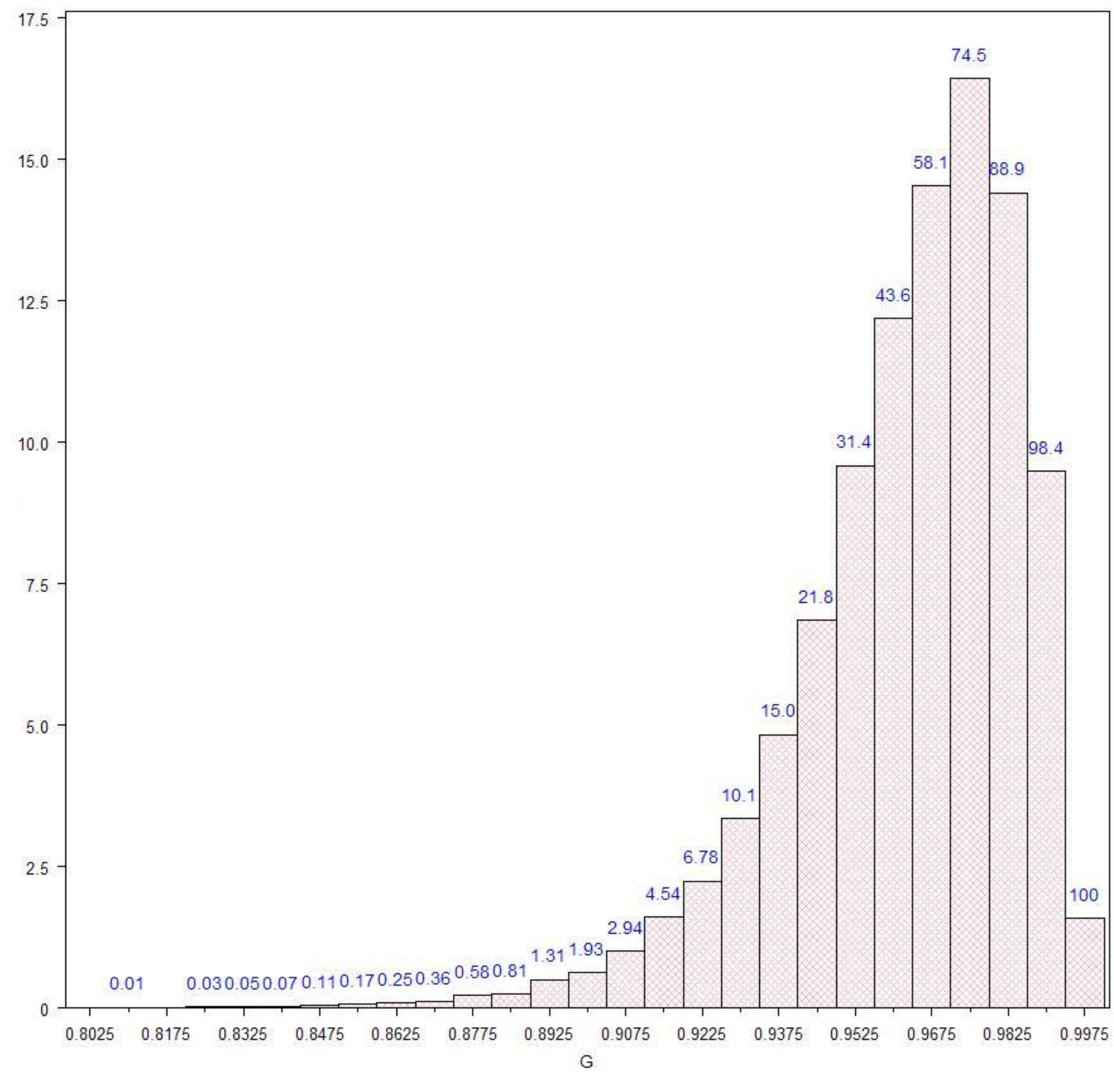

Figure 4 Histogram of G under the null hypothesis for the case

$w_1 = 1, w_2 = 2, \cdots, w_{10} = 10$

The hypotheses can be stated as

H_0: The network resource distribution process is under fair control.

H_1: The network resource distribution process is not under fair control.

From Table 3, the critical value is 0.9219 for the significance level 0.05.

In this example, a random sample of size 10 is drawn from a uniform distribution on [0, 1]. The sample data set is listed in Table 4. The test statistic (fairness score) G calculated by formula (3) is 0.9097.

Table 4 Sample drawn from the uniform distribution

w_i	x_i
1	0.1849626
2	0.9700887
3	0.3998243
4	0.2593986
5	0.9216026
6	0.9692773
7	0.5429792
8	0.5316917
9	0.0497940
10	0.0665666

Because the value of the test statistic 0.9097 is less than the critical value 0.9219, the null hypothesis is rejected. It is concluded with 95% confidence that the network resource distribution is not under fair control.

Similarly, a control chart of fairness score for this special case can be constructed as discussed in Section 4.3.

6. DISCUSSION AND CONCLUSION

Statistical process control (SPC) has its long history for its application in industry. The purpose of the statistical process control is to use statistical analysis to monitor processes to ensure that the processes are running properly. A typical tool in SPC is the control chart. To construct a control chart, one or more quality characteristics and the control limits/critical values are needed. To analyze the control chart, some sensitizing rules are needed.

In this research, a statistical analysis method is developed to monitor the fairness of network resource distribution. A newly developed fairness score function is adopted as the quality characteristic. The adopted fairness score function considers the case that the users have the same or different priority levels. Especially, this fairness score function possesses all the necessary properties required as a quality characteristic for the purpose of statistical process control. Monte Carlo simulation is designed and implemented to find the distribution and the critical values of the fairness score function. When users have the same priority level, a table of the critical values is given for different sample sizes and different significance levels. When users have different priority level, it is difficult to generate a similar table for the fairness score function since the users' priority levels vary. Therefore, the critical values are computed dynamically for given priority levels. In each case, an example is given to demonstrate how to apply the approach developed in this study.

When the control limits are available, network resource distribution researchers might construct a control chart of fairness score to monitor the fairness of network resource distribution. After the sample size and the significance level are decided, the lower control limit can be selected from Table 1 for the case that all the users have the same priority level. For the case that users have different priority levels, the control limits were computed for given priority levels as exemplified in Section 5. The upper limit is automatically set to be one. Then sample data are collected and the fairness scores are computed and put into the control chart. The frequency of sampling depends on the sample size. Generally, when the sample size is small, draw samples more frequently; when the sample size is large, draw samples less frequently.

Some sensitizing rules can be adopted from the Shewhart chart. The basic rule of detecting the out-of-control status of the process of network resource distribution is one or more points plot below the lower control limit. Some other criteria may be applied simultaneously to increase the sensitivity of the control chart.

The statistical process control method presented in this research is not only for the fairness process control of computer network resource distribution but also for any other fairness issues in resource allocation applications.

While the results of this study look quite promising, there is much left to address. In the future, real data of the network resource distribution will be collected to validate the method developed in this study.

REFERENCES

[1] James F. Kurose and Keith W. Ross (2009) *Computer Networking: A Top-Down Approach* (5th ed.). Addison Wesley.

[2] Doulas C. Montgomery (2005) *Introduction to Statistical quality Control* (5th ed.). John Wiley & Sons, Inc.

[3] Leonard A. Doty (1996) *Statistical Process Control* (2nd ed.). Industrial Press Inc.

[4] John S. Oakland (2003) *Statistical Process Control*. (5th ed.). Butterworth-Heinemann, Business & Economics.

[5] R. Jain, D. Chiu and W. Hawe (1984) A Quantitative Measure of Fairness and Discrimination for Resource Allocation in Shared Computer System, Technical Report, Digital Equipment Corporation, DEC-TR-301.

[6] Denda Robert, Banchs Albert and Effelsberg Wolfgang (2000) The Fairness Challenge in Computer Networks, Lecture Notes in Computer Science, Vol. 1922, 208-220.

[7] D. Bertsekas and R. Gallager (1987) *Data Networks*, Prentice Hall.

[8] D. Chiu and R. Jain (1989) Analysis of the Increase and Decrease Algorithms for Congestion Avoidance in Computer Networks, *Computer Networks and ISDN Systems*, 17, 1-14.

[9] F. Kelly, et al. (1998) Rate Control for Communication Networks: Shadow Prices, Proportional Fairness and Stability, *Journal of the Operational Research Society*, 49, 237-252.

[10] R. Mazumdar, L. Mason and C. Douligeris (1991) Fairness in Network Optimal Flow Control: Optimality of Product Forms, *IEEE Trans. on Communications*, 39, 775-782.

[11] Zhenmin Chen and Chi Zhang (2005) A New Measurement for Network Sharing Fairness, *Computer and Mathematics with Applications*, 50, 803-808.

[12] Zhenmin Chen and Chi Zhang (2010) Network Sharing Fairness Function with Consideration of Priority Levels, *Computer and Mathematics with Applications*, 60, 2548-2555.

[13] X. Fan, A. Felsovalyi, S. A. Sivo and S. C. Keenan (2002) SAS for Monte Carlo Studies: A Guide for Quantitative Researchers. Cary, NC: SAS Institute Inc.