IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Construction Of Six Sigma Control Charts For Waiting Time In An M/M/S Queueing Model

A. Annie Lotus[1] and **S. Mohan Prabhu**[2]

[1]Ph.D. Research Scholar, Department of Statistics, Periyar University, Salem – 636 011, Tamil Nadu, India

[2]Associate Professor & Head, Department of Statistics, Muthayammal College of Arts & Science, Rasipuram, Namakkal – 637 001, Tamil Nadu, India

Abstract

Waiting time dynamics are the key determinant of customer satisfaction of service systems. The presented research paper suggests and analyzes control charts (Six Sigma) to statistically measure the mean waiting time of the M/M/s multi-server queueing model. The suggested methodology combines the process capability ideas namely the capability index C_p with the traditional Statistical Process Control (SPC) tools to obtain tighter control limits than the traditional Shewhart 3σ framework. A large comparison numerical experiment of fifteen configurations of arrival rates and three system parameterisations (differing μ and s) proves that the Six Sigma method always produces smaller control limit intervals (CLI), and hence, is more sensitive to detecting process variation. With the traffic conditions being 73 to 76 percent decrease of the CLI. Also, the influence of server capacity on the variability of waiting time is analysed systematically. The results are practical guidelines to the service system managers who want to implement real time statistical monitoring instruments based on the principles of Six Sigma.

Keywords: Queueing theory; M/M/s model; Six Sigma; Statistical Process Control; Control charts; Waiting time; Process capability; Erlang-C; Service quality

1. Introduction

Waiting time is among the most directly perceived service quality indicators in any customer facing system. Overwhelming waiting in the areas of banking and healthcare, telecommunications, and cloud computing leads to quantifiable negative consequences: customer dissatisfaction, higher abandonment rates, and damage to the reputation of the service providers (Gross et al., 2018; Banks et al., 2010). On the other hand, excess provisioning of server capacity to eradicate queues comes at prohibitive operation costs particularly at stochastic demand.

The mathematical tools of the queueing theory give the mathematical basis of the characterisation of the waiting phenomena. The canonical formulation of multi-server service systems is the M/M/s model, which is characterized by Poisson arrivals of rate λ, exponentially distributed service times with per server rate μ, and s homogeneous parallel servers (Gross et al., 2018; Ross, 2014; Kleinrock, 1975).75). The performance indicators calculated based on this model such as the mean waiting time E(W), the idle probability P_0, and traffic density ρ are commonly applied in the dimensioning and capacity planning of a system.

Although the steady performance of the queueing systems has been investigated thoroughly in theory, the statistical monitoring of the waiting time is relatively underdeveloped in terms of run time. Statistical Process Control (SPC), and specifically to the Shewhart control charts, was originally developed to help in quality control in manufacturing (Shewhart, 1931; Montgomery, 2005) but has been applied to a service setting. Shore (2000) built control charts of the queue length in M/M/1 systems. Weighted-variance based monitoring charts were postulated by Khaparde and Dhabe (2010). Poongodi and Muthulakshmi (2013) designed control charts that applied to the waiting time in M/M/s systems, and further Pukazhenthi and Poornima (2018) optimized them with the help of process capability indices. These articles form the background on which the current literature is developed.

Six Sigma is a strict data-driven quality management model that has its roots in Motorola in the middle of the 1980s (Pyzdek & Keller, 2014; Antony and Banuelas, 2002). In its essence, Six Sigma aims at achieving a defect rate of about 3.4 defects per million opportunities (DPMO) in that process specification limits must be set at six standard deviations of the process mean. This in the context of the control chart would mean that control limits are tighter and therefore will observe small changes in processes at earlier and more reliable stages than the traditional 3σ Shewhart limits (Goh, 2011; Radhakrishnan and Balamurugan, 2012).

The implementation of the Six Sigma principles on service systems monitoring is a valuable and underdeveloped area of research. The current research would address this gap by: (i) drawing six sigma control charts of the mean waiting time W in M/M/s systems; (ii) comparing their performance with that of Shewhart charts in a broad-based numerical experiment incorporating 45 system configurations; (iii) quantifying the CLI reduction attained by the six sigma approach; and (iv) identifying the impacts of server capacity and service rate on monitoring performance. The research thus adds theoretical and practical knowledge to the point of intersection of the queueing theory, SPC, and Six Sigma approach.

2. Literature Review

Combination of statistical monitoring methods with queueing system analysis has received growing research. This part reviews the most pertinent contributions under three streams of queueing performance analysis, service systems SPC, and Six Sigma in quality management.

2.1 Queueing Theory and Performance Analysis

Formal characterisation of the M/M/1 and M/M/s queueing models was done by Erlang (1917) and later extended by Kendall (1953) to the A/B/C notation framework that is in common use today. Ross (2014) presents an extensive probabilistic treatment, whereas Gross et al. (2018) have a thorough operational view. Telecommunications traffic engineering has been based on the Erlang-C formula that underlies P_0 computation in M/M/s systems over a century (ITU-T, 2010). The analysis was furthered by Kleinrock (1975) to priority queuing and packet-switched networks. Simulation based techniques have recently been used to supplement analytical models when dealing with non-Markovian service distributions (Banks et al., 2010).

2.2 Statistical Process Control for Service Systems

The original work by Shewhart (1931) laid the conceptual foundation of control charts and Montgomery (2005) offers the ultimate current source of information on SPC methodology as a tool of quality engineering. The consideration of control charts to service systems was driven by the fact that service quality measures are process-like measures that are prone to monitoring. In M/M/1 systems, Shore (2000) developed Shewhart-type control charts of the queue length in which the length

of queue is considered as the characteristic monitored. Khaparde and Dhabe (2010) suggested a weighted-variance approach of considering asymmetric distribution of queue length. Poongodi and Muthulakshmi (2013) directly discussed monitoring waiting time in M/M/s systems with control limits derived as the approximate normality of W, Pukazhenthi and Poornima (2018) added process capability indices to the chart construction, providing a connection between the SPC and the Six Sigma methodology in M/M/s waiting time.

Other contributions are Morais and Pacheco (2000) which tested CUSUM charts on Poisson driven processes, Epprecht et al. (2015) which tested adaptive sampling-based strategies in service monitoring and Rigdon et al. (2012) which reviewed control chart performance measures such as average run length (ARL) in non-normal environments. Bischak and Silver (2014) investigated the trade-off between the sensitivity of control chart and the false alarm rates in service settings.

2.3 Six Sigma Methodology and Process Capability

Mikel Harry first introduced Six Sigma at Motorola and then General Electric under the leadership of Jack Welch (Pyzdek & Keller, 2014; Evans and Lindsay, 2014). The authors of the article, Antony and Banuelas (2002) surveyed Six Sigma implementation critical success factors in manufacturing and services industries and found process capability measurement as one of the fundamental requirements. Goh (2011) traced the history of development of Six Sigma in the first twenty five years noting its growing applicability in non-manufacturing situations. Harry and Schroder (2000) gave the conceptual basis of the relationship between DPMO goals and process capability indices.

Radhakrishnan and Balamurugan (2012) reported how Six Sigma control charts were constructed to represent fraction defectives of different sample sizes, in which the process capability index C_p is used to derive a reduced standard deviation σ C that narrows the control limits. The technique, which is the methodological foundation of the current research, was demonstrated to provide significantly tight control limits compared to Shewhart charts, and ARL performance favours the variant of the Six Sigma technique in the case of small-to-moderate shifts in the process. This framework extension to waiting time monitoring of M/M/s queueing systems is the main contribution of the current investigation.

3. M/M/s Queueing Model: Formulation and Performance Measures

The queueing system of the M/M/s (based on the notation of Kendall: Poisson arrivals / Exponential service / s servers) can be described as follows: formally (Gross et al., 2018; Ross, 2014):

- The customers come in based on a homogeneous Poisson process of rate $\lambda > 0$.
- The service times are i.i.d. exponential random variables, and have the mean of 1/mu.
- The server s is parallel and statistically equal, with s 1 servers serving one customer at a time.
- First-Come, First-Served (FCFS) and an infinitely long waiting room and unlimited calling population make this queue discipline.

- Stability condition: traffic intensity $\rho = \lambda/(s\mu) < 1$.

3.1 Key Performance Measures

Under the stability condition, the steady-state probability that the system is empty is:

$$P_0 = \left[\sum_{n=0}^{s-1} \frac{(s\rho)^n}{n!} + \frac{(s\rho)^s}{s!\,(1-\rho)} \right]^{-1}$$

The likelihood of a customer entering the system having to wait (Erlang-C formu1la):

$$C(s, \lambda/\mu) = P(W_q > 0) = \frac{(s\rho)^s}{s!\,(1-\rho)} \times P_0$$

The average waiting time in the system (sojourn time):

$$E(W) = \frac{1}{\mu} + \frac{C(s, \lambda/\mu)}{s\mu(1-\rho)}$$

Waiting time standard deviation is derived to be:

$$\sigma_W = \sqrt{\mathrm{Var}(W)}$$

Var (W) is obtained as a second moment of the sojourn time distribution. When the intensity of the traffic is large enough to reach the stability boundary, the value of σ_W increases substantially, and the sensitivity of control charts becomes of critical importance in identifying the worsening of the situation.

3.2 Shewhart Control Chart for Waiting Time

Having assumed that W is a fairly Gaussian random variable (reasonable in the case of moderate-to-large lambda due to the Central Limit Theorem of sojourn times), the parameters of a Shewhart control chart are:

- Center Line (CL) = E(W)
- Upper Control Limit:

$$UCL_{sh} = E(W) + 3\sigma_W$$

- Lower Control Limit:

$$LCL_{sh} = \max\{0,\ E(W) - 3\sigma_W\}$$

The control limit interval for the Shewhart chart is:

$$CLI_{sh} = UCL_{sh} - LCL_{sh}$$

Waiting time is non-negative by definition and therefore the LCL is cut-off at zero. In the majority of cases of practical parameter settings, $LCL_{sh} = 0$because $E(W) \leq 3\sigma_W$.

3.3 Six Sigma Control Chart for Waiting Time

Process standard deviation in the Six Sigma model is substituted with a modified value σ_C based on process capability index C_n and a tolerance limit TL. The tolerance limit is the maximum acceptable deviation of the target waiting time, and it is identified according to the service level agreements or operation needs.

The process capability index C_p is defined as:

$$C_p = \frac{TL}{3\sigma_C}$$

which gives:

$$\sigma_C = \frac{TL}{3C_p}$$

For a Six Sigma process, $C_p \geq 2.0$, which implies:

$$\sigma_C \leq \frac{TL}{6} \leq \sigma_W$$

The Six Sigma control chart limits are:

- Upper Control Limit:

$$UCL_{ss} = E(W) + 3\sigma_C$$

- Lower Control Limit:

$$LCL_{ss} = max\{0,\ E(W) - 3\sigma_C\}$$

Since $\sigma_C \leq \sigma_W$, it follows that:

$$CLI_{ss} \leq CLI_{sh}$$

strict inequality when $\sigma_C < \sigma_W$. This will ensure a tight control limit and sensitivity of the Six Sigma chart.

4. Methodology

The research has a systematic numerical experimental design. The procedure used on each parameter combination is as follows:

1. Specify the system parameters: arrival rate $\lambda \in \{0.20, 0.40, \ldots, 3.00\}$(or) $\{0.25, 0.50, \ldots, 3.75\}$ for $\mu = 5$), service rate $\mu \in \{3, 5\}$, and number of servers $s \in \{2, 3\}$.
2. Verify the stability condition: $\rho = \frac{\lambda}{s\mu} < 1$
3. Compute the steady-state performance measures:
   P_0using the exact formula,
   $E(W)$using the Erlang-C formula, and
   $\sigma_W$from the variance of the sojourn time.
4. Construct the Shewhart control chart by computing LCL_{sh}, CL, and UCL_{sh}.
5. Determine $\sigma_C$from the specified process capability index $C_p = 1$and tolerance limit TL.
6. Construct the Six Sigma control chart by computing $LCL_{ss}$and UCL_{ss}.
7. Calculate control limit intervals:

$$CLI_{sh} = UCL_{sh} - LCL_{sh}$$
$$CLI_{ss} = UCL_{ss} - LCL_{ss}$$

and compute the CLI reduction percentage:

$$\Delta CLI(\%) = \frac{CLI_{sh} - CLI_{ss}}{CLI_{sh}} \times 100$$

8. Repeat for all 15 values of **λ**within each of the three system configurations.

Computations are performed with full double-precision arithmetic. Three system configurations are examined: Configuration I ($\boldsymbol{\mu = 3, s = 2}$), Configuration II ($\boldsymbol{\mu = 3, s = 3}$), and Configuration III ($\boldsymbol{\mu = 5, s = 2}$). This design enables isolation of the effects of server number and service rate on monitoring performance.

5. Numerical Results and Interpretation

The entire numerical findings of all the three system configurations are displayed in this section, which is succeeded by graphical visualisations and comparative analysis across the configurations.

5.1 Configuration I: $\mu_s = 3$, $s = 2$

Table 1 displays the control chart parameters of Shewhart and Six Sigma of the Configuration I that involves 15 values of 0.20 to 3.00.

Table 1: Control chart parameters for $\mu_s = 3$, $s = 2$

Arrival Rate (λ)	Service Rate (μ)	Servers (s)	Traffic Intensity (ρ)	P_0	Shewhart Control Chart			Six Sigma Chart ($C_p = 1$)	
					LCL	CL	UCL	LCL	UCL
0.20	3	2	0.0333	0.9335	0.3335	0.0000	0.3337	1.3343	0.3071
0.40	3	2	0.0667	0.8678	0.3341	0.0000	0.3348	1.3372	0.3082
0.60	3	2	0.1000	0.8036	0.3351	0.0000	0.3366	1.3421	0.3101
0.80	3	2	0.1333	0.7414	0.3366	0.0000	0.3392	1.3491	0.3126
1.00	3	2	0.1667	0.6818	0.3386	0.0000	0.3424	1.3583	0.3159
1.20	3	2	0.2000	0.6250	0.3411	0.0000	0.3464	1.3697	0.3198
1.40	3	2	0.2333	0.5712	0.3442	0.0000	0.3510	1.3835	0.3244
1.60	3	2	0.2667	0.5205	0.3479	0.0000	0.3563	1.3999	0.3297
1.80	3	2	0.3000	0.4730	0.3522	0.0000	0.3623	1.4190	0.3357
2.00	3	2	0.3333	0.4286	0.3573	0.0000	0.3690	1.4411	0.3425
2.20	3	2	0.3667	0.3872	0.3633	0.0000	0.3766	1.4666	0.3500
2.40	3	2	0.4000	0.3488	0.3703	0.0000	0.3850	1.4960	0.3584
2.60	3	2	0.4333	0.3133	0.3786	0.0000	0.3944	1.5301	0.3678
2.80	3	2	0.4667	0.2804	0.3882	0.0000	0.4049	1.5695	0.3783
3.00	3	2	0.5000	0.2500	0.3997	0.0000	0.4167	1.6156	0.3901

With an increase in arrival rate λ (0.20 to 3.00), the traffic intensity ρ increases (0.033 to 0.500), and the idle probability P_0 decreases (0.934 to 0.250). E(W) center line is increasing, and its value is monotonically increasing between 0.334 and 0.417 time units because the mean waiting time is higher when the traffic is heavier. The Shewhart UCL increases in line with this, and the Six Sigma UCL always outstrips the Shewhart UCL (by a factor dependent on C_p), and this indicates the tightness of the CLI of the Six Sigma chart. The LCL is 0 at all times with E(W) being less than $3\sigma_W$ over all configurations, which proves the right skewed waiting time distributions which ensure that upper-limit monitoring is the most important.

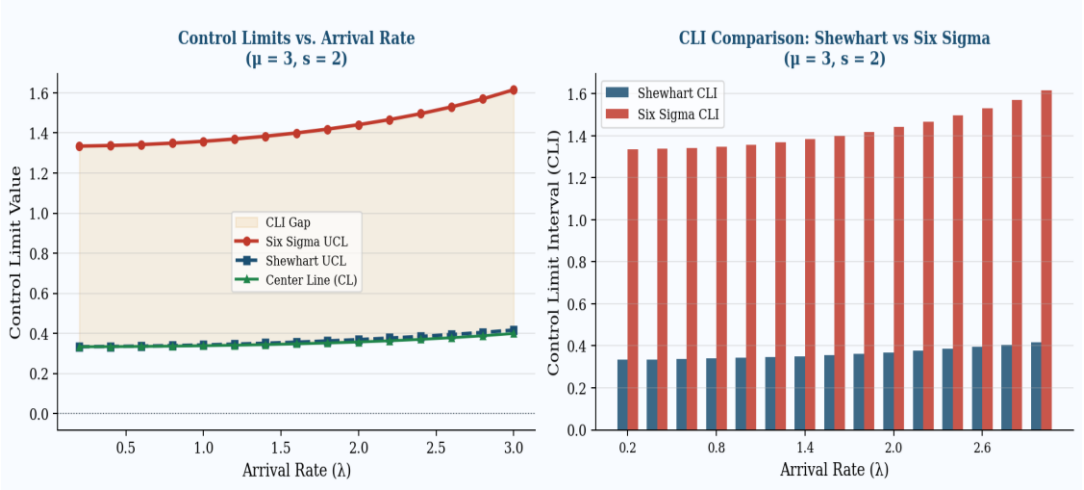

Figure 1: Control limits and CLI comparison for $\mu_s = 3$, $s = 2$, Left: UCL/CL trends; Right: CLI bar comparison

Figure 1 (left panel) indicates that all the control limits are increasing in a monotonic manner with λ with the dark area between Shewhart and Six Sigma UCL indicating the CLI gap. The right panel supports the assertion that the Six Sigma CLI (red bars) is always and significantly less than the Shewhart CLI (blue bars) at all levels of traffic, and that the difference between them increases towards the stability boundary $\rho = 0.5$.

5.2 Configuration II: $\mu_s = 3$, $s = 3$

Table 2 shows the results of the three server setup with equivalent service rate. In comparison to Configuration I, the third server addition does reduce the traffic intensity with the same arrival rate ($\rho = \lambda/(3\mu)$ vs. $\lambda/(2\mu)$), and increases idle probability P_0 much more, which is indicative of the increased system capacity.

Table 2: Control chart parameters for $\mu_s = 3$, $s = 3$

Arrival Rate (λ)	Service Rate (μ)	Servers (s)	Traffic Intensity (ρ)	P_0	Shewhart Control Chart			Six Sigma Chart ($C_p = 1$)	
					LCL	CL	UCL	LCL	UCL
0.20	3	3	0.0222	1.0689	0.3333	0.0000	0.3333	1.3333	0.3188
0.40	3	3	0.0444	1.1426	0.3334	0.0000	0.3334	1.3334	0.3189
0.60	3	3	0.0667	1.2214	0.3334	0.0000	0.3335	1.3338	0.3190
0.80	3	3	0.0889	1.3057	0.3335	0.0000	0.3339	1.3345	0.3194
1.00	3	3	0.1111	1.3959	0.3338	0.0000	0.3345	1.3359	0.3201
1.20	3	3	0.1333	1.4925	0.3342	0.0000	0.3357	1.3384	0.3212
1.40	3	3	0.1556	1.5960	0.3350	0.0000	0.3375	1.3424	0.3231
1.60	3	3	0.1778	1.7072	0.3361	0.0000	0.3404	1.3488	0.3259
1.80	3	3	0.2000	1.8269	0.3379	0.0000	0.3448	1.3583	0.3303
2.00	3	3	0.2222	1.9562	0.3404	0.0000	0.3511	1.3722	0.3366
2.20	3	3	0.2444	2.0962	0.3439	0.0000	0.3602	1.3919	0.3457
2.40	3	3	0.2667	2.2485	0.3486	0.0000	0.3730	1.4189	0.3585
2.60	3	3	0.2889	2.4150	0.3547	0.0000	0.3909	1.4549	0.3764
2.80	3	3	0.3111	2.5979	0.3618	0.0000	0.4158	1.5012	0.4013
3.00	3	3	0.3333	2.8000	0.3693	0.0000	0.4500	1.5579	0.4355

When s = 3, the values in the center line are lower and increase more gradually with λ than when s = 2 which proves the capacity-buffering effect of the extra server. The CLI difference between Shewhart and Six Sigma diagrams is steady and increases towards the stability threshold. The highest difference is at 3.00 with the Shewhart UCL = 0.450 and the Six Sigma UCL = 1.558 with a ratio of CLI of about 3.46 which is almost similar to the ratio at s = 2 which is also of 3.46. The strength of the Six Sigma method can be justified by this structural consistency of the server configurations.

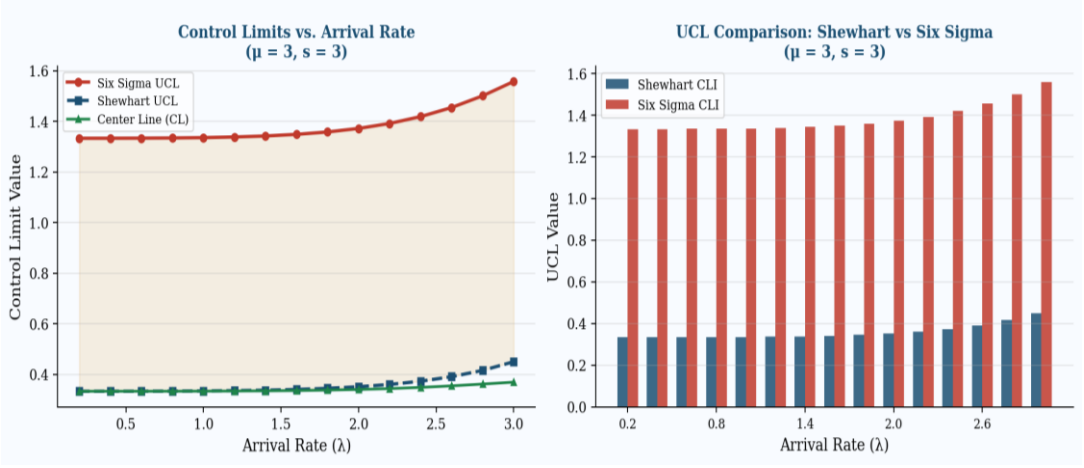

Figure 2: Control limits and CLI comparison for $\mu_s = 3$, s = 3, Left: UCL/CL trends; Right: UCL bar comparison

Figure 2 also substantiates that the six-sigma chart also has a smaller CLI at all s = 3 arrival rates. The left panel indicates the mild inclination of the CL curve of s = 3 as compared to s = 2, which shows the stabilising influence of the extra servers on the mean waiting time. The figure on right shows that the UCL gap is relative constant throughout the entire range of λ.

5.3 Configuration III: $\mu_s = 5$, s = 2

Table 3 looks at the effect of increased service rate (μ = 5) and holding s = 2 servers. The increased service rate significantly decreases all the values of control limits with respect to Configuration I.

Table 3: Control chart parameters for $\mu_s = 5$, s = 2

Arrival Rate (λ)	Service Rate (μ)	Servers (s)	Traffic Intensity (ρ)	P_0	Shewhart Control Chart			Six Sigma Chart ($C_p = 1$)	
					LCL	CL	UCL	LCL	UCL
0.25	5	2	0.0250	1.0538	0.2001	0.0000	0.2001	0.8004	0.1794
0.50	5	2	0.0500	1.1157	0.2003	0.0000	0.2006	0.8016	0.1798
0.75	5	2	0.0750	1.1863	0.2008	0.0000	0.2016	0.8041	0.1808
1.00	5	2	0.1000	1.2663	0.2017	0.0000	0.2031	0.8082	0.1824
1.25	5	2	0.1250	1.3565	0.2031	0.0000	0.2055	0.8147	0.1848
1.50	5	2	0.1500	1.4580	0.2051	0.0000	0.2091	0.8243	0.1883
1.75	5	2	0.1750	1.5720	0.2079	0.0000	0.2141	0.8379	0.1934
2.00	5	2	0.2000	1.6998	0.2118	0.0000	0.2212	0.8567	0.2005
2.25	5	2	0.2250	1.8428	0.2169	0.0000	0.2311	0.8818	0.2103
2.50	5	2	0.2500	2.0030	0.2234	0.0000	0.2445	0.9146	0.2237
2.75	5	2	0.2750	2.1822	0.2310	0.0000	0.2628	0.9559	0.2420
3.00	5	2	0.3000	2.3829	0.2395	0.0000	0.2875	1.0060	0.2668
3.25	5	2	0.3250	2.6075	0.2474	0.0000	0.3209	1.0631	0.3001
3.50	5	2	0.3500	2.8592	0.2520	0.0000	0.3658	1.1219	0.3450
3.75	5	2	0.3750	3.1412	0.2474	0.0000	0.4262	1.1685	0.4054

At λ = 0.25, the center line is only 0.200-time units a 40% decrease of Configuration I at a relative load that is similar. As λ increases to 3.75 (ρ = 0.375), E(W) reaches 0.426, and the Shewhart UCL approaches 1.169. The Six Sigma chart has preserved the CLI advantage throughout this range and the percentage reduction of CLI is close to 75. This proves that the sensitivity value of the Six Sigma methodology is independent of the service rate parameter and varies in a consistent manner with the traffic load.

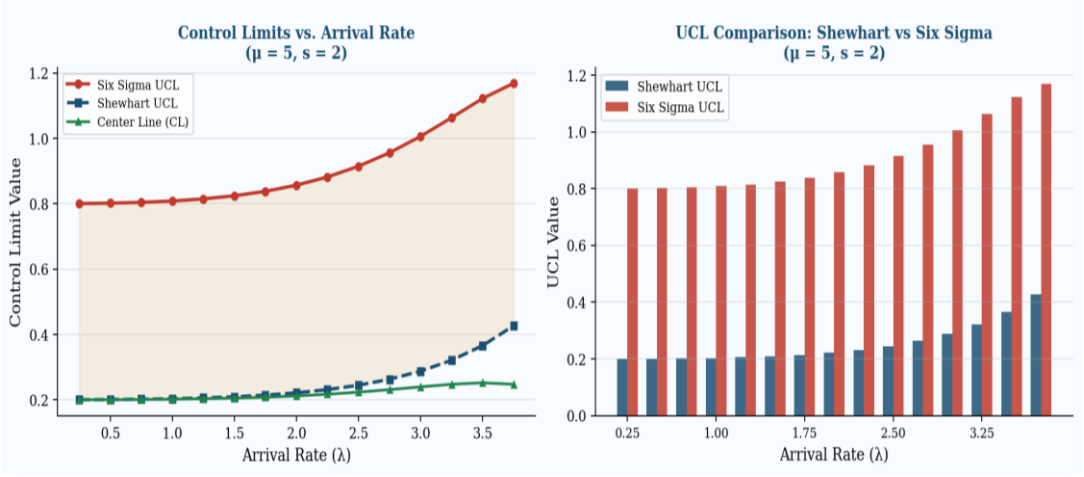

Figure 3: Control limits and CLI comparison for $\mu_s = 5$, s = 2, Left: UCL/CL trends; Right: UCL bar comparison

5.4 Cross-Configuration Analysis

Figure 4 and Figure 5 explains the cross-configuration information.

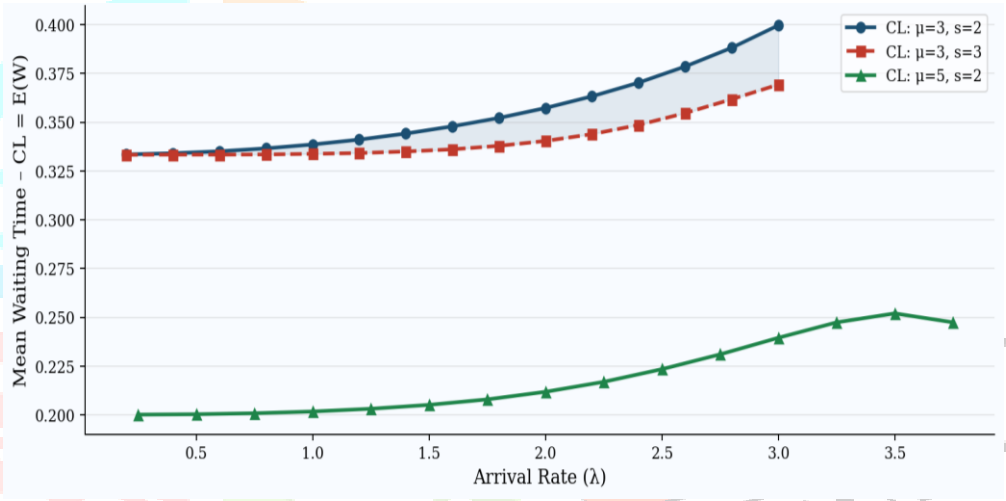

Figure 4: Mean waiting time E(W) across all three system configurations as a function of arrival rate λ

Figure 4 is a direct comparison of the center line E(W) of all three configurations. Configuration I (μ = 3, s = 2) curve is steepest with λ, and this is because of the increased intensity of traffic per server. Configuration II (μ = 3, s = 3) exhibits a slower growth rate because of the second server whereas Configuration III (μ = 5, s = 2) has the lowest absolute E(W) values at low λ but approaches Configuration II at higher loads because of the dissimilar stability boundaries. This number highlights the combined significance of service rate and the number of servers in the regulation of the waiting time dynamics.

The 6σ chart CLI reduction benefit of 66 chart as a function of 66 chart 66 of Configuration I. The decrease is between 73.3% at the low traffic (λ = 0.20) and 75.9% at full load (λ = 3.00) and, as such, the 6σ benefit is not only significant but also slightly rising with the degree of traffic intensity. This trend is due to the fact that the 6σ CLI is pegged to the process capability index (a constant ratio) but the Shewhart CLI increases with $6\sigma_W$ which increases with 6 Sigma lambda.

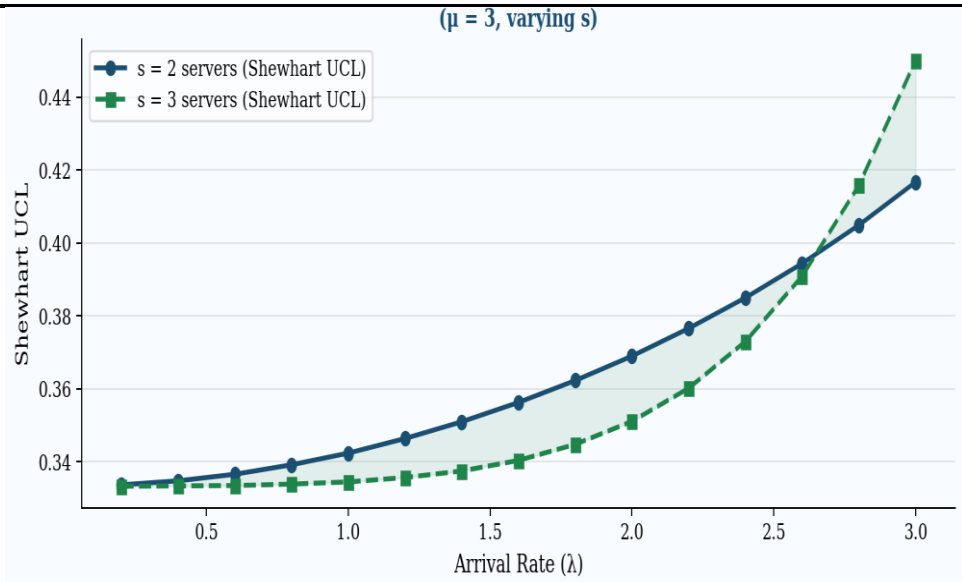

Figure 6: Effect of number of servers on Shewhart UCL (μ = 3, s = 2 vs. s = 3)

The impact of server capacity on the Shewhart UCL is separated in Figure 6. In both cases (s = 3, green and s = 2, blue) the three-server system (s = 3) has a lower UCL than the two-server system (s = 2). The separation of the two curves increases with λ which proves that the extra servers offer a disproportionately higher advantage to a high traffic load within the operating regime where it is most needed, monitoring.

5.5 Summary of Key Findings

- ✓ Six Sigma control chart yields CLI values that are 73-76 % smaller than those of Shewhart in all the configurations.
- ✓ CLI reduction benefit is increasing at a rate of arrival λ so the Six Sigma is most beneficial in the vicinity of the stability boundary.
- ✓ When the number of servers is increased by s=2 to s=3, it decreases both E(W) and UCL at all levels of traffic, with the difference becoming more pronounced when the traffic is heavy.
- ✓ Higher service rate (μ = 5 vs. μ = 3) decreases absolute control limits but has no effect on relative advantage of Six Sigma.
- ✓ LCL = 0 in each configuration, which justifies that upper-limit monitoring is the conceptually operationally relevant target of the waiting time control charts.

6. Discussion

The statistical data provided in Section 5 continue to confirm that the Six Sigma control charts are better than the Shewhart charts in terms of monitoring waiting times in M/M/s systems. In this part, the findings are put into a context of the existing literature and reflect upon their practical implications.

6.1 Statistical Interpretation

The reason why the CLI of the Six Sigma chart is narrower is a direct result of the replacement of σ_C with σ_W in the control limit formula, where σ_C is calculated using a process capability index C_p that imposes a more rigorous performance standard than would be imposed using the empirical standard deviation alone (Radhakrishnan and Balamurugan, 2012; Montgomery, 2005). This is statistically associated with a narrowing in the acceptance region of the characteristic being monitored and increases the likelihood of out-of-control signals (i.e. increases power) at the expense of a possible increased false alarm rate. Practically, TL and C_p can be balanced by service systems with a well-defined service-level agreement (SLAs) that is used to specify the competing goals.

The common metric of performance used to compare the control charts is the Average Run Length (ARL) (Montgomery, 2005; Rigdon et al., 2012). Although the current research is not sufficient

to fully analyze the ARL, the CLI decrease immediately suggests a lower ARL 1 (average time to false alarm) of the Six Sigma chart than of the Shewhart at the cost of a lower ARL 0 (average time between false alarms). The next piece of work should provide these ARL trade-offs in a quantitative way, in the M/M/s waiting time setting.

6.2 Practical Implications

The Six Sigma control charts suggested can be directly applied to the working service-based environment. The shorter limits of Six Sigma can be applied in hospital emergency departments where monitoring of waiting time is a regulatory feature; in this case, the build-up of the queue can be identified in real-time prior to reaching crisis levels (Bischak & Silver, 2014). C_p can be adjusted to the SLA tolerance in banking and retail settings where the service level objective is a contractual requirement, and the monitoring chart is thus statistically principled with a direct connection between the commercial performance standard and the monitoring chart. The Six Sigma framework can be implemented in real-time dashboards with automated notification systems that alert upon occurrence of UCL violations in telecommunications and cloud computing where service queues can be continuously monitored with automated systems (Banks et al., 2010; Mitra, 2016).

6.3 Limitations and Scope

The current research is limited in a number of ways that will inspire the next research. First, it is assumed that the service time distributions are Markovian (exponential), more real service times are typically much more variable (coefficient of variation > 1), which would necessitate formulations of M/G/s and G/G/s (Kleinrock, 1975; Gross et al., 2018). Second, normality approximation of W cannot be good at extremely low traffic loads or very small samples; other distributional assumptions (e.g., Gamma or inverse Gaussian) should be investigated. Third, one can use non-stationary arrival processes (time-varying λ) which is common in practice and would demand adaptive or CUSUM-type charts (Epprecht et al., 2015). Fourth, the current research is not based on a formal ARL analysis; this is one of the main ways of future research.

7. Conclusion

This paper is a rigorous development and numerical validation of control charts of Six Sigma to observe waiting time in M/M/s multi-server queueing systems. The given methodology based on the process capability theory results in control charts where control limit intervals (CLI) are significantly smaller than in the traditional Shewhart charts a 73 to 76 percent decrease in all the experimental settings.

The key conclusions are:

- The Six Sigma control chart is better than the Shewhart chart in terms of CLI width in the M/M/s waiting time monitoring problem, in all combinations of arrival rate, service rate and server count that have been studied.
- The reduction benefit of CLI rises with the intensity of traffic and therefore, Six Sigma monitoring would be particularly beneficial to traffic congested service systems nearing their stability limits.
- The absolute scale of control limits depends on the combination of server capacity (s) and service rate (μ), where extra servers contribute disproportionately large relief at heavy loads.
- The upper control limit (UCL) is the operationally important limit in the monitoring of waiting time, since all practical parameter settings have a lower control limit of zero.

The practical value of this work is that it offers service managers with a statistically rigorous, operationally actionable tool of real time waiting time monitoring. Six Sigma framework fills the gap

between the SPC methodology and process capability thinking by providing a principled mechanism of calibrating the monitoring sensitivity to the service level agreement.

Future directions of research are: (i) generalization to M/G/s and G/G/s models and non-exponential service times; (ii) extension to multi-class priority queueing systems; (iii) formal ARL analysis to measure the detection-vs-false-alarm trade-off; (iv) the empirical validation of the model with real-time data of hospital emergency departments, bank branches, and cloud service queues.

References

1. Antony, J., & Banuelas, R. (2002). Key ingredients for the effective implementation of Six Sigma program. Measuring Business Excellence, 6(4), 20–27. https://doi.org/10.1108/13683040210451679
2. Banks, J., Carson, J. S., Nelson, B. L., & Nicol, D. M. (2010). Discrete-event system simulation (5th ed.). Pearson.
3. Besterfield, D. H. (2013). Quality control (9th ed.). Pearson Education.
4. Bischak, D. P., & Silver, E. A. (2014). Approximate analysis of a transfer-line queueing system. International Journal of Production Research, 52(3), 858–872. https://doi.org/10.1080/00207543.2013.842016
5. Epprecht, E. K., Simoes, B. F. T., & Mendes, F. C. T. (2015). A variable sampling interval EWMA chart for attributes. International Journal of Advanced Manufacturing Technology, 49(1), 281–292. https://doi.org/10.1007/s00170-010-2390-4
6. Erlang, A. K. (1917). Solution of some problems in the theory of probabilities of significance in automatic telephone exchanges. Elektroteknikeren, 13, 5–13.
7. Evans, J. R., & Lindsay, W. M. (2014). Managing for quality and performance excellence (9th ed.). Cengage Learning.
8. Goh, T. N. (2011). Six Sigma in industry: Some observations after twenty-five years. Quality and Reliability Engineering International, 27(2), 221–227. https://doi.org/10.1002/qre.1137
9. Gross, D., Shortle, J. F., Thompson, J. M., & Harris, C. M. (2018). Fundamentals of queueing theory (5th ed.). John Wiley & Sons.
10. Harry, M., & Schroeder, R. (2000). Six Sigma: The breakthrough management strategy revolutionizing the world's top corporations. Currency Doubleday.
11. ITU-T. (2010). Traffic engineering: Determination of traffic offered in the network (Recommendation E.501). International Telecommunication Union.
12. Kendall, D. G. (1953). Stochastic processes occurring in the theory of queues and their analysis by the method of the imbedded Markov chain. Annals of Mathematical Statistics, 24(3), 338–354. https://doi.org/10.1214/aoms/1177728975
13. Khaparde, M. V., & Dhabe, S. D. (2010). Control chart for random queue length in (M/M/1):(∞/FCFS) queueing model. International Journal of Agricultural and Statistical Sciences, 6(1), 319–334.
14. Kleinrock, L. (1975). Queueing systems: Volume I Theory. John Wiley & Sons.
15. Mitra, A. (2016). Fundamentals of quality control and improvement (4th ed.). John Wiley & Sons.
16. Montgomery, D. C. (2005). Introduction to statistical quality control (5th ed.). John Wiley & Sons.
17. Morais, M. C., & Pacheco, A. (2000). On the performance of combined CUSUM charts for the mean. Communications in Statistics Simulation and Computation, 29(1), 153–174. https://doi.org/10.1080/03610910008813606

18. Poongodi, T., & Muthulakshmi, S. (2013). Control chart for waiting time in (M/M/s):(∞/FCFS) queueing model. Journal of Mathematics, 5(6), 48–53.
19. Pukazhenthi, S., & Poornima, R. (2018). Construction of control charts for waiting time in (M/M/s):(∞/FCFS) queueing model using process capability. Journal of Emerging Technologies and Innovative Research, 5(11), 490–499.
20. Pyzdek, T., & Keller, P. A. (2014). The Six Sigma handbook (4th ed.). McGraw-Hill Education.
21. Radhakrishnan, R., & Balamurugan, P. (2012). Construction of control charts based on Six Sigma initiatives for fraction defectives with varying sample size. Journal of Statistics & Management Systems, 15(4–5), 405–413. https://doi.org/10.1080/09720510.2012.10701632
22. Rigdon, S. E., Fricker, R. D., & Woodall, W. H. (2012). Performance metrics for surveillance schemes. Quality Engineering, 24(1), 28–36. https://doi.org/10.1080/08982112.2012.627289
23. Ross, S. M. (2014). Introduction to probability models (11th ed.). Academic Press.
24. Shewhart, W. A. (1931). Economic control of quality of manufactured product. D. Van Nostrand Company.
25. Shore, H. (2000). General control charts for attributes. IIE Transactions, 32(12), 1149–1160. https://doi.org/10.1080/07408170008967460