International Journal of New Computer Architectures and their Applications (IJNCAA) 6(1): 23-33
The Society of Digital Information and Wireless Communications, 2016 ISSN 2220-9085 (Online); ISSN 2412-3587 (Print)

Evaluation of Performance of Smart Devices in Closed System Models

Cecilia Twinamatsiko[1], Kyanda Swaib Kaawaase[2], Michael Okopa[3]
Makerere University[1,2], International University of East Africa[3]
Kampala, Uganda
hopececilia1@gmail.com, kswaibk@cis.mak.ac.ug, okopamike@gmail.com

ABSTRACT

Smartphones and tablets are becoming dominant devices in the present market. Apart from offering the same features as traditional mobile phones, smart devices provide several other features as well. Designing Smartphones to balance the tradeoff between minimum power dissipation and increased performance becomes a challenge. This is due to the fact that users try to optimize for mobility and lightness rather than for maximum computing performance and throughput. In an effort to study the performance of smart devices, we modeled the components of smart devices using an $M/M/1//N$ closed queue system. In this model, the number of requests in the system is fixed. The performance of requests under the closed and open system models are compared in terms of average response time. We observe that average response time generally increases with increase in arrival rate, and increase in load values. It is further observed that requests experience lower average response time under the open system model at low arrival rates, low service times and low load compared to closed system model. However, at high arrival rates, high service rates and high load values, requests experience lower response time under the closed system model than under the open system model. The analytical results indicate that closed system models offer better performance at high arrival rates and high load values, whereas open system models offer better performance at low arrival rates and low load values.

KEYWORDS

Closed system model, open system model, smart devices, embedded systems, workload generators

1 INTRODUCTION

Despite the market's heterogeneity, smart devices, wireless broadband, and network-based cloud computing constitutes a perfect storm of opportunity for application developers, luring their attention towards the new platforms [1].
Smart devices are becoming dominant devices in the present market. Apart from offering the same features as traditional devices, smart devices provide several other features. Some of these features include amongst others high speed access to Internet using Wi-Fi and mobile data network, running several applications & games, capturing & sharing pictures to social platforms, audio-video capturing and sharing [2].
The performance limitation of smart devices originate mainly from the fact that they have embedded systems [3]. An embedded system is a computer system made for specific control functions inside a larger system frequently having real-time computing constraints.
Normally, embedded systems have different design constraints than desktop computing applications. Instead embedded systems try to optimize for mobility and lightness rather than for maximum computing performance and throughput. Since many users want their smartphones to be more powerful, have high throughput, and at the same time, small and light, it presents a constraint to designers [4]. This constraint acts as a recipe to analyze the factors that affect system performance, for example, number of processor cores, clock frequency, processor types, load, service rate, arrival rate and so on.

System researchers are aware of the importance of representing a system accurately and this involves accurately representing the scheduling of requests, service distributions, arrival of requests into systems, etc. Representing systems accurately involves many things, including accurately representing the scheduling of requests, service distributions, correlation between requests, etc. One factor that researchers pay little attention to is whether the job arrivals obey a closed or an open system model.

In a closed system model, new job arrivals are only triggered by job completions [3]. In addition, in a closed system model, it is assumed that there is some fixed number of users, who use the system forever. These users are called the multiprogramming level (MPL) [3, 5, 6]. Examples of closed systems include TPC-W [8] used as a database benchmark for e-commerce, RUBiS [9] which is an Auction website benchmark, AuthMark [10] which is a Web authentication and authorization benchmark, etc.

On the other hand, in an open system model, new jobs arrive independent of job completions [11, 12]. In an open system model there is a stream of arriving users and each user is assumed to submit one job to the system, wait to receive the response, and then leave. The differentiating feature of an open system is that a request completion does not trigger new requests, instead a new request is only triggered by a new user arriving.

The differences between open and closed system models motivate the need for researchers to investigate the performance of devices in either open or closed systems. This background therefore provides a simple recipe for investigating the performance of smart devices in closed system models.

2 RELATED WORK

Schroeder *et al.* [3] studied the performance of closed versus open workload generators under many applications, like static and dynamic web servers, database servers, auctioning sites, and supercomputing centers. The study noted that there was a huge difference between the two systems. For example, under a fixed load, the mean response time for an open system model can exceed that for a closed system model by a high magnitude. In addition, while scheduling to favour short jobs is extremely effective in reducing response times in an open system, it has very little effect in a closed system model. The deduction is that variability in the job sizes has a higher effect in an open system than a closed one.

Bondi et al [14] studied a general network of FCFS queues and concluded that service variability is more dominant in open systems and less pronounced in closed systems (provided the MPL is not too large). However, this study was primarily restricted to FCFS queues.

In a similar study, Schatte et al. [15, 16] studied a single FCFS queue in a closed loop with think time. The study revealed that, as the MPL grows to infinity, the closed system converges monotonically to an open system. This result provides a fundamental understanding of the effect of the MPL parameter.

M. Choi *et al.* [1] developed a queuing model to analyze the performance of the components of Smart devices in terms of average waiting time and mean queue length using an open system model. The components analyzed include; Central processing Unit (CPU), RAM, GSM, LCD, and Graphics. The authors assumed an open queuing system model. Although many applications can be modeled using an open system model, open system models do not capture interactivity in systems which is well captured by closed system models. For example, interactive behaviors like human users interacting with a system, threads contending for a lock, and networked servers waiting for a response message are so common. In a closed network, it is possible to model a set of users submitting requests to a system, waiting for results, and

then submitting more requests.
The expressions for average response time and number of requests in the system respectively are derived in [1] and given below:
Average response time, $E[S]$ is given as:

$$E[S] = \frac{1}{\mu(1-\rho)} = \frac{\rho}{\lambda(1-\rho)} \quad (1)$$

Similarly, the mean number of requests, $E[X]$ is given as:

$$E[S] = \frac{\rho}{(1-\rho)} = \frac{\lambda}{(\mu-\lambda)} \quad (2)$$

The differences between open and closed system models shown across a range of applications motivates the need for system designers to be able to determine how to choose if an open or closed model is more appropriate for the system they are targeting. Although many applications can be modeled using an open system model, open system models do not capture interactivity in systems which is well captured by closed system models.
In this paper, we model the performance of Smart devices in a closed system in terms of average response time and mean queue length so as to capture the interactivity.
The contributions of this paper are two fold. Firstly, we developed closed system model using queuing theory to analyze the performance of Smart devices in terms of average response time and mean queue length. Secondly, this research evaluated the performance of the closed system models while comparing it to the open system models.
The rest of the paper is organized as follows: in the next section, we present the system models. We evaluate the proposed models in section 4 and finally conclude the paper in section 5.

3 SYSTEM MODELS

We consider a closed system model, where new job arrivals are only triggered by job completions as shown in figure 1. In a closed system model, new job arrivals are only triggered by job completions.

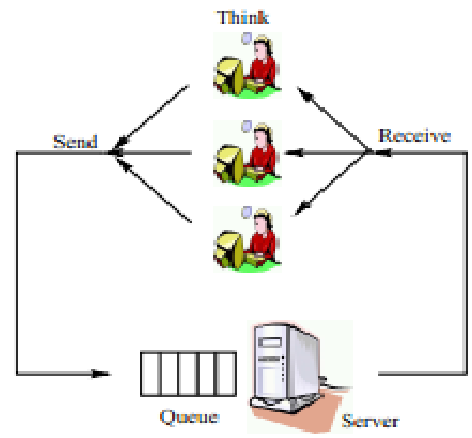

Figure 1. Closed system. Adopted from [3]

In a closed network, we can model a set of users submitting requests to a system, waiting for results, then submitting more requests. Examples of closed systems include TPC-W [8] used as a database benchmark for e-commerce, RUBiS [9] which is an Auction website benchmark, AuthMark [10] which is a Web authentication and authorization benchmark, etc. In addition, we consider the following;

(a) a smart phone with the following components; CPU, RAM, graphics, GSM, LCD, Backlight, and Rest.

(b) we assume that the throughput is known (to be equal to the arrival rate), and also that there is no probability of incomplete transfer in this system, so there is no retrial path to go back.

(c) Although the CPU components of recent smartphones can have more than one CPU, known as dual-core or quad-core, however we assume a smartphone with single-core in this study.

In the next section, we derive the expression for average response time and average queue length.

3.1 Expression for average response time and average queue length

Specifically, we use an $M/M/1//N$ queue system, where the first M denotes Poisson distribution which best models random arrivals into systems. The *pdf* (probability density function) of a Poisson distribution is given by [13]:

$$P(x = k) = \frac{\lambda^k e^{-\lambda}}{k!} \tag{3}$$

The second M represents an exponentially distributed service time. The *pdf* (probability density function) of an exponential distribution is given as [13]: $f(x) = \mu e^{-\mu x}, \quad x \geq 0, \ \mu \geq 0$, where μ is the mean service rate. One server and FCFS service discipline are assumed for the system model, while at the same time the population of the system is finite.
We used the tagged job approach where we track the experience a job undergoes, i.e, when a new customer arrives to the queue, there will be some number of other customers already present and waiting, including a customer that is in service.

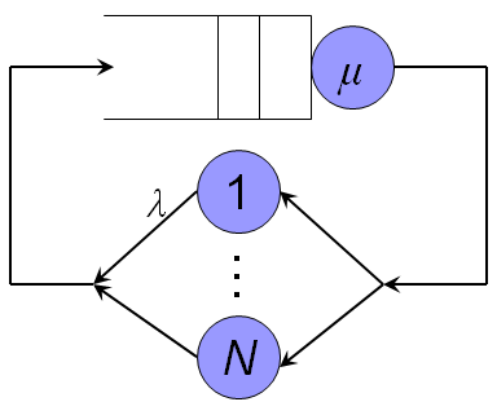

Figure 2. $M/M/1//N$ closed queue system

Figure 2 shows an $M/M/1//N$ closed queue system. The corresponding state transition diagram is given in figure 3.

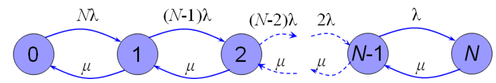

Figure 3. State transition diagram for $M/M/1//N$ closed queue system

Using the birth-death result we can obtain individual probabilities as:

$$\mu\pi_1 = N\lambda\pi_o \tag{4}$$

$$\mu\pi_2 = (N-1)\lambda\pi_1 \tag{5}$$

$$\mu\pi_3 = (N-2)\lambda\pi_2 \tag{6}$$

Combining equations 4, 5, and 6 we obtain;

$$\pi_3 = N(N-1)(N-2)(\frac{\lambda}{\mu})^3 \tag{7}$$

Equation 7 can be generalized to any probability state i as:

$$\pi_j = \frac{N!}{(N-j)!}.\rho^j.\pi_o, j = 1, 2, ..., N. \tag{8}$$

Using the fact that total probability is one we obtain;

$$\sum_{j=1}^{N} \frac{N!}{(N-j)!}.\rho^j \pi_o = 1 \tag{9}$$

and hence,

$$\pi_o = \left[\sum_{j=0}^{N} \frac{N!}{(N-j)!}.\rho^j\right]^{-1} \tag{10}$$

Using little's law [13], the average queue length, $E[X]$ is given by;

$$E[X] = \mu(1 - \pi_o)E[S] \tag{11}$$

where $E[S]$ is the average response time. For the thinking part of the closed system,

International Journal of New Computer Architectures and their Applications (IJNCAA) 6(1): 23-33
The Society of Digital Information and Wireless Communications, 2016 ISSN 2220-9085 (Online); ISSN 2412-3587 (Print)

$E[N - X] = \mu(1 - \pi_o).\frac{1}{\lambda}$.
$E[N] - E[X] = \mu(1 - \pi_o).\frac{1}{\lambda}$, from which we obtain $E[X] = N - \mu(1 - \pi_o).\frac{1}{\lambda}$.
Using equation 11 we obtain;

$$E[S] = \frac{E[X]}{\mu(1 - \pi_o)} \tag{12}$$

Therefore, the average response time is given as;

$$E[S] = \frac{N - \mu(1 - \pi_o).\frac{1}{\lambda}}{\mu(1 - \pi_o)} = \frac{N}{\mu(1 - \pi_o)} - \frac{1}{\lambda} \tag{13}$$

where π_o is as given in equation 10.
Similarly, the average number of requests in the system can be deduced from little's theorem as follows: $E[X] = \lambda E[S]$. Dividing equation 13 by λ we obtain the expression for the average queue length of requests in the system is given as:

$$E[X] = \left(\frac{N.\rho}{(1 - \pi_o)} - 1\right), \tag{14}$$

where π_o is as given in equation 10.
In the next section, we present the performance evaluation of closed against open system models.

4 PERFORMANCE EVALUATION

In this section, we use the derived models to evaluate its performance. In particular, we analyze the variation of average response time and average queue length with arrival rate of requests, load in the system, and service rate. In each case, we compare the performance under closed and open system models.
For numerical evaluation we considered the following hypothetical data, these data is the same as the configurations used in [1]. Average arrival rate used is 1000 to 1500 requests/second, average service rate is 1500 to 1800 trequests/second, load used is in the range 0 to 1.

4.1 Comparison of open and closed systems in terms of average response time

In this section, we compare the performance of requests under open and closed system models in terms of average response time. We used equations 1 and 13 to plot the following graphs.

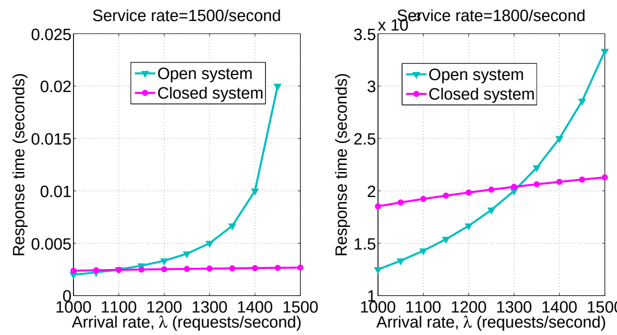

Figure 4. Response time against arrival rate.

Figure 4 shows the variation of average response time against arrival rate of requests into the system. We observe that average response time increases with increase in arrival rate regardless of the service rate. We further observe that requests experience lower response time under the open system at low arrival rates, however at high arrival rates, requests experience lower response time under closed systems. The difference in performance of requests between open and closed systems is more pronounced at high service rate compared to low service rate. This is shown when service rate of 1500 requests/second is used compared to 1800 requests/second. Whereas the performance of requests under open and closed systems are the same at arrival rate of 1100 requests/second when the service rate is 1500 requests/second, it is the same at arrival rate of 1300 requests/second when the service rate is 1800 requests/second.
Figure 5 shows the graph of average response time against load for different service rates. We observe that average response time increases with increase in load in the system regardless of the service rate. We also note that requests experience lower average response time under the open system at low load, however when the

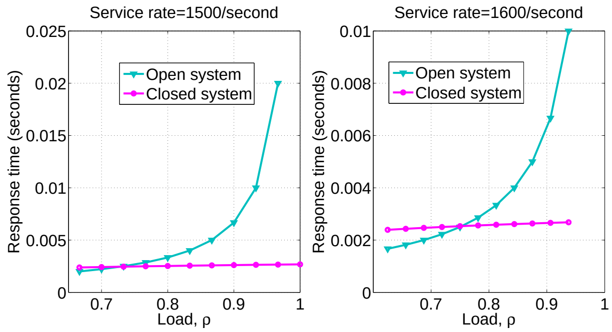

Figure 5. Response time against load.

load increases, requests experience lower average response time under the closed system model. The difference in performance between open and closed system is more pronounced at high service rate compared to low service rate.

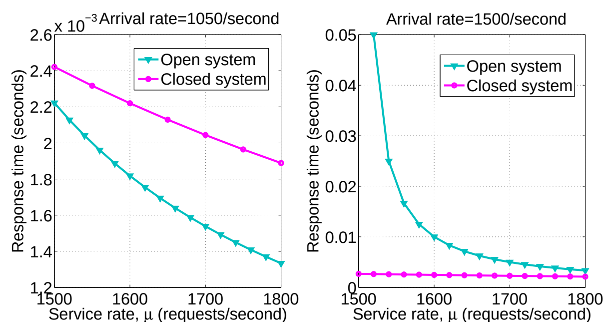

Figure 6. Response time gainst service rate.

Figure 6 shows the graph of response time against service rate for different arrival rates. We observe from the graph that response time decreases with increase in service time. In other words, as the service rate increases, the response time decreases. Increase in service rate implies increase in the number of requests served per unit time, hence decrease in average response time. We also observe that requests experience lower response time under the open system at low service rate, however at high service rate, requests experience lower response time under the closed system compared to under the open system.

4.2 Comparison of open and closed system models in terms of average queue length

In this section, we compare the performance of requests under open and closed system models in terms of average queue length. We used equations 2 and 14 to plot the following graphs.

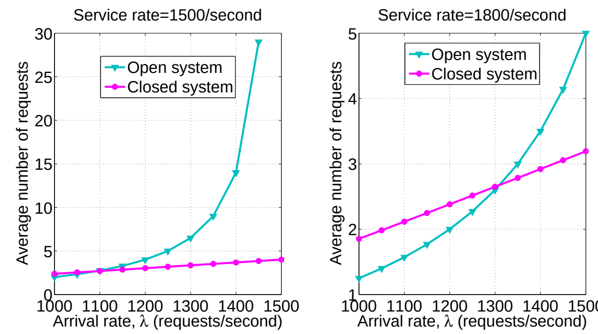

Figure 7. Average queue length against arrival rate.

Figure 7 shows the variation of average queue length against arrival rate of requests into the system. We observe that average queue length increases with increase in arrival rate regardless of the service rate. We further observe that the number of requests in the system under the open system model is lower than under the closed system model at lower arrival rates. However, at higher arrival rates, the number of requests in the system under the open system model is higher than under the closed system model. Whereas the performance of requests under open and closed systems are the same at arrival rate of 1100 requests/second when the service rate is 1500 requests/second, it is the same at arrival rate of 1300 requests/second when the service rate is 1800 requests/second.
Figure 8 shows the graph of average queue length in the system against the load. We observe that the number of requests in

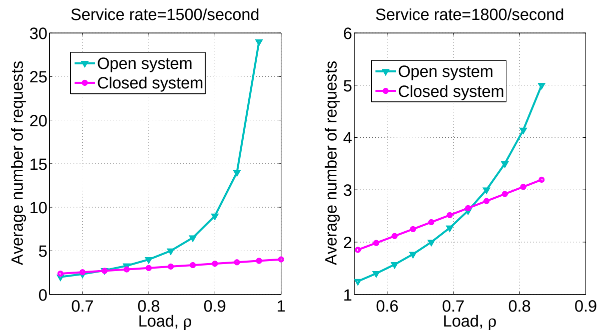

Figure 8. Average queue length against load.

the system increases with increase in load regardless of the service rate. We further observe that the number of requests under the open system model is lower than under the closed system model at lower load values. On the other hand, the number of requests under the open system model is higher than under the closed system model at higher load values. Figure

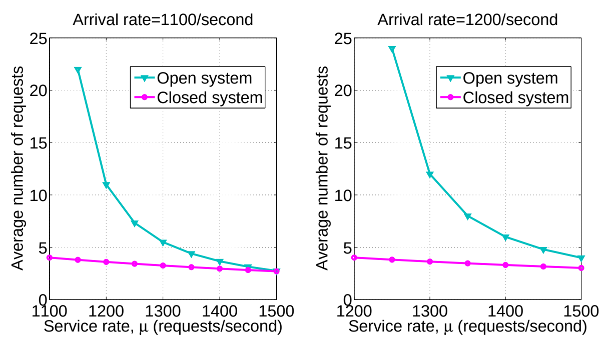

Figure 9. Average queue length against service rate.

9 shows the graph of average queue length in the system against the service rate for different arrival rates. We observe from the graph that the average queue length decreases with increase in service time. We also observe that the average queue length in the system under the open system is generally higher than under the closed system model.

4.3 Performance under the closed system model

In this section, we investigate the effect of varying parameters on the performance of requests under the closed system model. We used equation 13 and equation 14 to plot a graph of average response time and average queue length against arrival rate, service rate, and load while varying the number of requests in the system.

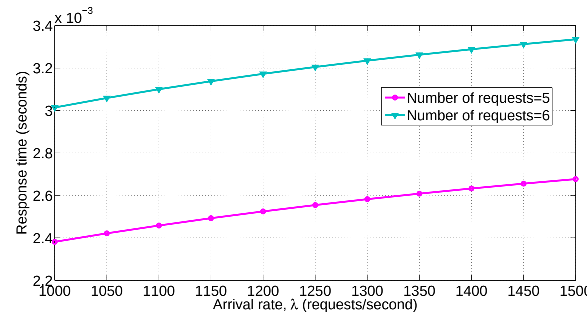

Figure 10. Response time against arrival rate.

We observe from figure 10 that average response time increases with increase in arrival rate. We further observe that the average response time is higher when the number of requests in the system is higher as compared to when the number of requests in the system is lower. For example, when the arrival rate of requests into the system is 1300 requests/second, the average response time when the number of requests in the system is 5 is approximately $2.6 * 10^{-3}$ seconds while when the number of requests in the system is 6, its approximately $3.2 * 10^{-3}$ seconds. The increase in average response time as arrival rate increases is due to the fact that increase in arrival rate leads to increase in the number of requests in the system and therefore an increase in average response time.

We observe from figure 11 that the average response time of requests generally decreases with increase in service time. We further observe that average response time is lower when the number of requests in the system is lower compared to when it is high. When the arrival rate is high, the average response time is lower. On the other hand, when the average arrival rate is low, the average response time is lower.

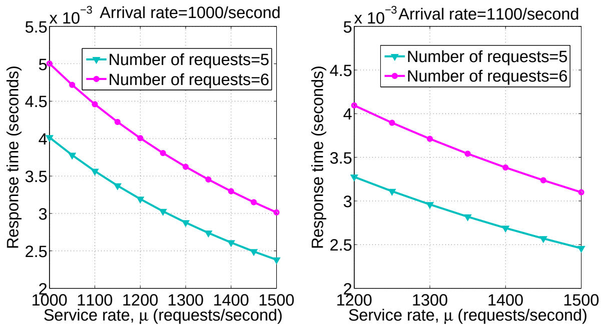

Figure 11. Response time gainst service rate.

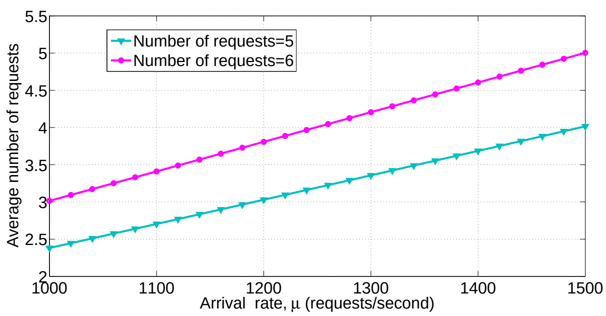

Figure 12. Average queue length against arrival rate.

We observe from figure 12 that the average queue length increases as the arrival rate increases. In addition, the average queue length is higher when the number of requests in the system is high compared to when it is low. For example, when the arrival rate is 1200 requests/second, and the number of requests in the system is 5, the average queue length is observed to be 3. On the other hand, when the number of requests in the system is 6, the average queue length is 3.8 requests.

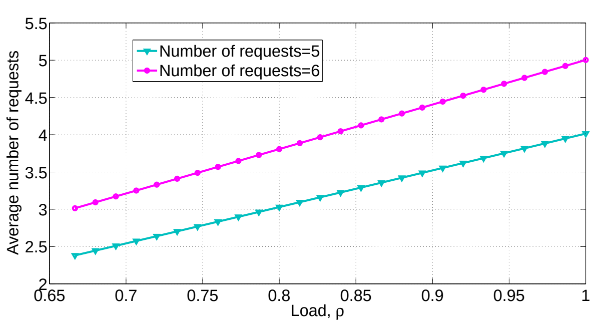

Figure 13. Average queue length against load.

We observe from figure 13 that average queue length generally increases with increase in load. We further observe that the average queue length is higher when the number of requests in the system is high compared to when it is low. For example when the load is 0.9, and the number of requests in the system is 5, the average queue length is 3.5 requests. On the other hand, when the load is 0.9, and the number of requests in the system is 6, the average queue length is 4.4 requests.

We observe from figure 14 that the average queue length decreases with increase in service rate regardless of the number of requests in the system. We also observe that the average queue length is lower when the number of requests in the system is low compared to when it is high. This is noted when the number of requests in the system is 5 compared to when it is 6. For example, when the service rate is 1200 requests/second, the average queue length when the number of requests in the system is 5 is 3.6 requests whereas when the number of requests in the system is 6, the average queue length is 4.5 requests.

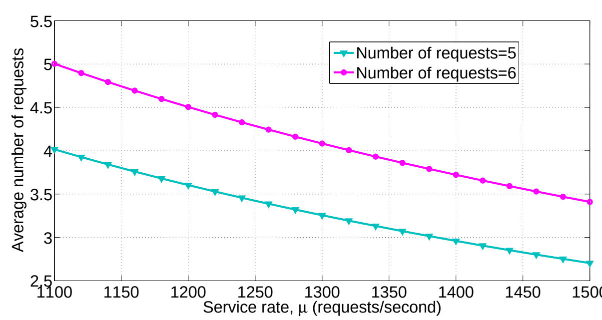

Figure 14. Average number of requests against service rate

5 CASE STUDY FROM FRANCE FOOTBALL WORLD CUP 1998 TRACE

In this section, we conducted performance test on smartphones.

We used trace data from Internet sites serving the 1998 World Cup. The data used are available from the Internet Traffic Archive (see [17] and http://ita.ee.lbl.gov/html/traces.html). This repository provides detailed information about the 1.3 billion requests received by World Cup sites over 92 daysfrom April 26, 1998, to July 26, 1998. We assume that all requests submitted must first pass through the Internet sites for providing HTTP service before moving on to the smart devices.
The data trace indicated in [17] exhibits arrival-rate fluctuations corresponding to light, medium and heavy loadings in this temporal order. More specifically, the trace allows us to study performance under various load conditions, as follows:

- Light load. In the time interval [0, 1200], the arrival rate is relatively low (below 1200 requests/s), and the resultant utilization is also low (40%).
- Medium load.In the time interval (1200, 2400], the arrival rate is between 1200 and 1800 requests/s, and the resultant utilization is intermediate (50%).
- Heavy load. In the time interval (2400, 3600], the arrival rate often exceeds 2000 requests/s, and the corresponding utilization is high (75%).

We use two performance metric to test the performance of smart devices, that is, average response time and mean queue length.

5.1 Comparison of open and closed systems in terms of average response time

In this section, we compare the performance of open and closed systems in terms of average response time using trace data from Internet sites serving the 1998 World Cup.

5.1.1 Performance under light load

Under light load, we consider request average arrival rate of below 1200 requests/s and service rate of 1500 requests/s. Figure 15

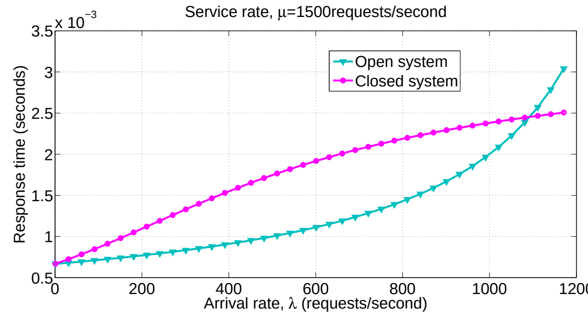

Figure 15. Response time against arrival rate.

shows the variation of average response time against arrival rate of requests into the system when the load is light. We observe that average response time generally increases with increase in arrival rate. We note that requests experience lower average response time under open systems at lower arrival rates, however at higher arrival rates, requests experience lower response time under closed systems. The difference in performance of requests between open and closed systems is more pronounced at arrival rate of between 400 requests/second and 900 requests/s. On the other hand, the performance of requests under open and closed systems are the same when the arrival rate is approximately 1100 requests/s.

5.2 Performance under medium load

Under medium load, we consider request average arrival rate of between 1200 and 1800 requests/s and service rate of 1800 requests/s to ensure the arrival rate is less than service rate ($\lambda < \mu$), otherwise the queue grows infinitely long. Figure 16 shows the variation of average response time against arrival rate of requests into the system under medium load. We observe that average response time generally increases with increase in arrival rate. We further observe that requests experience lower average response time under open systems at lower arrival rates, however at higher arrival rates, requests experience lower response time under

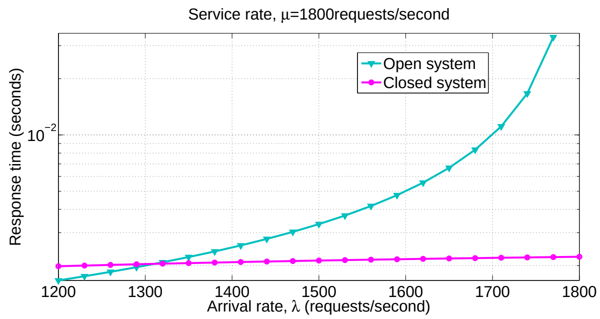

Figure 16. Response time against arrival rate.

closed systems. The difference in performance of requests between open and closed systems is more pronounced at higher arrival rates. It is also observed that the performance of requests under open and closed systems are the same when the arrival rate is approximately 1300 requests/s, after the arrival rate of 1300 requests/s, requests perform better under closed systems.

5.3 Performance under heavy load

Under heavy load, we consider request average arrival rate of between 2000 and 9000 requests/s and service rate of 9000 requests/s to ensure the arrival rate is less than service rate ($\lambda < \mu$), otherwise the queue grows infinitely long. Figure 17 shows the variation of average

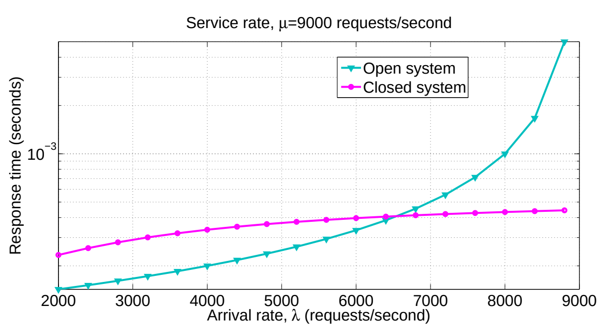

Figure 17. Response time against arrival rate.

response time against arrival rate of requests into the system under heavy load. We observe that average response time generally increases with increase in arrival rate.

We further observe that requests experience lower average response time under open systems at lower arrival rates, however at higher arrival rates, requests experience lower response time under closed systems. It is further observed that the performance of requests under open and closed systems are the same when the arrival rate is approximately 6500 requests/s, after the arrival rate of 6500 requests/s, requests perform better under closed systems.

6 CONCLUSION

In an effort to study the performance of smart devices, we modeled the components of smart devices using an $M/M/1//N$ closed queue system. In this model, the number of requests in the system is fixed. The performance of requests under the open and closed system models are compared.

We observe that average response time generally increases with increase in arrival rate, and increase in load values. The average response time however, decreases with increase in the service rate. It is further observed that requests experience lower average response time under the open system model at low arrival rates, low service times and low load. However, at high arrival rates, high service rates and high load values, requests experience lower response time under the closed system model.

A similar trend is observed with the average queue length which generally increases with increase in arrival rate, and increase with increase in load values. The average queue length however decreases with increase in the service rate. It is further observed that requests experience lower average queue length under the open system model at low arrival rate, low service rate and low load. However, at high arrival rate, high service rate and high load values, requests experience lower average queue length. Therefore, closed system models is observed to offer better performance at high arrival rates and high load values than open

system models.
These results are also confirmed when trace data from Internet sites serving the 1998 World Cup was used. The analytical results indicate that closed system models offer better performance at high arrival rates and high load values, whereas open system models offer better performance at low arrival rates and low load values.

ACKNOWLEDGMENT

This work was partially funded by the International University of East Africa, Kampala, Uganda Research Fund.

REFERENCES

[1] M. Choi, ”Power and Performance Analysis of Smart Devices,” in Proc. International Journal of Smart Home, vol. 7, no. 3, May 2013.

[2] ”http://www.gartner.com/newsroom/id/1924314 ,” Feb. 2012, press release.

[3] B. Schroeder, A. Wierman, and M. Harchol-Balter, ”Open Versus Closed: A Cautionary Tale,” in Proc. the 3rd conference on Networked Systems Design & Implementation, vol. 3, 2006, pp. 8-18.

[4] J. Kro, A. Brunnert, C. Prehofer, T.A. Runkler, H. Krcmar, ”Model-based Performance Evaluation of Large-Scale Smart Metering Architectures,” ACM 2015, PP. 9-12.

[5] K. Kant, V. Tewari, and R. Iyer, ”GEIST: Generator of e-commerce and internet server traffic,” in Proc.International Symposium on Performance Analysis of Systems and Software, 2001.

[6] P. Barford and M. Crovella, ”The surge traffic generator: Generating representative web workloads for network and server performance evaluation,” in Proc.the ACM SIGMETRICS, 1998.

[7] E. Nahum, M. Rosu, S. Seshan, and J. Almeida, ”The effects of wide-area conditions on www server performance,” in Proc. ACM SIGMETRICS, 2001.

[8] ”Transaction Processing Performance Council, TPC benchmark W(webcommerce),” Feb. 2002, number Revision 1.8.

[9] C.Amza, E.Cecchet, and W. Zwaenepoel, ”Specification and implementation of dynamic website benchmarks,” 2002.

[10] Mindcraft, ”The authmark benchmark http://www.mindcraft.com/authmark/,” Feb. 2002.

[11] M. Yuksel, B. Sikdar, K. S. Vastola, and B. Szymanski, ”Workload generation for ns simulations of wide area networks and the internet,” in Proc.Comm. Net. and Dist. Sys. Mod. and Sim, 2000.

[12] Z. Liu, N. Niclausse, and C. Jalpa-Villanueva., ”Traffic model and per-formance evaluation of web servers,” in Proc.Performance Evaluation, 2001, pp. 77-100.

[13] L. Kleinrock, Queueing Systems, Volume I& II. Computer Applications. John Wiley & Sons, 1976.

[14] A. B. Bondi and W. Whitt, ”The influence of service-time variability in a closed network of queues,” Proc. Performance Evaluation, 1986, PP. 219-234.

[15] P. Schatte, ”On conditional busy periods in n/M/GI/1 queue,” Proc. Mathematics Operations for Optimization, 1983.

[16] P. Schatte, ”The M/GI/1 queue as limit of closed queueing systems,” Proc. Mathematics Operations for Optimization, 1984.

[17] A. Martin and J. Tai, ”Workload Characterization of the 1998 World Cup Web Site”, IEEE, 2000. pp.30-37.