Sensitivity of Blocking-based Adaptive Capacity Control in a Dynamic Traffic Network Environment

Brenda Groskinsky and Deepankar Medhi
Department of Computer Networking, University of Missouri-Kansas City

ABSTRACT

We consider a dynamically reconfigurable network environment where dynamically changing traffic is offered. Rearrangement and adjustment of network capacity can be performed to maintain Quality of Service (QoS) requirements for different traffic classes in the dynamic traffic environment. In this work, we specifically consider the case of a single, dynamic traffic class scenario in a loss mode environment. We have developed a numerical, analytical tool which models the dynamically changing network traffic environment using a time-varying, fluid-flow, differential equation that can be used to study the impact of an adaptive capacity adjustment control scheme. We present the results for a simple, blocking-based capacity adjustment, control scheme. In particular, we show that a purely blocking-based capacity adjustment control scheme can be very sensitive to capacity changes, and can lead to network instability.

Keywords: Dynamic Traffic, Capacity Control, Fluid-Flow Differential Equation, Stability

1. INTRODUCTION

Network traffic tends to show dynamic behavior due to time of the day factor, burst in traffic due to a failure and so on. An important issue is how to maintain the quality of service in the dynamic traffic environment. An approach, for this case, is to engineer the network for a given QoS by allocating capacity for the highest load over a time window (e.g. 24-hour period). The down side of this approach is that the capacity may be vastly underutilized when the load is significantly below the peak load within the time window. With the advent of dynamically reconfigurable networking idea and the virtual network concept, it will be possible to allocate and de-allocate capacity from the underlying transport network for different virtual networks to respond to network traffic changes so as to meet and continue to meet QoS requirements. For instance, virtual networks may be defined for different service classes and differing QoS requirements. Thus, if the traffic is low for a particular service class, at a certain point of time, then, as long as QoS is met, some capacity may be de-allocated for use by other services or assigned to a shared pool for use by other services. Similarly, the reverse situation is also possible. In addition, if network resources are affected due to a component failure in the virtual network for a certain traffic class, then even additional capacity to maintain acceptable QoS under the failure situation may be requested from the transport network. Such dynamic capacity adjustment, in response to or in anticipation of changes in traffic, can lead to better use of network resources, since idle capacity for a service class can be de-allocated for use by another service class.

While the idea of dynamic adjustment of capacity to address dynamically changing traffic is appealing, there is limited work on understanding how the system behaves, and/or how often to perform updates, and/or what specific control strategies should be used for updating. Towards this end, we have considered a system with a single traffic class that has dynamic (time-based) traffic and traffic arrival is considered at the level of the connection/call arrival. The QoS parameter considered is the connection/call blocking and we assume that the virtual network for this traffic class operates in a loss system; i.e. if a connection does not find capacity, the connection is blocked and cleared from the system. While there is some work that proposes update control schemes, there is virtually no work that models the system behavior along with update strategies, especially to observe network performance, even for a single traffic class system. In this work, we present a numeric, first-order, differential-based, fluid-flow model that captures the traffic dynamics and where various adaptive capacity adjustment strategies can be implemented. We then provide computational results on the behavior of the system, and specifically, show that under a specific control scheme the allocation may lead to an oscillatory behavior.

It is worth noting that, a network operating in a connection-blocking environment where capacity allocation/de-allocation is possible, it may be tempting to request additional capacity from the transport network as soon as a new connection for this traffic class does not find capacity within its currently specified, allocation of bandwidth for this traffic class. This situation will usually require the generation of a set of signaling messages to request additional capacity from the transport network and execution of an update function and at the same time, the specific connection request will still be waiting to be connected. This wait for the virtual capacity update function will lead to additional increase in the connection set-up time over the normal procedure. Thus, in our system, we consider the case where the capacity adjustment is done independent of the arrival of a specific connection; rather, it is dependent on the overall arrival and change in traffic.

We now briefly discuss the existing literature. Analyzing network performance under non-stationary traffic has been addressed for a queuing system using a fluid-flow-based approach[4,11]. A good summary on fluid-flow approach for different queueing systems with non-stationary traffic can be found in Wang *et. al.*[12]. These works do not address control schemes. In Qian *et. al.*[9], admission control with non-stationary traffic is addressed using Chapman-Kolmogorov's equation for a multi-rate loss system. None of this sited research however, considers the dynamic adjustment of available capacity. Research has been done which proposes connection blocking, capacity adjustment, based on changing traffic conditions[1,5,8,10]. From a network dimensioning perspective, it has shown that the overall network capacity requirement is less under a dynamic traffic condition using the dynamic virtual path concept in ATM networks as compared to static allocation[7]. To our knowledge, models to study the dynamic traffic case, under a dynamically, reconfigurable, capacity environment, have not received much attention.

2. MODEL

In order to study the effects of dynamic capacity adjustment under a dynamic traffic scenario, we consider a non-stationary, offered load to a single-link, loss system where the capacity adjustment is time-dependent. The dynamic offered load is the ensemble offered load, $a(t)$ erlangs, at time t. The connection holding time is assumed to be exponentially-distributed, with mean, $1/\mu$. Thus, $\lambda(t) = a(t)\mu$ is the ensemble arrival rate at time, t, and follows a time-dependent, Poisson process. The connection arrivals find the capacity, at time, t, to be $C(t)$, expressed as discrete integer values. If there is enough available capacity allotments, out of the present value of $C(t)$, to accommodate the connection arrivals, then the connections are accepted and used as an unit of bandwidth. Otherwise, the connections are blocked and cleared. $C(t)$ is time-dependent and represents the adjustable capacity in the network link but is independent of the offered load, $a(t)$. This allows us to update or adjust $C(t)$ based on a specific time-dependent, control scheme. Thus we have a time-dependent loss system $M(t)/M/C(t)/C(t)$.

Let $x(t)$ represent the average number of connections that are present on the link at time, t. Interpreting the "flow" of connections made to the link, $f_i(t)$, and the "flow" of disconnections from the link, $f_o(t)$, as a dynamic, fluid-flow model[3,11], we can describe the average rate of change of the number of connections, with respect to t, as a first-order, differential equation,

$$\frac{dx(t)}{dt} = -f_o(t) + f_i(t)$$

Let $\pi_0(t)$ denote the probability that there are no connections at time, t, and $\pi_{C(t)}(t)$, denote the probability that a connection is blocked at time t, given the capacity of the system $C(t)$. Then $f_i(t)$, is the arrivals that were not blocked and is given by

$$f_i(t) = \lambda(t)(1 - \pi_{C(t)}(t))$$

and connections that are completed, $f_o(t)$, is given by

$$f_o(t) = \mu(1 - \pi_0(t)).$$

Computing exact solutions for $\pi_0(t)$ and $\pi_{C(t)}(t)$ in the non-stationary case is extremely difficult. We use the point-wise, stationary approximation method[11,12] to relate the ensemble number of connections in the system, $x(t)$ to the stationary offered load, a, by the relation

$$x(t) = a[1 - E(a, C(t))]$$

(where $E(\cdot,\cdot)$ is the Erlang-B loss formula) which can be rewritten as

$$a = \frac{x}{1 - E(a, C(t))}.$$

This fixed point equation can be solved to obtain the offered load which is dependent on $x(t)$. Thus, the stationary load is denoted by $a(x(t))$. Here, the stationary values of $\pi_0(t)$ and $\pi_{C(t)}(t)$ for this gives stationary load $a(x(t))$ are then given by

$$\pi_0(t) = \frac{1}{\sum_{k=0}^{C(t)} \frac{a(x(t))^k}{k!}}$$

and

$$\pi_{C(t)}(t) = E(a(x(t)), C(t)) = \frac{\frac{a(x(t))^{C(t)}}{C(t)!}}{\sum_{k=0}^{C(t)} \frac{a(x(t))^k}{k!}}.$$

$f_o(t)$, in the stationary case, is equivalent to $\mu a(t)[1 - E(a(x(t)), C(t)]$(shown in the Appendix). Thus, the point-wise, stationary, fluid-flow approximation (PSFFA) of the system is given by

$$\frac{dx(t)}{dt} = -\mu a(t)[1 - E(a(x(t)), C(t))] + \lambda(t)[1 - E(a(x(t)), C(t))].$$

We solve this differential equation using the Runge-Kutta-Fehlberg numerical method[2], where the step size parameter is optimized based on prescribed tolerance values.

The control scheme used to implement capacity adjustment using the PSFFA model is based on network performance measures, which are calculated at each time interval t. These performance measures include, $x(t)$, the number of connections in the system, and $b(t)$, the blocking probability at time t.

As a starting place for investigation, we have implemented a capacity adjustment scheme that relies solely on the objective blocking range and no other conditional constraint. At each time value t, after the initial start-up period, the PSFFA model calculates a set of network performance measures. A conditional statement is tested based on QoS specified blocking range, $\bar{b} \pm b_\delta$. If capacity adjustment is warranted, the adjustment is made by altering the capacity, $C(t)$. The capacity is never allowed to drop lower than a value that can handle the current number of connections on the link. This ensures that current connections are never lost and is also needed for the solvability of the fixed point equation. The scheme determines if the blocking is not within the range of the objective blocking. The capacity is then increased or decreased by a pre-set value k. The scheme can be summarized as

$$C(t) = C(t) - k \text{ if } b(t) < \bar{b} - b_\delta$$

$$C(t) = C(t) + k \text{ if } b(t) > \bar{b} + b_\delta.$$

3. RESULTS

The implementation of the PSFFA model involves user supplied objective network conditions and an initial start-up period. For the purposes of this study, we establish, average offered load, $\hat{a} = 15$; the service rate $\mu = 1$; and the objective blocking range $\bar{b} \pm b_\delta = 0.02 \pm 0.01$. Runs were done for $t = 0$ to 80 with the first 10 time units as the start-up period. The initial capacity of the system, immediately following the start-up period, reflects the user supplied objective network conditions. In other words, if $\bar{b}$ is specified to be 0.02, the capacity of the system, after the start-up period, will reflect a value which accommodates a blocking value of not greater than 0.02, i.e. in our case here, $C_{start} = 23$. A sine function is used to represent the periodic, dynamic, offered load $a(t) = \hat{a} + 3\,sin(0.01(t + 20))$. Fig. 1a and Fig. 1b show performance-measure output using the PSFFA model without capacity adjustment. Fig. 1a plots the number of connections in the system, $x(t)$, versus time, t, where $\mu = 1$, and $\hat{a} = 15$. The graph reflects dynamic traffic flow. Fig. 1b plots blocking versus time for the same user specified input as Fig. 1a. The blocking shown in Fig. 1b is consistent with the dynamic traffic flow shown in Fig. 1a and ranges from a minimum value of 0.0016 occurring at $t = 28.1$ to a maximum value of 0.0486 occurring at $t = 58.8$. The weighted average blocking, b_{avg}, with respect to the load, is 0.0208. Because capacity adjustment is not implemented in this example, the capacity remains constant at $C(t) = 23$. Fig. 1b shows the potential for the implementation of capacity adjustment when the blocking falls outside of a QoS specified blocking range, $\bar{b} \pm b_\delta = 0.02 \pm 0.01$, or, when $b(t) < 0.01$ or $b(t) > 0.03$.

Now let us look at the results in which capacity adjustment is performed if blocking is not within $\bar{b} \pm b_\delta$. In our example case, the capacity allotment is increased or decreased by k, where $k = 1, 2$, and 3. The user supplied objective network conditions are the same as above and $C_{start} = 23$. Fig. 2 shows the output for blocking versus time for $k = 1$ and $k = 2$. We carefully walk through the plotted results in Fig. 2. Notice that after the start-up period, blocking reaches the minimally allowed value of 0.01 when $t = 14.1$ and as a result of the capacity adjustment scheme, the capacity allotment is decreased by 1 to $C(t) = 22$. At $t = 14.2$, the decreased capacity allotment shoots the blocking up to a value of 0.0166. As the traffic load continues to decrease, the blocking decreases until once again, blocking reaches a minimally allowed value of 0.01 when $t = 17.1$. The capacity allotment is then, decreased by 1 to $C(t) = 21$, and the blocking value increases to 0.0166. This continues during the duration of the decreasing traffic load. As traffic starts to increase, somewhere around $t = 28$, capacity adjustments are not made until blocking has exceeded a maximally allowed value of 0.03 at $t = 36.9$. The capacity allotment is increased by 1 and blocking is lowered to 0.0182. This process of reaching minimum and maximum blocking values, which in turn, leads to a capacity modification by ± 1 is repeated throughout the course of the traffic flow. For the duration of the run, the capacity is adjusted 14 times. For this period, b_{avg} is 0.0190. For $k = 2$, also shown in Fig. 2, the capacity allotment is increased or decreased by 2 when $b(t) < 0.01$ or $b(t) > 0.03$. For $k = 2$, b_{avg} is 0.0192 and capacity is adjusted 7 times. Notice that at $t = 14.2$, the blocking increased to 0.0272, versus 0.0166, due to the larger capacity decrease, compared with $k = 1$.

Now we compare the plotted output in Fig. 2 with the plotted output in Fig. 3. Fig. 3 is a plot of the blocking output using the same objective network conditions as in Fig. 2 and the same capacity adjustment scheme, except $k = 3$. At $t = 14.1$, the capacity allotment is decreased by 3 to $C(t) = 20$ when the falls below the objective range. However at $t = 14.2$, since $C(t)$ is decreased to $C(t) = 20$, the blocking increases to 0.0440 causing the adjustment scheme to increase $C(t)$ back to 23. Again blocking falls below 0.01 and $C(t)$ is adjusted to 23 at $t = 14.3$. This capacity adjustment oscillation from $C(t) = 20$ to $C(t) = 23$ back to $C(t) = 20$ causes a thrashing effect on the network blocking as shown in Fig. 3. The weighted average blocking with respect to the load, when $k = 3$, is $b_{avg} = 0.0184$. the number of capacity adjustments made is 169.

Fig. 4 plots the output of the number of connections in the system versus time, for $k = 1, 2$, and 3, using the capacity adjustment scheme and objective network conditions described above. It is interesting to note, that even for the case of $k = 3$, it is not apparent that a thrashing effect of the blocking has occurred, because the output for all three plots significantly overlap.

4. SUMMARY

We consider a dynamic virtual network environment where dynamically changing traffic is offered. In this work, we consider the case of a single, dynamic traffic class scenario, in a loss mode environment. This study shows the utility of our numerical, analytical tool for modeling a dynamic network traffic environment using a time-varying, fluid-flow, differential equation, PSFFA model. A simple capacity adjustment scheme was tested using the numerical network model. Investigation of the output of the capacity adjustment scheme shows us the importance of understanding the full implication of capacity adjustment control schemes and their sensitivities. It can be inferred from the results that a purely blocking-based, capacity adjustment control scheme can work very well in some cases but can also be very sensitive to capacity changes and can lead to network instability. The results also show us that network capacity adjustment thrashing may not be evident from examination of the flow of connections into the system as shown in Fig. 4. We have been conducting further work on the impact of several capacity adjustment schemes that condition on multiple network parameters instead of solely on a blocking range constraint; this work can be found in a separate technical report[6]. We plan to continue work in this area to explore robust and optimal capacity adjustment schemes.

5. ACKNOWLEDGMENTS

This work is supported by DARPA and Air Force Research Lab, Air Force Materiel Command, USAF, under agreement No. F30602-97-1-0257. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the Defense Advanced Research Projects Agency (DARPA), Air Force Research Laboratory, or the U.S. Government.

6. APPENDIX

Flow out of the multi-server loss system in the stationary case can also be represented as:

$$
\begin{aligned}
f_o(t) &= \mu\pi_1 + 2\mu\pi_2 + 3\mu\pi_3 + \ldots + C(t)\mu\pi_{C(t)} \\
&= \mu\left[a(t)\pi_0 + \frac{2a(t)^2}{2!}\pi_0 + \frac{3a(t)^3}{3!}\pi_0 + \ldots + \frac{C(t)a(t)^{C(t)}}{C(t)!}\pi_0\right] \\
&= \mu\left[a(t)\pi_0 + a(t)^2\pi_0 + \frac{a(t)^3}{2!}\pi_0 + \ldots + \frac{a(t)^{C(t)-1}}{C(t)-1}\pi_0\right] \\
&= \mu a(t)\pi_0\left[1 + a(t) + \frac{a(t)^2}{2!} + \ldots + \frac{a(t)^{C(t)-1}}{(C(t)-1)!}\right] \\
&= \mu a(t)\pi_0 \sum_{k=0}^{C(t)-1} \frac{a(t)^k}{k!} \\
&= \mu a(t)\pi_0\left[\sum_{k=0}^{C(t)} \frac{a(t)^k}{k!} - \frac{a(t)^{C(t)}}{C(t)!}\right] \\
&= \mu a(t)\left[\frac{\sum_{k=0}^{C(t)} \frac{a(t)^k}{k!}}{\sum_{k=0}^{C(t)} \frac{a(t)^k}{k!}} - \frac{\frac{a(t)^{C(t)}}{C(t)!}}{\sum_{k=0}^{C(t)} \frac{a(t)^{C(t)}}{k!}}\right] \\
&= \mu a(t)[1 - E(a(t), C(t)].
\end{aligned}
$$

7. REFERENCES

1. A. Arvidsson, "Management of Reconfigurable Virtual Path Networks," *Proc. 14th International Teletraffic Congress*, pp. 931-940, 1994.

2. R. Burden, J. D. Faires, and A. C. Reynolds, *Numerical Analysis, Second Edition*. Prindle, Weber & Schmidt, Boston, MA, 1981.

3. J. Filipiak, *Modelling and Control of Dynamic Flows in Communication Networks*. Springer-Verlag, Berlin, 1988.

4. J. Filipiak and P. Chemouil, "Modeling and Prediction of Traffic Fluctuations in Telephone Networks," *IEEE Transactions on Communications* **COMM-35**, pp. 931–941, 1987.

5. G. Gopal, C. Kim, and A. Weinrib, "Algorithms for Reconfigurable Networks," *Proceedings 13th International Teletraffic Congress (ITC-13)*, pp. 341-347, 1991.

6. B. Groskinsky, D. Medhi, and D. Tipper, "An Investigation of Capacity Control Schemes In A Dynamic Traffic Environment," Technical Report, Computer Science Telecommunications, University of Missouri–Kansas City, July 1998.

7. D. Medhi, "Multi-hour Network Design for Dynamically Reconfigurable Wide-Area ATM Networks," *IEEE/ACM Trans. On Networking* **3**, pp. 808-819, 1995.

8. M. Logothetis, S. Shioda, "Medium-Term Centralized Virtual-Path Bandwidth Control Based on Traffic Measurements," *IEEE Transactions on Communications* **43**, pp. 2630-2640, 1995.

9. Y. Qian, D. Tipper, and D. Medhi, "An Analysis of Access Control Schemes for Multirate Loss Networks Under Nonstationary Conditions," *Proc. of IEEE INFOCOM'96*, pp. 730-737, 1996.

10. S. Shioda, H. Toyoizumi, H. Yokoi, T. Tsuchiya, and H.. Saito, "Self-sizing Network: A New Concept Based on Autonomous Network VP Bandwidth Adjustment," *Proceedings 15th International Teletraffic Congress(ITC-15)*, pp. 997-1006, 1997.

11. D. Tipper, and M. K. Sundareshan, "Numerical Methods for Modeling Computer Networks Under Nonstationary Conditions," *IEEE Journal on Selected Areas in Communications* **8**, pp. 1682-1695, 1990.

12. W. Wang, D. Tipper, S. Banerjee, "A Simple Approximation for Modeling Nonstationary Queues," *Proceedings of IEEE INFOCOM'96*, pp. 2d.3.1-2d.3.8, 1996.

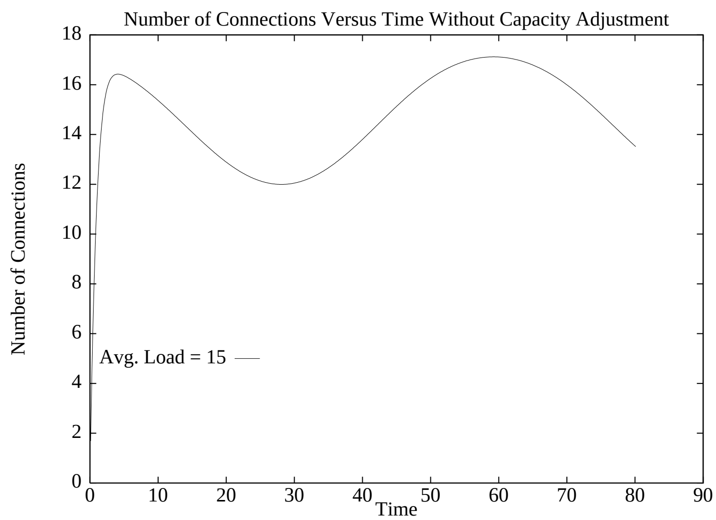

Fig. 1a: Number of connections vs. time without capacity adjustment

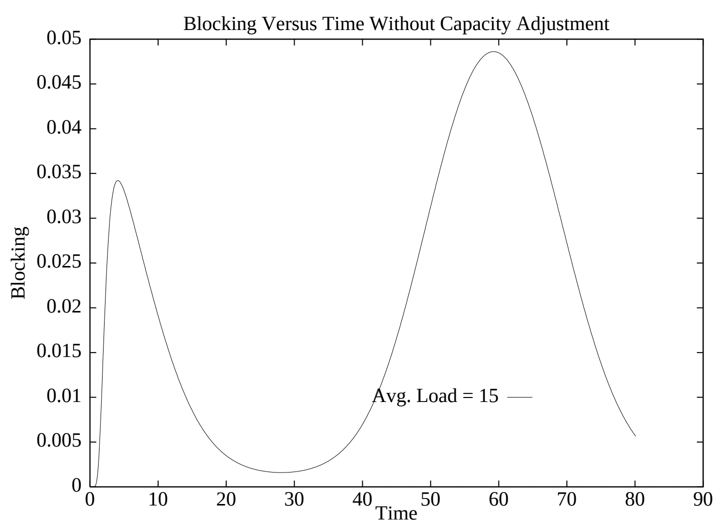

Fig. 1b: Connection blocking vs. time without capacity adjustment

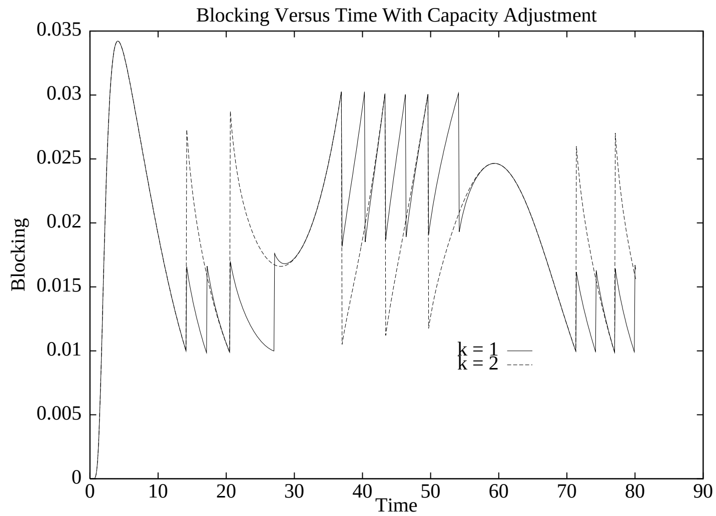

Fig. 2: Connection blocking vs. time with capacity adjustment when $k = 1$ and $k = 2$

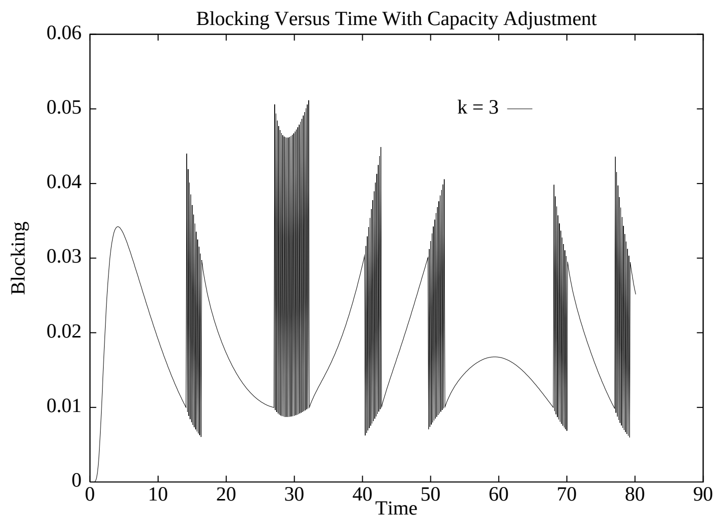

Fig. 3: Connection blocking vs. time with capacity adjustment when $k = 3$

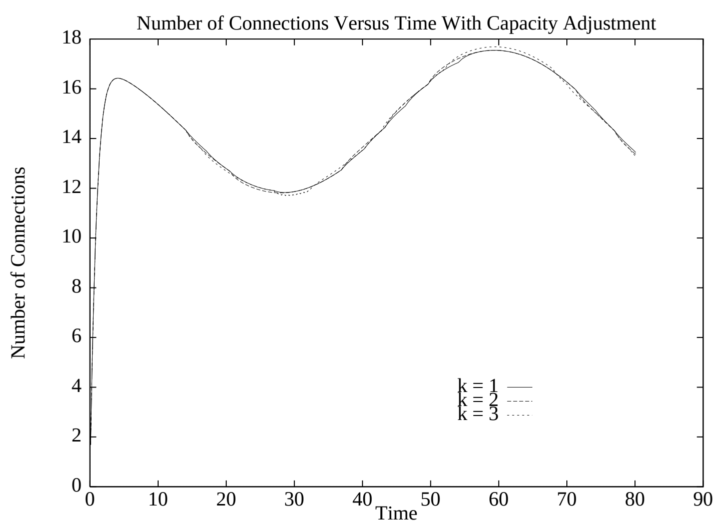

Fig. 4: Number of connections vs. time with capacity adjustment when $k = 1, 2, 3$