http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 7 (2): 306-312, 2008
ISSN 1812-5638

Simulation of Scalability and Congestion Control of Broadband Intelligent Networks

[1]S. Sankara Gomathi and [2]M.A. Bhagyaveni
[1]Department of Computer Science and Engineering, Sri Venkateswara College of Engineering,
Sriprumbudhur, Tamilnadu, India
[2]Department of Electronics and Communication Engineering, College of Engineering,
Anna University, Chennai, Tamilnadu, India

Abstract: Broadband Intelligent Networks (B-IN) are ought to sustain a large number of mobile and fixed users and a varying amount of data while maintaining high availability of resources and acceptable performance. In this study, we address the impact of increasing the number of users supported by a network node and provide an experimental study to test B-IN scalability. We also address the congestion control issue becoming necessary to protect a B-IN during overload situations. Existing algorithms do not operate optimally with respect to important performance criteria such as responsiveness, efficiency and fairness. Here, we identify some of their shortcomings and propose a new algorithm called Enhanced Call Gapping (ECG) able to provide consistently better performance than current algorithms in terms of above-mentioned criteria. Experiments have been conducted using NS-2 simulator.

Key words: Broadband intelligent mobile networks, scalability, congestion control, service control point, service switching point

INTRODUCTION

The ultimate goal of mobile communication services is to provide information to any person, in any form, at any time and at any place in the world. This can be realized through the provision of terminal, personal and service mobilities. Third generation mobile communications systems such as the International Mobile Telecommunications-2000 (IMT-2000) have been recommended to fully support these mobility requirements. Intelligent Network (IN) concepts are used in these third generation mobile communication systems to provide call/connection functions and facilitate a rapid service introduction. The IN architecture allows the rapid and cost effective deployment of new services by separating the service control and service switching currently located in switches. On the other hand, Broadband Integrated Services Digital Network (B-ISDN) is intended as a future wired telecommunication network. It is service independent and based on the Asynchronous Transfer Mode switching concept.

Because of the expansive growth of the available capabilities in telecommunications, it is expected that many services that are currently provided by other media (e.g., video films), will be taken over by telecommunication networks. In this study, we are interested in studying Broadband Intelligent Networks (B-IN) model designed to provide new services to fixed and mobile users and seen as an integration of the B-ISDN and IN. Particularly, we aim at studying the scalability and congestion control of B-IN model.

Kawahara and Asaka (1996) has argued that traffic on the telephone and signalling network will become much more unpredictable and volatile, with the consequence that the overload control becomes a necessity in the network of present and future. As per (Brennan *et al*., 2000) scalability is one of the most important factors in the design of a distributed system. It can be defined as the ability to increase the size of the network, while maintaining Quality of Service and network performance criteria. On the other hand, the congestion occurs when the core network processes an incoming load larger than its capacity. Congestion problems and associated performance degradation can be solved by using appropriate congestion control mechanisms suggested by (Karagiannis and Nicola, 2000). Congestion control algorithms are used to protect the Service Control Point (SCP-contains data and service logic required by many telecommunication network services) during overload situations and to ensure continuous and fair processing of requests from different sources. Author (Bos and Leroy, 2001) suggest a balanced Island structure to

Corresponding Author: S. Sankara Gomathi, Department of Computer Science and Engineering,
Sri Venkateswara College of Engineering, Sriprumbudhur, Tamilnadu, India

evaluate scalability in B-IN. The algorithm of (Kryvinska and Harmen, 2002) does not operate optimally with respect to important performance criteria related to their responsiveness, efficiency and fairness. Authors (Pham and Betts, 2004), describe ACG algorithm where the duration interval is not used and the value of the gap interval is determined adaptively by the SCP.

In this study, we verify the scalability of B-IN via a set of experiments conducted using a NS-2 Simulator (Fall and Varadhan, 2000). We also identify some shortcomings of current congestion control algorithms and propose a new algorithm called Enhanced Call Gapping (ECG) that can effectively remove these shortcomings. ECG gives the possibility to the communication network operator to provide adaptive robust and fair admission policies to different user request by simultaneously achieving high session (or call) throughput while protecting the network from overloads. It provides better performance than current congestion control algorithms.

TERMINOLOGY

In essence, several common measures are used to compute the performance of a network particularly,

- Utilization is the fraction of time that a processor, a physical entity, or a network, is busy (non-idle).
- Throughput: is the number of processed service requests per second.
- Blocking rate represents the number of rejected (or blocked) service requests (or procedures) per second.
- Reaction time is the duration of time needed by the congestion control algorithm to bring a congested SCP back to its referential operating load.
- Robustness is the ability of the congestion control algorithm to provide nearly optimal throughput regardless of variations of the congestion detection and notification mechanisms of the B-IN architecture. The congestion control algorithm should be able to avoid oscillations (over- utilization and under-utilization) of the SCP independently of the amount of traffic overload.
- Fairness is the ability of the congestion control algorithm to adjust the traffic sent by a source depending on its required SCP capacity relative to the reference capacity set by the network provider. For example, conforming sources need not to be penalized during overload periods.
- Island is a network topology. A Balanced Island consists of one SCP providing service control facilities to four Point SSPs (Kihl and Nyberg, 2003).

EXISTING WORK

Here, we detail several congestion control algorithms closure to the work apart from the one given in the literature. SCP overload controls (Kolyvas *et al.*, 1998) use a rate based congestion control mechanism, which is a closed loop control with responsive global feedback. Due to its characteristics, this type of congestion control algorithm is typically considered to be the most efficient type algorithm for Intelligent Networks.

Commonly used in IN, is the ACG mechanism allows reducing the call attempt rate by allowing only one call attempt per gap interval (g). This reduction takes place when the mechanism has been activated due to congestion and lasts for gap duration (d). Two ACG mechanisms are provided: table driven ACG and adaptive ACG (Pham and Betts, 2004).

Table driven ACG: Here, each SSP (Kolyvas *et al.*, 1998) maintains a view of the congestion level of the SCP, which is sent by the SCP to each individual SSP. A typical way to do this is by sending an integer value, which directly reflects the congestion level. Within the SSP, a table is stored which associates a gap duration and gap interval with each congestion level. The SCPs must monitor regularly their overall and subsystem congestion status. Different congestion level counters are maintained for the overall SCP and for each individual subsystem of the SCP. The congestion level is determined by an overload rule, which is used to determine the overall SCP congestion levels is shown in the following pseudo-code:

```
if (overall_utilization > upper_threshold)
{
        overall_congestion_level ++
}
elseif (overall_utilization < lower_threshold)
{
        overall_congestion_level —
}
```

Despite its several advantages, the table driven ACG algorithm is not adaptive and therefore the values of the congestion control parameters have to be predefined. Moreover, a non-conforming source may lead to the throttle of all sources of the same type.

Adaptive ACG: To overcome the shortcoming of the above, an adaptive ACG algorithm has been described by Pham and Betts (2004), where the duration interval is not used and the value of the gap interval is determined adaptively by the SCP. The calculation of the gap interval is based on the values of one or more measures observed in each measurement period. The following measures are proposed

- The average response time or delay
- The dropped message count
- The incoming message rate

One or a combination of these measures can be used to determine the overload or congestion level. Here also, several shortcomings are identified. First, adaptive ACG does not provide the network operators with the possibility to specify a maximum utilization threshold for each source. Therefore, a network operator would not be able to control the SCP capacity used by each source. Furthermore, despite its adaptability, this algorithm is not able to implement flexible fairness criteria. When a source exceeds its threshold and the SCP becomes overloaded, then all sources, even those that do not exceed their limit, will be punished.

STUDY OF SCALABILITY

A balanced Island structure is considered in the scalability experiments. The (B-SSPs) Service Switching Point is the local exchange in the telephone network, and (B-SCPs) is an entity in the intelligent network, that implements a service control function. These are clustered into B-IN islands, where each consists of 4 SSPs and one SCP. The impact of increasing the number of users supported by a network node, links in the network, on the quality of service is also considered. First, the bottleneck is initiated by one or more physical entities approaching their maximum processing capacity. In order to maximize the network throughput, the bottleneck is removed by balancing the processing speed of the B-IN network physical entities, such that their utilizations are approximately equal. The balanced network is then used to perform two sets of scalability experiments.

First set of experiments: The experiments were accomplished in several steps. The relationship between the processing speed and the throughput, the processing speed and the mean delay, is investigated in the first set of experiments with a fixed utilization (max = 0. 8). A linear relationship implies high degree of scalability. In Fig. 1, the throughput is plotted for different values of the processing speed. The throughput increases linearly with the processing speed because, when the processing speed is increased, the service time for a request is decreased, which in turn increases the number of requests serviced. In Fig. 2, the mean delays are plotted for different values of the processing speed. The mean delay decreases with the processing speed, when the processing speed is increased, the requests are processed faster due to the reduction in service time. This reduces the amount of time that a request has to wait for service and hence there is a fall in the delay with increase in the processing speed.

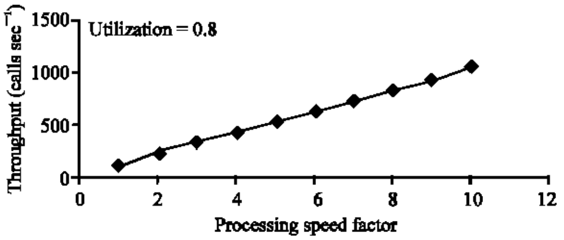

Fig. 1: Processing speed vs throughput

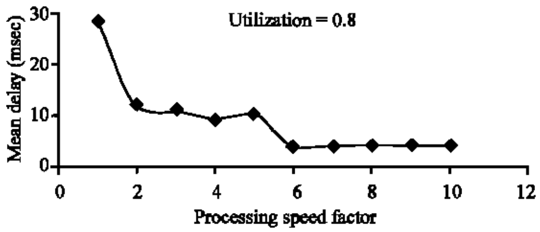

Fig. 2: Processing speed vs mean delay

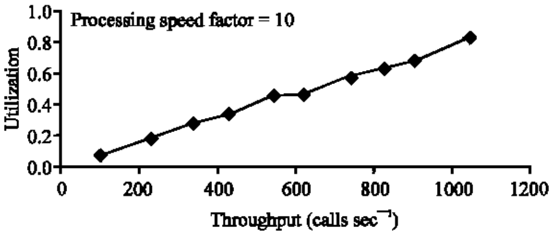

Fig. 3: Throughput vs utilization

From this set of experiments, we conclude that the network is scalable, with the only limit that the processing speed which is due to physical or practical limitations, cannot increase indefinitely. Such a practical limitation could be the occurrence of new bottlenecks in the network.

Second set of experiments: The second set of experiments is done with a fixed processing speed factor (α = 10). Figure 3 shows that the utilization increases linearly with the throughput. When the number of service requests increases and the SCP has enough processing capacity, the number of processed service requests increases. Thus the time, the SCP spends on processing the requests increases. Hence, we obtain a linear relation between throughput and utilization.

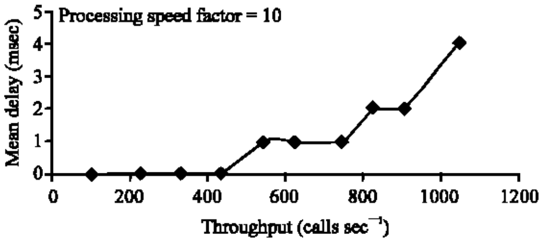

Fig. 4: Throughput vs mean delay

Figure 4 shows the relationship between the throughput and the mean delay. The delay increases sharply with the throughput. When the service request to be processed increases with fixed processing capacity, then, the waiting time of the service requests in the queue also increases. This implies that the B-IN network is scalable.

Further the research deals about the novel Enhanced Call Gapping algorithm which will able to protect B-SCP during overload situations and insure continuous and fair processing of service request.

Novel congestion control algorithm and experimental study: Due to the fact that the SCP is used to provide service control in a centralized way, it is expected that this physical entity can limit the maximum throughput of the B-IN network. By using efficient congestion control algorithms, the SCP physical entity can be protected from being overloaded. Here, we present an Enhanced Call Gaping (ECG) algorithm able to provide better performance than current algorithm.

Enhanced call gaping (ECG) algorithm: The ECG algorithm is an extension of ACG. It shares the advantages of ACG and overcome its disadvantages, by allowing the communication network operator to provide efficient, adaptive and fair admission policies to different call request types, while maintaining high session throughput and protecting the SCPs from overloads. ECG uses a fixed gap interval as in ACG and allows a source to adjust the gap interval value in order to achieve the right transaction rate. The SCP computes gaps using the following Eq. 1.

$$g = T_{out}-T_{in} \qquad (1)$$

where, g is the gap interval, T_{out} is the estimated mean time between call requests from the SSP to the SCP and T_{in} is the estimated mean time between call requests at a source (towards the SSP).

In this a measurement period is set to measures the average response time or delay, dropped (or blocked) message count, incoming (or offered) message rate, the overall utilization, Subsystem utilization. One or a combination of these measures can be used to determine the overload or congestion level of the SCP. By adding the subsystem utilization rule the fairness has been brought into the network.

Congestion level is determined based on the following as,

```
CL (1): = 0
For all t>1: (Trigger AND CL (t+ 1): = MIN
(Cmax, CL(t)+1)) OR (NOT trigger AND
CL (t+1): = MAX (O, CL (t)-1)
```

That is, the congestion level, CL(t), associated with measurement period t, can vary from zero to a predefined maximum Cmax. Initially, the congestion level is set to zero. If a trigger occurs during a measurement period, then the congestion level is incremented by 1. If no trigger occurs during a measurement period then the congestion level is decremented by 1. The congestion level is updated by the rule given above. The algorithm for determining subsystem utilization is given as follows.

```
if ((subsystem utilisation >subsystem-threshold)
AND(overall_utilisation > upper_thershold))
{
subsystem congestion-level =
subsystem congestion level + 1
}
else if (overall_utilisation < lower-threshold)
{
subsystem congestion-level =
subsystem congestion level - 1
}
```

This gives the additional characteristics such as the congestion control functionality in SCP is capable of measuring the SCP utilization due to each individual source connected to a SSP, the sum of which is the overall SCP utilization. The network operator may specify a SCP utilization bound (threshold) for each source or source type. The reduction factor r is not constant but, depends on the source parameters and the traffic. In this way, we obtain higher reductions at sources that demand higher processing capacity from the SCP. Furthermore, ECG provides reductions only to sources that require more capacity than their pre-specified (by the operator) utilization bound. If the SCP is under-utilized (congestion situation is not triggered), then some sources may use more SCP capacity than their respective utilization bounds, thus achieving higher overall SCP utilization.

Load balancing can be easily provided by re-directing call requests that are rejected by a congested SCP to another non-congested one.

Experimental study: Each physical entity in the B-IN network can be modeled as a single server queue with infinite message buffer. In our simulation, we have modeled the SCP as a single server queuing system with infinite buffer and performed the overload experiments to compare the performance of ECG to current one. The experiments are performed on a single B-IN island. Two types of measures are used for the comparison of different congestion control algorithms:

- Qualitative measures such as reaction time, robustness and fairness.
- Quantitative measures such as Utilization, Throughput and blocking rate.

We assume that an Intelligent Network call begins with an initial query. If the initial query is accepted, the call is served. This assumption does not have significant effect on the validity of the results because only the initial request is relevant to the congestion control algorithms.

Experimental settings: For a stable system, the average service rate should always be higher than the average arrival rate, otherwise the queues will rapidly race towards infinity. In telecommunication services, the arrival process of requests is often approximated by a Poisson process. In our experiments, the arrivals follow a Poisson process. The inter-arrival times are independent of each other and each has an exponential distribution with mean inter-arrival time. The service times are also independent of each other and have an exponential distribution with mean service time. All simulation runs last 450 sec. From time 0 to 100, the offered call rate is 100 calls sec^{-1}. Then, it jumps to 400 calls sec^{-1} at 100 and continues until 300. The system is overloaded in order to observe the performance of the congestion control algorithms. At 300, the call rate returns to 100 calls sec^{-1}. The SCP capacity is 250 calls sec^{-1} and therefore the SCP service time for a call is 4 msec. Table 1 recapitulates these settings.

Threshold settings: The model aggregates all operating system and high priority non-call related processes into a single overhead process. The overhead process runs 20% of time during the simulation to represent that portion of time during which the SCP cannot process calls. Therefore, the utilization threshold is set to 0.8 in the overload experiments. The first set of experiments has been performed with a single class of service. Only one type of call arrives at the SSP and is sent to the SCP. This is mainly performed to observe the robustness and the reaction time of the congestion control algorithms. The second set of experiments with multiple classes of service, denoted as individual overload experiments, has been performed to observe the fairness of the congestion control algorithms. In this set of experiments, the overload situation is caused by sources generating one class of service.

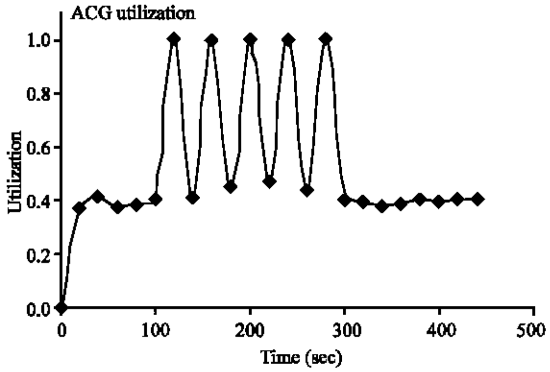

Fig. 5: ACG utilization

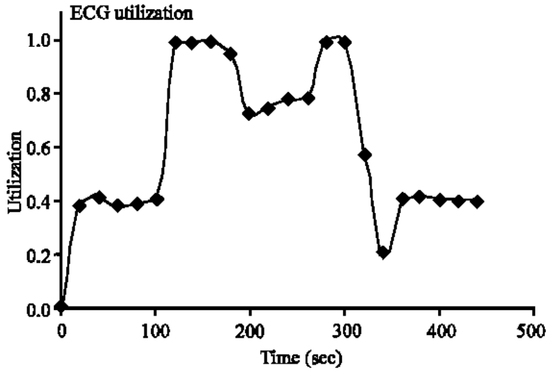

Fig. 6: ECG utilization

Table 1: Simulation settings

	Normal	Overload
Arrival rate	100 calls sec^{-1}	400 calls sec^{-1}
Service rate	250 calls sec^{-1}	250 calls sec^{-1}

Analysis and results: In Fig. 5 and 6, the server usage under ACG oscillates between 40 and 100% during the time interval from 100 to 300 and the queue at the SCP alternately grows and empties. ACG alternately under-controls and over-controls the calls leading to idleness of SCP even during a period of overload. It over-controls traffic because its gap settings are too large. When congestion is not observed for one measurement period, the abrupt removal of control recreates an overload and gives the control an undesirable on-off character, where

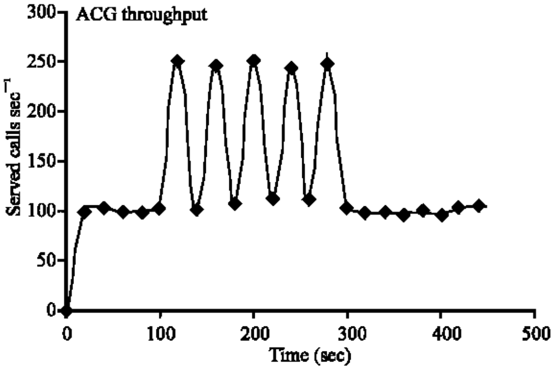

Fig. 7: ACG throughput

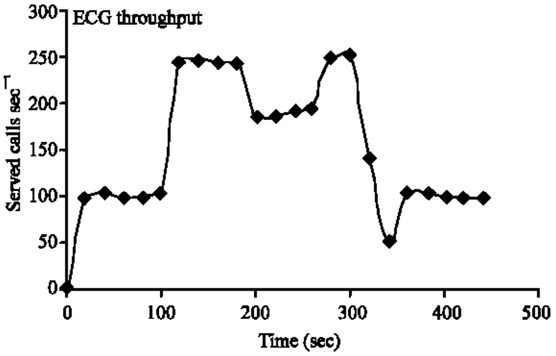

Fig. 8: ECG throughput

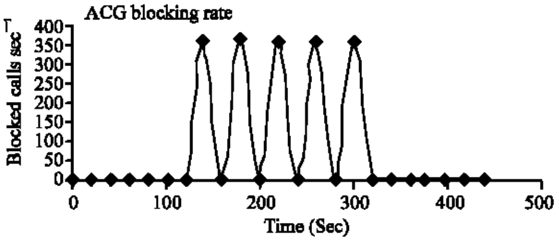

Fig. 9: ACG blocking rate

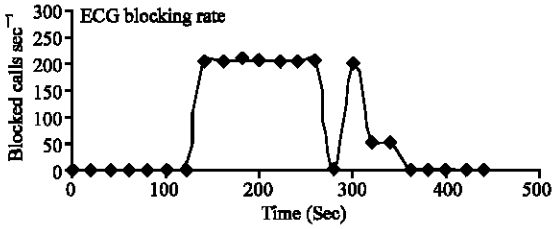

Fig. 10: ECG blocking rate

in a period of throttling alternates with a period of no control at all. In contrast, ECG drives the SCP to 100% utilization and maintains 75% utilization from time 100 to 300 that the SCP is working at full capacity during overload.

The call throughput is significantly better with EGC than with ACG. The difference is clearly visible in Fig. 7 and 8. The total throughput rate under ACG oscillates with a mean near 225 calls sec^{-1} from time 100 to 300. By contrast, the total throughput rate for ECG is at most 175 during the interval 100 to 300. Figure 9 and 10 shows that the ECG control keeps the dropped calls down when compared to ACG.

CONCLUSION

The implementation and performance evaluation in relation with scalability and congestion control of the Broadband Intelligent Network (B-IN) have been addressed. The conducted scalability experiments show that B-IN is scalable. A new congestion control algorithm called Enhanced Call Gaping (ECG) is provided and compared to current ones. The congestion simulation experiments shows that the enhanced algorithm performs better than the existing algorithms in terms of reaction time, robustness, utilization, throughput and blocking rate.

Future experiments will be conducted with several B-INs and with multiple islands to get a real feel of the network. Furthermore, the enhanced congestion control algorithm will be extended with load balancing functions to redirect the user requests that are rejected by a congested SCP to another non-congested SCP. In addition, the new experiments will be carried out with various other distribution models and two or three combined measures to determine the node or network overload.

REFERENCES

Bos, L. and S. Leroy, 2001. Toward an All-IP-Based UMTS system architecture. IEEE. Network, 15: 36-45.

Brennan, R., B. Jennings, C. McArdle and T. Curran, 2000. Evolutionary trends in intelligent networks. IEEE. Commun. Mag., 38 (6): 86-93.

Fall, K. and K. Varadhan, 2000. Network simulator notes and documents, the VINT project. UC Berkeley, LBL, USC/ISI and Xerox ARC, http://www. isi. edu/nsnam/ns/ns-dcumentation. html.

Karagiannis, G. and V. Nicola, 2000. Evaluation of scalability in a broadband mobile communication system. In: Proceedings of IEEE. Intelligent Network Workshop.

Kawahara, R. and T. Asaka, 1996. Overload control for intelligent networks based on an estimation of maximum number of calls in a node. IEEE Intelligent Network Workshop 96, Vol. 1, Melbourne, Vic., Australial.

Kihl, M. and C. Nyberg, 2003. Investigation of overload control algorithms for SCPs in the intelligent network. IEEE Process Commun., 144 (6): 419-424.

Kolyvas, G.T., G.K. Mamais and I.S. Venieris, 1998. Overload and congestion control for IN based broadband call processing systems. In: Proceedings of IEEE International Conference on Communications (ICC”98), 1: 259-264.

Kryvinska, N. and R.V. Harmen, 2002. Queuing system models for performance analysis of Intelligent Networks. In: Proceedings of the London Communications Symposium, pp: 169-172.

Pham, X.H. and R. Betts, 2004. Congestion control for intelligent networks. In: Proceedings of International Conference on Networking, pp: 375-386.