

A Strategy For The Resolution of Intelligent Network (IN) and Signalling System No. 7 (SS7) Congestion Control Conflicts

Brendan Jennings, Fiona Lodge, Thomas Curran.
Teltec Ireland, Dublin City University, Ireland.

Abstract

Signalling System No. 7 (SS7) networks are used to transport call control messages for both traditional telephony and advanced Intelligent Network (IN) based services. A large number of IN implementations have been successfully deployed in recent years, however a key factor in ensuring their ongoing commercial success will be that they continue to meet specified performance targets. Because network congestion may have a serious adverse effect on IN service performance, much research effort has been focused on IN congestion control. However, in general, this research has not taken into account operation of the standardised congestion controls operating at the SS7 plane of the network. This paper outlines an investigation that demonstrates that the interaction between SS7 and IN congestion controls is not always harmonious and can lead to serious performance degradation. A congestion control strategy, which resolves the problems identified, is presented and evaluated.

1. Introduction

The introduction of advanced service architectures such as the Intelligent Network (IN) into the telecommunications infrastructure has resulted in the offering of many different services to the end-user. As the number, complexity and usage of these services increases, a critical factor in ensuring their commercial success will be that their performance meets specified targets, even when the network is under strain. A number of network failures resulting from congestion has led to in-depth investigations of congestion control issues for IN and SS7. In particular, media-stimulated mass calling events and focused overloads have called into question the effectiveness of the independent congestion controls that reside in the SS7 and IN planes of the network [Atai96]. This paper investigates the joint operation of these controls and finds that in a focused overload scenario they can conflict with each other, resulting in serious degradation of service performance. Furthermore, a strategy for combating the identified shortcomings in existing mechanisms is presented and discussed.

2. SS7 Congestion Control Strategies

Congestion control in Signalling System No.7 networks is realised by a number of overload detection and traffic throttling mechanisms residing at different levels of the SS7 protocol. A central role is played by the Message Transfer Part Level 3 (MTP3) Transfer Controlled (TFC) mechanism, a reactive control that aims to maximise throughput during periods of overload by instructing traffic sources to throttle the volume of signalling messages they offer the network. The TFC control detects the overload level of a signalling link by ascertaining

when the size of the link's MTP Level 2 transmission queues surpass a pre-defined congestion onset threshold. TFC congestion messages are then sent back to source Signalling Points, whose User Parts throttle their load towards the affected destinations in a stepwise manner, using (non-standardised) sets of message priorities. A comprehensive discussion of the operation of the TFC mechanism and the other SS7 congestion controls can be found in [Manfield93].

The introduction of advanced service architectures, supporting intelligent services and mobility (which utilise SS7 for message transfer), has resulted in significant changes in the characteristics and volume of signalling traffic, changes which have been shown ([Zepf94]) to have significant negative impact on the performance of the SS7 controls. There is broad consensus that the best solution to this problem is the deployment of application level controls which can be designed to be sufficiently adaptive to operate efficiently in a signalling network supporting a large number of heterogeneous service types.

3. IN Congestion Control Strategies

To date most documented research in the area of IN congestion control has been focused on developing ways to maximise IN Service Control Point (SCP) efficiency.

Two different types of strategy have generally been investigated and compared. In the first, independent throttling of calls takes place at both the SCP and Service Switching Points (SSPs). When overload occurs at the SCP, it throttles all new arrivals until the overload condition has been alleviated. No notification of overload is sent to the SSPs. At SSPs, an algorithm is in place which monitors the response time of the SCP to requests from that SSP. If the response delay to a request becomes excessive, overload is deemed to have occurred and incoming calls are throttled. The most widely used algorithm of this type is the Window mechanism (described in [Pham92]).

In the second type of strategy, an overload detection algorithm is located at the SCP and works in conjunction with a throttle located in the SSP. When SCP overload is detected, a control message indicating the severity of the overload is sent, via SS7, to the SSP, where it is interpreted and an appropriately severe throttle is put in place to restrict incoming IN traffic. Types of throttling mechanism include Call Gapping (CG), Percent Thinning, and Leaky Bucket.

The operation of the two above control methods has been compared (see [Pham92], and [Nyberg94]) with differing results, although it has generally been found that the interactive scheme is superior. The primary reasons for this are:

- For the interactive scheme, no SCP processor time is spent rejecting calls, thus maximising SCP efficiency; while, in the

SSP, the throttle is inactive when no overload condition exists, thus minimising SSP processing overheads,

- Window is, by its nature, very slow to detect and respond to the onset of congestion.

However in all of these studies the detailed operation of the underlying SS7 network (and in particular its congestion controls) has been omitted when investigating the performance of a given IN congestion control strategy.

4. Description of Simulation Model

In order to analyse as completely as possible the interaction between SS7 and IN congestion controls the simulation model represented in Figure 1 below was developed. The model incorporates all the entities needed to model the provision of IN and non-IN services as well as both the SS7 and IN congestion control mechanisms.

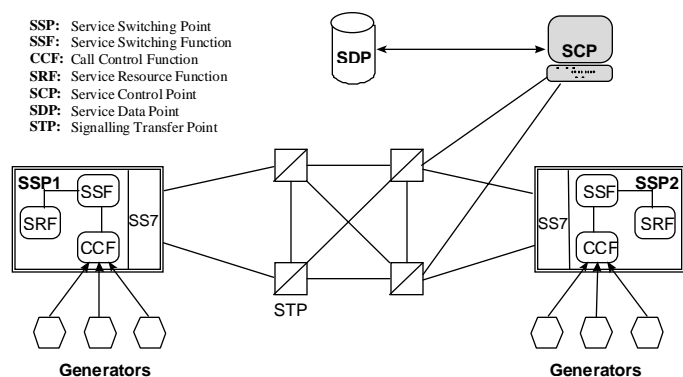


Figure 1: The IN/SS7 Simulation Model

The following attributes of the simulation model should give a good indication of the modelling assumptions that have been made in this study:

- Three service types are modelled, these are standard telephony (modelled in accordance with the SS7 ISDN User Part, ISUP), IN Freephone and IN Televoting. Freephone requests require the SCP to translate a number by requesting information from the SDP, while televoting requires two interactions with the user (handled by the SRF) and one SDP database update.
- In normal operation the percentages of signalling traffic belonging to the three service types are: Telephony: 75%, Freephone: 15% and Televoting 10%.
- The IN Service Resource Function (SRF) is shown as being integrated into the SSP – modelling the SSP as having control over SRF execution simplifies the model but without affecting the sequence of events involved in service processing.
- The SSP has been dimensioned so as to ensure that it will never become overloaded.

- Telephony calls are fully modelled on a call-by-call basis with all timers for the generation of re-attempts of circuit release messages in place.
- The SS7 congestion controls follow the International Option, using transmission queue congestion threshold sizes as recommended by [Manfield94].
- ISUP message traffic is throttled based on two priority levels: Initial Address Messages (IAMs) have lower priority and are throttled first, while all other call progress messages are of the higher priority.
- Messages are assumed to incur a set delay traversing the STPs, the processors of which are assumed to not be in danger of overloading.
- IN messages are assumed to be 4 times longer than ISUP messages.
- The IN-plane controls are based on Load Measure Control (described in [Korner91]) at the SCP with a Call Gapping throttle residing in the SSPs.
- The IN services are assumed to utilise the Global Title Translation (GTT) addressing capability of the SCCP in the SSP. For this reason the IN User Parts will not be able to react to incoming Congestion Indications (which specify congestion destinations in terms of SS7 Point Code addresses) and therefore will not be capable of any traffic throttling. This factor will have significant negative impact on the operation of SS7 controls ([McMillan96]). It is also assumed that no traffic throttling takes place in the SCCP.

5. Analysis of Interactions between SS7 and IN Congestion Controls

The simulation model described in Section 4 above has been used to examine a number of different focused overload scenarios caused by rapidly increasing the volume of requests for the televoting service – corresponding, for example, to a media-stimulated mass call-in event. In the simulation runs the number of televoting requests was increased in a stepwise fashion over a period of duration 180s, starting at time $t = 900s$. After this time the input traffic levels are approximately 10 times their original values. The input traffic was kept at this level until $t = 1500s$, at which time it was decreased (over 300s) back to its original value. The two cases discussed below provide good examples of the major findings of interest.

Case 1:

In this scenario the increase in televoting traffic is sufficient to trigger an overload of the SS7 links which connect the SCP with its local STP pair (other links, such as those connecting STPs to each other do not become overloaded), but which is not of sufficient magnitude to cause the SCP itself to become overloaded. Note that in this case the SS7 congestion control mechanisms will be invoked whereas the IN ones will not. Figures 2 and 3 show STP to SCP link queue size and SCP load respectively during the overload period.

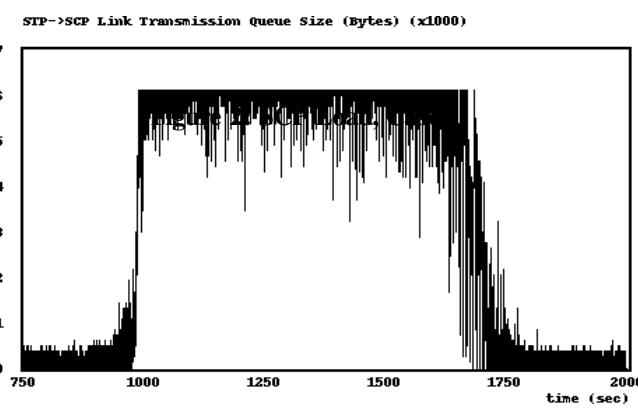
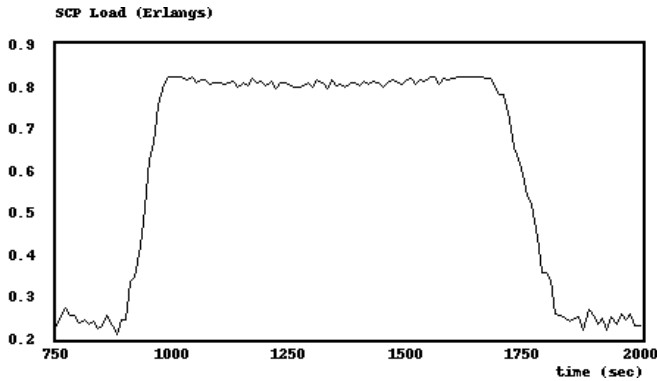


Figure 3: STP→SCP Link Transmission Queue, Case 1

As is evident from Figure 2 the links from the SCP to the STPs remain overloaded for the duration of the overload, even though the SS7 MTP3 Transfer Controlled (TFC) control mechanism have been invoked and will have sent messages back to the SSPs, indicating that they should throttle messages destined for the SCP. However, as we assume that the IN User Parts utilise the SCCP GTT addressing service at the SSPs, they will not be able to respond to these congestion indications, thus no new televoting requests will be throttled and the link will remain overloaded.

Meanwhile, as Figure 3 shows, the SCP load will not surpass its overload onset threshold (set here to be 0.9 Erlangs.) However, because of the link queue build-up, messages will experience significant delays reaching the SCP (in addition, those arriving when the buffer is full will be discarded). Results not presented here show that response delays experienced by users to their initial service are typically in the range 3-6 seconds, which will result in a large number of users abandoning their call attempts. The magnitude of these delays is largely dependant on the size of the link transmission buffers – a value of 6 Kbytes was used for the above simulation, but higher values would result in even more unacceptable delays.

Case 2:

The second set of results we present correspond to the scenario in which the input traffic causes overload of both the SS7 links and the SCP itself. This happens when the capacity of the SCP is less than or equal to that of the two signalling links connecting

it to the local STP pair. Link overload occurs as before, however messages arriving at the SCP cause its load to pass the threshold value causing congestion controls to be invoked, resulting in a Call Gapping (CG) message being sent to both SSPs. Upon receipt of the CG messages (they will arrive promptly as the links are not overloaded in the reverse direction) the SSP starts to throttle new requests for IN services. This has the effect of the STP→SCP links not being offered any new traffic, so they have an opportunity to service the message backlog contained in their queues. Therefore, during this time, the SCP receives a constant stream of the delayed messages from the SS7, keeping its load constant at 1 Erlang during the time required to process them. Eventually its load will fall back below the overload threshold, which triggers the sending of a CG message to the SSPs instructing them to discontinue request throttling. Figures 4 and 5 show, that because of this, the traffic level the SSP offers the network is increased, resulting in the entire cycle of link and SCP overload starting again. Queue lengths on the SS7 links, and on the SCP processor input queues have the effect of greatly increasing the delays experienced by messages traversing the network, to the extent that an extremely large percentage of service requests will be abandoned.

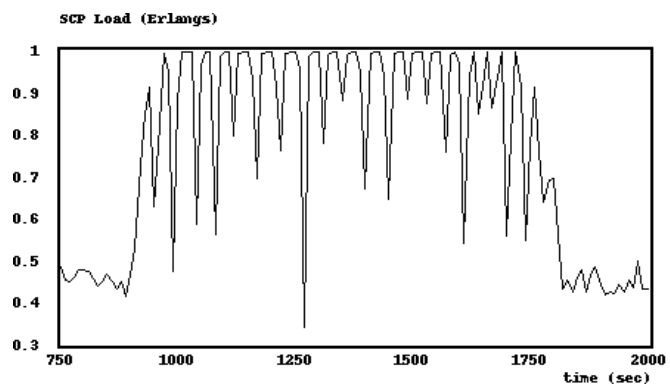


Figure 4: SCP Load, Case 2

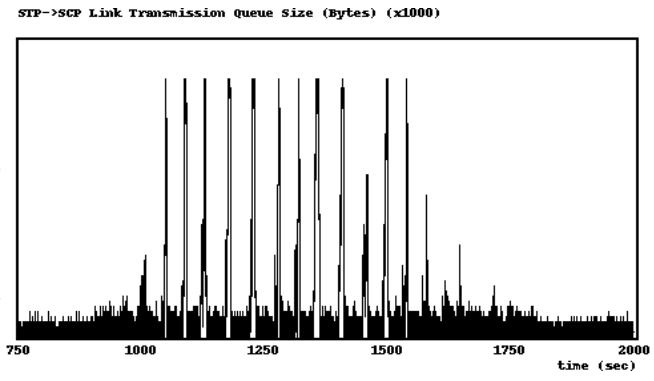


Figure 5: STP→SCP Link Transmission Queue, Case 2

The two cases described above indicate that signalling link congestion can cause a serious degradation in IN performance. Even in Case 2, where the SCP does become overloaded, it can be seen that the IN controls have been invoked too late to prevent SS7 link congestion. Furthermore, the fact that their operation leads to oscillatory behaviour and that they do not

solve the problem of long service delays renders them ineffectual. This points towards the need for an IN congestion control strategy which can react quickly to overload of both IN and SS7 resources and which is not solely dependant on the prompt exchange of congestion indication messages.

6. A Unified Congestion Control Approach

The solution we propose is to use a Window-like SS7 congestion control strategy in the SSP combined with an interactive IN control strategy where overload is detected by the SCP itself and throttles are put in place in the SSPs.

The new SSP SS7 control mechanism is a form of Call Count Control (CCC) and operates as follows:

A counter C_x is defined for each service type x . This corresponds to the number of initial messages of this service type which have been sent but for which a response remains outstanding. Each time a new initial message of type x is sent by the SSP C_x is incremented and a timer is set to a time appropriate to the service type – for example the timer duration for telephony calls will be based on the guidelines defined in [E.721], while for IN services it will be similar to those defined in [MacD94]. Each time an appropriate response is received from the SCP C_x is decremented and the corresponding timer is reset. At the end of a time interval T , all counters are examined and an overload level is assigned for each service type based on its counter value. An appropriate throttle (we used CG in our study) is then put in place for each service type.

Preliminary results showed that whilst this algorithm is very efficient at preventing the onset of SS7 overload it can be quite slow to react in situations where only the SCP becomes overloaded – as indicated in Section 3 this is a drawback of all Window-like IN control strategies. A good solution to this problem is to have a separate IN interactive control comprising a detection algorithm (such as Load Measure Control) in the SCP, which, upon detection of SCP overload, sends a control message to all SSPs defining the level of SCP overload. At the SSP a set of throttles (separate from those managed by the new SS7 controls above) are then activated to operate on all IN service types.

The result of these two algorithms working in parallel is that:

1. When the SS7 is overloaded but the SCP is not, response delays to initial messages will increase and this will be reflected in the value of C_x for each service type. Throttles will therefore be put in place for both telephony and IN services.
2. When the SCP is overloaded, but the SS7 is not, a control message from the SCP will cause throttles to be put in place for all IN services. Telephony calls will not be throttled.
3. When both SS7 and SCP are overloaded, SS7 controls will put telephony service throttles in place, while both algorithms will put separate throttles in place for the IN calls resulting in calls of this type being subject to two independent throttles.

It should be noted that throttling occurs only on the initial message of a service session from the SSP. There are two reasons for this:

- a) It is a general rule in congestion control theory that once processing time has been expended on a service session, it must be allowed to complete successfully to make the most efficient use of system processor resources.
- b) If there are multiple query/response pairs between an SSP and the SCP, this generally signifies SRF interaction with the calling party. Therefore, if the response to the first query of a service of this type is returned successfully before the corresponding timer expires, interaction will take place with the caller. As outlined in [MacD94], this should increase caller patience and, by extension, the probability that the caller will not abandon after this point.

6.1. Evaluation of suggested Approach

The two focused overload cases discussed in Section 5 demonstrate the important role played by the relationship between the capacity of the SCP and the capacities of the signalling link pair connecting it to the rest of the SS7 network. If the capacity of the SCP is greater than that of the two links combined, then link overload may occur without SCP overload, but if the SCP capacity is less than that of the links, both will tend to overload. In order to evaluate the new controls, their operation was simulated over a range of values of relative SCP/SS7 link pair capacity (by simply varying the SCP service rate), where in all cases the level of overload was enough to trigger SS7 link overload. It should be noted that in the case where the relative capacities and level of overload are such that only the SCP overloads, results not presented here show the algorithm to be similar in operation to previously studied interactive IN congestion control strategies ([Lodge94]).

The criteria used for evaluation of the new control strategy were based on its ability to satisfy the following objectives during overload periods:

1. Load of the SCP and signalling links should be kept at acceptable levels. For the SCP the load should not exceed 0.9 Erlangs (the de-facto level for overload thresholds), whilst the SS7 link load should be in the region of 0.4 to 0.5 Erlangs (to cope with the possibility of a doubling of this value due to link failure).
2. Service completion rates should be kept as high as possible to maximise generated revenue and customer satisfaction.
3. Response delays for users should be kept within acceptable bounds.
4. The algorithm should be fair in its treatment of services – if a service type is not the cause of the overload then throttling of requests for this service should be kept to a minimum.

Figure 6 shows the mean values (during overload) of the load of the SCP and one of the SS7 links, over a range of relative SCP/SS7 link capacities, with the suggested control strategy in turn absent ('Old Controls') and present ('New Controls').

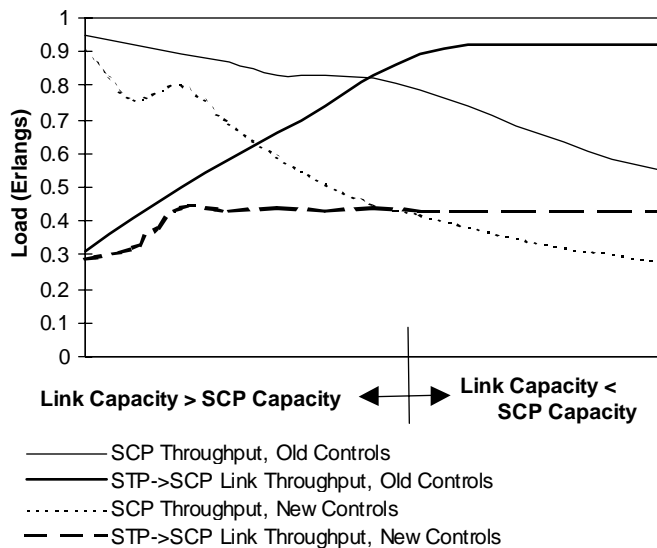


Figure 6: SCP / SS7 Link Loads

The results presented in Figure 6 can be evaluated by considering at the two distinct regions of the graph,

- SCP Capacity < Combined Link Capacity: When the new controls are absent the oscillatory behaviour described in Section 5 occurs, hence there are high values for SCP mean load and, although the values for link load are lower, undesirable oscillations occur. However, when the algorithm is in place, for all but very low ratios of SCP to link capacity, the load of the links is maintained at approximately 0.4 Erlangs, with the SCP load varying appropriately.
- SCP Capacity > Combined Link Capacity: this region of the figure corresponds to the case where only the SS7 links become (and remain) overloaded, with the result that when the algorithm is not in place the mean link load is above 0.9 Erlangs. For the case where the algorithm is in place, it is once again seen to behave as desired – average link load is maintained at approximately 0.4 Erlangs, which means that SCP load stays well below its overload threshold.

Figure 7 presents a diagram of the SCP load for a value of SCP/SS7 link capacity of 0.65 (because this value is less than 1, overload of both SS7 links and the SCP will occur under the Old Controls). The two traces represent the situation with and without the new controls in place. This figure indicates that the new controls keep the load at a near constant value, avoiding any variances which might push it over the overload onset threshold.

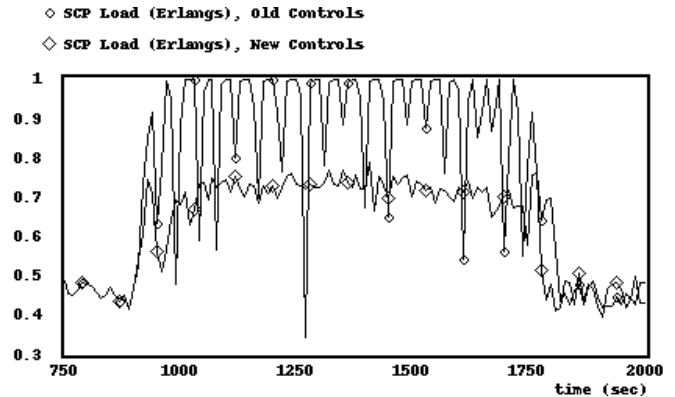


Figure 7: SCP Loads, Capacity Ratio 0.65

Figure 8 presents the service completion rates (expressed as percentages of total service requests) during the overload period for the freephone and televoting services (with new controls absent/present) over a range of relative SCP/SS7 link capacities. Service sessions where the response to the initial message exceeded predefined thresholds (3s for freephone, 4.5s for televoting) were modelled as abandonments, thus the values represented correspond to successful completions.

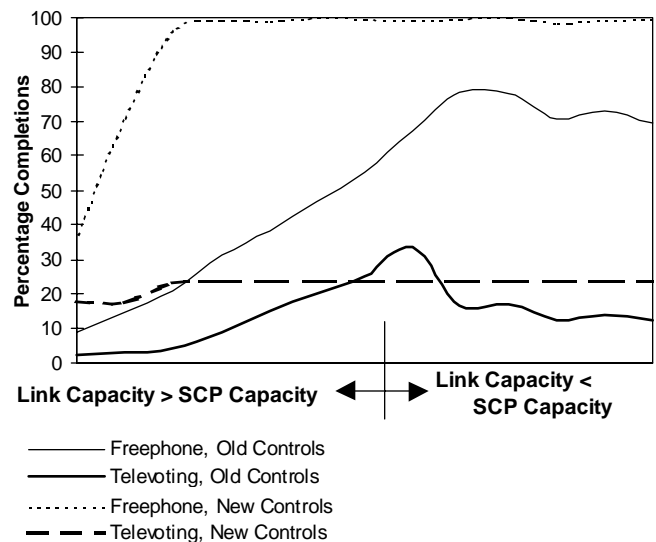


Figure 8: Service Completion Percentages

A comparison of the service completion rates plotted in Figure 8 shows that the presence of the new controls results in more completions (and hence more generated revenue) for the majority of relative capacity values. It should be noted that the seemingly low level of televoting completions is due to the magnitude of the overload which is in any case too great to be handled by the network. For the freephone service the algorithm provides very high levels (near 100% in most case) of successful completion during a televoting-induced overload. This is because the use of independent controls for different service types means that in general only televoting requests are throttled. This attribute of the new controls ensures fairness of treatment between different service types.

The only region of the graph in which the algorithm behaves badly with respect to teletovoting completions is in the area where the capacities of the SCP and its incoming SS7 links are well matched. In this situation the SCP and SS7 will tend to overload almost simultaneously, resulting in invocation of all the throttles for IN services together. However, because of the matched capacities, only the invocation of the SCP overload detection mechanism and associated throttle (i.e. the one operating on all IN services) at the SSPs is needed to solve the problem. Therefore, when the other throttles are also in place, the traffic level tends to be over-controlled. This result would seem to indicate that if the network is properly dimensioned to ensure a matching between SCP and SS7 link capacities, the suggested strategy would not be required. However this is not the case, because if such an approach were to be taken there would be many significant drawbacks: this sort of dimensioning may prove difficult in a dynamically changing network environment; the SS7 links will be running at higher than recommended load levels during overload; and services not contributing to the overload will be unfairly penalised. An additional advantage of the suggested strategy is that it would provide significant protection from other problems arising from overloads caused by non-IN services.

7 Conclusion

The study described in this paper shows that there is potential for severe performance degradation of IN services during focused overloads caused by SS7 network congestion. It demonstrates that in such scenarios SS7 and IN based congestion controls do not interact desirably, with the result that user delays become large enough to cause very high levels of service session abandonments. Although focused overloads of this nature occur infrequently, it is important that they be dealt with effectively, not only so that generated revenue can be maximised but also to avoid any adverse publicity that may affect the network operator's public image.

An integrated SS7/IN congestion control strategy which maximises re-use of existing control algorithms (whose efficiency has been verified in previous studies) was presented and its performance was compared to that of existing IN and standard SS7 control strategies using a simulation model. The results acquired show that the suggested strategy outperforms existing strategies in terms of maintaining resource loads at desired levels, maximising call completions, minimising user delays and ensuring fairness in its treatment of separate service types. A key factor that the study highlighted is the importance of the correct dimensioning of IN resources relative to SS7 resources in preventing congestion-related performance degradation.

Future work will concentrate on analysis of the suggested strategy under a broader range of network congestion scenarios, such as those where SSP and or STP processors become overloaded as a result of increased traffic levels. There is also potential to supplement the throttling algorithm with intelligence which will enable it to optimise network revenue during periods of high traffic loading.

References

- [Atai96] A.H. Atai, B.S. Northcote, "AIN Focused Overloads - A review of USA CCS network failures and lessons learned," ITC Mini-Seminar on Engineering and Congestion Control in Intelligent Networks, Melbourne, April 1996.
- [E.721] ITU-T Recommendation E.721, "Network Grade of Service Parameters and Target Values for Circuit-Switched Services in the Evolving ISDN."
- [Lodge94] F. Lodge, T. Curran, M. Gulyani, A. Newcombe, "Intelligent Network Congestion Control Strategies and their Impact on User-Level Quality of Service," Proceedings of Australian Telecommunication Networks & Applications Conference, pp627-632, Melbourne, December 1994.
- [Korner91] Ulf Korner, "Overload Control of SPC Systems," Proceedings of ITC-13, pp 105-114, Copenhagen, 1991.
- [MacD94] D. MacDonald, S. Archambault, "Using Customer Expectation in Planning the Intelligent Network," ITC-14, Juan-les-Pins, 1994.
- [McMillan96] D. McMillan, M. Rumsewicz, "Analysis of Congestion Control for SCCP Traffic and the Impact on Intelligent Network Services," Intelligent Networks '96, Melbourne, April 1996.
- [Manfield93] D.R. Manfield, G.K. Millsted, M. Zukerman, "Congestion Controls in SS7 Signaling Networks," IEEE Communications Magazine, June 1993.
- [Manfield94] D.R. Manfield, G.K. Millsted, M. Zukerman, "Performance Analysis of SS7 Congestion Controls under Sustained Overload," IEEE Journal on Selected Areas in Communications, Vol. 12, No. 3, April 1994.
- [Pham92] X.H. Pham, "Congestion Control for Intelligent Networks," Proceeding of 1992 International Zurich Seminar on Digital Communications, Intelligent Networks and their Applications, Zurich, 1992.
- [Nyberg94] H. Nyberg, B. Olin, "On Load Control of an SCP in the Intelligent Network," ATNAC, Melbourne, 1994.
- [Zepf94] J. Zepf, G. Rufa, "Congestion and Flow Control in Signalling System No. 7 - Impacts of Intelligent Networks and New Services," IEEE Journal on Selected Areas in Communications, Vol. 12, No. 3, April 1994.