CONTROL LOOP FOR TRAFFIC MANAGEMENT OF NETWORK UNDER FOCUSED OVERLOADS

Xuan Huy Pham

Standard Elektrik Lorenz - ALCATEL-, Stuttgart
Federal Republic of Germany

Network Management is an indispensable part of any successful telecommunications network. It consists of realtime surveillance and control of the network. This paper describes an effective dynamic control loop to ensure a stable and good Grade of Service in network under focused overload. Aspects which have qualitative and quantitative influences on the declination of the network goodness will be first studied. It will be shown that network performance deterioration phenomena may be classified into some distinct congestion types, each of them may be dealt by appropriate control measures. The most important control is the call rate throttling mechanism. Two well known principles, the Call Gapping and the Leaky Bucket principle, will be analysed and compared. The way how parameters of the used controls have to be dynamically calculated will be also one of the subject of the paper. Simulation results on a hierarchical 3-levels network containing exchange with distributed control architecture should verify the effects of the dynamical control loop.

1. INTRODUCTION

In telecommunication networks, resources such as exchanges, trunks and other facilities are planned and dimensioned so that grade of service is guaranteed for a certain load level corresponding e.g. the main busy hours. Practically loads offered to the network are subject to great variation often exceeding the dimensioned level. Such traffic disturbances may be classified into two classes of problems.

a) Overload problems due to e.g. natural disasters, holidays, failures, telecasting directed to broadcasting stations... The last mentioned problem shows an increasing tendency more and more.

b) Change of load profile due to the introduction of new network services, of network applications such as Intelligent Network, Private Virtual Network which may impact the existing overall public telephone network. Another reason influencing also the load profile is the change of user behaviour.

Such traffic problems cause a severe service degradation resulting in low completion rates, low revenue and long setup time. To cope with the above situation, network management centers with appropriate operations are required. In order to ensure the effective utilization of the network resources in any traffic situation, dynamic control loops under guidance of network management centers are needed performing the following tasks:

- Real-time supervision of the network
- Analysing for detection of traffic trouble situations
- Help the network manager by activation of suitable controls
- Regulate the loop for better control.

In practice the design and application of control loop is faced with some difficulties concerning the specification of load items to be observed, the analysis algorithm on these and the decision of effective control to be activated. Further difficulties consist e.g. of the calculation of parameters and the determination of location in the network at which the control is to be started.

2. NETWORK CONGESTION CHARACTERISTICS

Detection and understanding of causes of network troubles is one of the most important tasks for designing effective control loops. Several works about the phenomena in network during severe overload are published e.g. in /2/, /3/, /4/... In this chapter phenomena related to focused overload are described in a way, so that suitable controls may be derived. Large scale simulations are used to verify that. Traffic congestion considered here consists of focused overload characterized by a high increase of call attempts from a wide range of the network and destined to one certain exchange.

2.1 NETWORK GOS DEGRADATION

Degradations of network GOS under focused overload is expressed by evaluating the following CCITT items /3/:

- Network throughput
- Network revenue
- Setup time for successful calls
- Utilization profile of network resources

2.1.1 THROUGHPUT DEGRADATION

Throughput degradation is a well known phenomenon expressed by a fast decrease of the overall completion rate with increasing call attempts towards the considered focused area (fig 2.1).
b) Increase of setup time

Fig 2.1 Network GOS time under focussed overload

This deterioration of throughput can be explained related to the following knowledges:

- There are network wide two congestion mechanisms, a trunk congestion and a switching congestion. Trunk congestion is characterized by a high blocking due to the lack of lines and switching congestion by a high blocking due to the limitation of dynamical switching resources. Trunk congestion occurs in light overload level. For higher load the switching congestion dominates more and more. This effect is shown in fig 2.2 representing the network rejected call rate classified by its reject cause. Congestion occurs first in the direct environment of the focussed area, from which it extends rapidly to the remaining part of the network.

- The main cause of the decrease of the network completion rate is however the contention of resources. Due to large number of call attempts towards the HTR (Hard To Reach) destination code, an essential part of network use (specially long distant calls) will result in blind load. Therefore ETR (Easy To Reach) call attempts towards non overloaded codes will be considerably affected in respect of completion rate, see fig 2.1 a).

- An other effect influencing the declination of network throughput is the call reattempt rate caused by blocked calls. The magnitude of repeat rate depends in general on the behaviour of users, on the considered overload problem and also on the feature of subscriber type used. High repeat rate up to 20% is often registered during focussed overload /10/. In that situation the throughput variation may show hysteresis effect.

\[\text{Fig 2.2 Reject rate classified by its cause}\]

### 2.1.2 INCREASING OF SETUP TIME

Setup time as the duration a subscriber has to wait after off hook until a connection is setup is an important item presenting the actual network goodness. Setup time varies also greatly during overload situations. It depends on the state of the switching system dealing as large scale queuing networks. Results show that with increasing HTR call attempts the mean setup time for completed calls becomes greater up to 10 times of the dimensioned one while for incomplete calls it falls back after a peak value (see fig 2.1 b).

### 2.1.3 INEFFECTIVE UTILIZATION OF LINKS

An additional bottleneck observed during high overload is the ineffective usage of connection links. The frequency that calls take an alternate route during its setup phase increases with growing input call rate. At the same time the quota of successful calls decreases considerably. This means that alternative routing results more and more in blind traffic. This effect is specially distinct for HTR traffic streams. The success quota of long distant call attempts decreases stronger in comparison with those of shorter distance.

### 2.1.4 DECREASE OF REVENUE

Decrease of revenue is of course a consequence of the throughput decrease. However the declination may be more critical due to the strong deterioration of long distant call attempts. This emphasizes the prior goal of Network management from the operator point of view is to keep the network revenue at an acceptable level.

### 2.2 INFLUENCING ASPECTS

There are some impact factors influencing the speed and grade of the deterioration of network GOS. Planning aspects which ensure an optimal GOS in normal load level will result in opposite effect in overload due to the exhaustion of network resource. Therefore principles which allow a high degree of freedom for resource access become ineffective. Impact factors such as network architecture, network dimensioning and the used signaling technique are fixed elements, which cannot be easily adjusted dynamically during network operation. Other impact factors such as routing procedure, call setup procedure and switching system characteristics allow some regulation possibilities. Following experiences may be noted:

- Network configuration: network with hierarchical routing has more reserve as non-hierarchical meshed network. The congestion point identifying the begin of the fall of GOS will be first reached at a larger number of offered calls.

- Network dimensioning: An overdimensioning of trunk resource and switching capacity is no solution for heavy focussed overload. Tab 2.1 shows that an overdimensioning of 20% in an overload level of 500% causes stronger throughput deterioration (10%) compared with the normal dimensioned level. This is due to the fast increase of call attempts with short holding time on trunks which result in blind load.

<table>
<thead>
<tr>
<th>Dimensioning</th>
<th>Throughput in [CAPS]</th>
</tr>
</thead>
<tbody>
<tr>
<td>level in %</td>
<td>100% load 250% load 500% load</td>
</tr>
<tr>
<td>normal</td>
<td>9.43</td>
</tr>
<tr>
<td>110%</td>
<td>9.66</td>
</tr>
<tr>
<td>115%</td>
<td>9.71</td>
</tr>
<tr>
<td>120%</td>
<td>9.77</td>
</tr>
</tbody>
</table>

10% overdimensioning means 10% more trunks and node capacity

Tab 2.1 Influences of overdimensioning on throughput
Signaling technique and call setup procedure have an impact on the network performance so far it specifies how the resource will be seized during the setup phase. The use of Nr. 7 ISDN signaling technique e.g. is more effectivite because the more efficient channel assignment procedure. Other impact factor are e.g. time-out conditions. Time-out values are in general installable. Great time-out show advantage in underload, short time-out in overload situation.

Routing strategies: non-hierarchical routing scheme has a lower blocking rate than hierarchical routing in underload domain. In strong overload it results in opposite effect. The more alternate routes exist or the more exchange resource is shared, the more rapid the network throughput will fall in overload situation.

Switching system structure has an influence on the speed of switching congestion. Switching system with fully distributed control as considered in the paper has better characteristics in focused overload through the isolation of bottleneck on some control elements. The other impact factor is the local overload strategy which consists mainly of protecting the exchange against too many call attempts. Thereby neither a selection of calls by its destination code nor a use of network wide criterion for blocking judgement is made. The installation of parameters of local overload control has of course as shown in Tab 2.2 some influences on the throughput.

<table>
<thead>
<tr>
<th>OR</th>
<th>IC</th>
<th>OG</th>
<th>TG</th>
<th>100%</th>
<th>250%</th>
<th>500%</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>15(S) 20</td>
<td>20(S)  20</td>
<td>15(S) 20</td>
</tr>
<tr>
<td>R</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>0.40(R) 0.45(R)</td>
<td>20 (S)  20</td>
<td>0.30(R) 0.22(R)</td>
</tr>
<tr>
<td>0.30(R) 0.22(R)</td>
<td>20</td>
<td>7(S)  9.8</td>
<td>8.5</td>
<td>7.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.30(R) 0.22(R)</td>
<td>20</td>
<td>5(S)  9.6</td>
<td>8.6</td>
<td>7.4</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2.2 Influences of overload control parameters on throughput

In the table control at different stages such as OR (Originating), IC (Incoming), OG (Outgoing), or on TG (Trunkgroup) are considered. 100% load corresponds to the normal dimensioned load level, 15(S) means a State limiter on 15 Call attempts per Control element or Trunk group, 0.8(R) means a Rate limiter on one call request per 0.8 sec. Tab 2.2 shows:

The installation of local overload control has impacts on the performance only in normal load. In overload however it cannot replace network management control. A strong limitation of incoming traffic causes a needless blocking of calls in underload level. A light limitation of incoming traffic causes high completion rate in underload but strong declination in overload. Delimiter on a rate limiter basis has a better tuning than a state limiter, therefore is more effective for parameters adjusting in normal load level. The use of trunkgroup control through limitation the number of setup requests in outgoing trunk group shows a surprising positive effect in overload which may prevent strong throughput deterioration.

2.3 NETWORK SIMULATION MODEL

Simulation are done to test the effect of control loop. The considered network structure was defined in accordance to the DBP public Telephone network containing all features such as: Trunks, subscribers, routing scheme and real modeled switching nodes.

2.3.1 NETWORK STRUCTURE

Fig 2.3 shows the considered 3 levels hierarchical network with 3 central toll exchanges, which are responsible each for one of the 3 districts. Each district represents a subnetwork with 4 nodal and 2 main exchanges. Central toll exchanges have only transit traffic. Connection links between Nodes are one-way Trunk groups, which capacity is so defined, that a trunk blocking lower than $10^{-3}$ for busy hour is ensured. The routing scheme uses corresponds to a hierarchical routing allowing an number up to 6 alternate routes.

2.3.2 QUEUEING MODEL FOR EXCHANGES

Exchange nodes with distributed control according to ref./1/ are assumed. Essential functions such as local setup procedure, local routing processing and local overload control are considered. Controls for Network traffic management are also introduced adjusted to the distributed structure. A corresponding queueing model satisfying the mentioned aspects is shown in fig 2.4.

CALL HANDLING PROCEDURE

Subscriber Call attempts will be first entered into the peripheral Control Elements. It will be than transmitted to the associated Line- (LACE) or Trunk Control Element (TACE) according to its originating source. A free Process Control Block (PCB) is looking for, whose task is the coordination of the whole local setup procedure within the exchange. The next step is the processing of dialed digits within the PSAC Control Element. Digits may be processed on block or digit per digit. In the last case it is considered that the interarrival time for digits is described by a normal distribution with a mean value of 1 sec. PSAC forwards then terminating calls to the appropriate LSAC and transit calls to the corresponding TSAC. LSAC assigns the calls through its numbering code to the related outgoing LACE (now working at B-side), from which a connection to the end subscriber is done. For transit calls a same procedure is carried out by TSAC which hunts a free trunk to the next node. The call setup process will continue in the next station in the same way until a connection through the network is build. PSAC and LSAC are modeled as a multi server system due to its load sharing scheduling principle. The service time of CEs are assumed as neg. exp. distributed whose mean values are got by measurements in laboratory. The distribution of originating calls is of type poission. The number of CEs per type is calculated in relation to the number of lines and/or trunks installed. It is further assumed that blocked calls have a repeated rate probability of 10%. Following local overload controls are considered in the model:

Peripheral Control Elements control the input and output traffic by a state or rate delimiter. State limiter means e.g. that only a limited number of call requests is admissible at the same time.
Fig 2.4 Queueing model of exchange with distributed control

Rate limiter means e.g. that only a fixed number of call requests per time unit is accepted. Rate limiter implemented in the model works following the Leaky Bucket principle.

- LACE and TACE Control Elements process as well incoming as outgoing calls. Incoming calls are to be rejected if the number of PCBs occupied exceeds a defined threshold. The threshold must be so determined, that outgoing calls can never be blocked due to the lack of PCB. A call request has to be completely setup within a defined holding time limit. Outgoing calls receiving no free trunk after exhausting all defined alternate routes will be blocked. Blocking due to the bottleneck within the digital switching network is assumed to be negligible.

- Trunk Group control consists of the limitation of the number of setup processes occurring at a same time in each trunk group. On this way blind load caused by a high access rate on trunk groups directed towards the focused area may be restricted.

NETWORK MANAGEMENT controls considered in the model consist of: Routing control such as e.g. the restriction of alternate routing or the limitation of the number of hops are implemented in TSAC control elements. Call rate control is done selective on destination code basis after analyzing the numbering code in the PSAC control elements. Trunk access control on basis of trunk reservation are handled in the outgoing TACE control element.

3. CONTROL LOOP FOR FOCUSED OVERLOAD

Control loop is defined as a chain of iterative and independent tasks performed periodicaly e.g. each 5 minutes in the whole network during its operation, see Fig. 3.1. The results of each task are informations, which are needed for triggering the next task. The task for measurement has to be performed on the exchange level. The other tasks such as Data analysis, Network state presentation, problem detection and decision rules have to be carried out in Traffic Management Centers. Some solution approaches for control loop may be seen in [4], [5], [6]. Activities regarding the standardization of interfaces (Q3 interface) for network management in the scope of the Telecommunication Management Network (TMN) are actually worldwide in working [7].

Fig 3.1 Principle and components of a control loop

3.1 DETECTION OF FOCUSED OVERLOAD

Detection of focused overload is one of the most important element of the control loop. It needs analysis algorithms fullfilling the following goals:

- definite and fast detection of HTR destinations
- detection of trunkgroup congested
- detection of nodes congested
- detection of routing congestion

Detection algorithms require however that the network state must be observed permanently. Due to reason of systematic it should be classified into a range of network objects:
Items for the object Call Streams $X_{ij}$ are:

- $X_{ij}$ offered call rate from origin $i$ to destination $j$
- $X_{ij}$ seizure call rate from origin $i$ to destination $j$
- $X_{ij}$ answer call rate from origin $i$ to destination $j$
- $X_{ij}^j$ part of call rate from origin $i$ to destination $j$ blocked due to the reason 'c' (= 'i':incoming limiter, 'd':lack of dynamic resource, 't':lack of trunk or line, 'o':outgoing limiter, 'n':Network Management controls)

$T_{ij}$ setup time for call attempts from origin $i$ to destination $j$ registered in node $e$

Items registered for the object Trunkgroup $tg_k$ are:

- $X_{tg_{succ}}$ successful call rate on $tg_k$
- $X_{tg_{rej}}$ rejected call rate on $tg_k$
- $T_{tg_{succ}}$ Portion of holding time for successful calls on $tg_k$
- $T_{tg_{rej}}$ Portion of holding time for blocked calls on $tg_k$

Items registered for the object Node $e$ are:

- $X_{nod}^j$ offered call rate observed at the node $e$
- $X_{nod}^j$ seizure call rate observed at node $e$
- $X_{nod}^j$ answer call rate observed at node $e$
- $X_{nod}^j(c)$ blocked call rate from origin $i$ to destination $j$ due to the reason $c$ as described for object stream

$T_{nod}^j$: Setup time for successful call attempts in node $e$

$T_{nod}^{rej}$ Setup time for blocked call attempts in node $e$

Items registered for the object routing are:

- $Palt_{ij}$ Percentage of call rate from $i$ to $j$ attempting one or more alternate routes
- $Palt_{ij}^{succ}$ Percentage of successful calls from $i$ to $j$ attempting one or more alternate routes

Detection algorithm of HTR streams: A stream is identified as HTR stream if its completion ratio is considerably low. This fact is expressed by a Reject Answer Ratio (RAR), whereby the portion of rejects due to TM controls are to be excluded.

$$RAR(X_{ij}) = \frac{\sum X_{rij}(c)}{X_{a_{ij}}}$$

Experiences with simulation show that RAR algorithm is a very effective rule allowing a fast and definite detection of true HTR. RAR is also able to detect several HTR streams occurring at the same time in the network. Other analysis methods based e.g. on the seizure bid ratio or the answer seizure ratio may be also used for recognition of the presence of HTR problem, but not suitable for the definite identification of HTR code. The analysis of stream setup time is also suitable as well for the detection as for the identification of HTR streams. The use of setup time has of course a disadvantage regarding the variable threshold, which is in general different from one call stream to the other. The detection of Trunk congestion is based on the analysis of $Y_{tg_k}$ defined as the quota of trunk group related blind- to its carried load

$$Y_{tg_k} = \frac{X_{tg_{succ}} \cdot T_{tg_{succ}}}{X_{tg_{succ}} \cdot T_{tg_{succ}} + X_{tg_{rej}} \cdot T_{tg_{rej}}}$$

The use of $Y_{tg_k}$ for the judgement of the effectiveness of trunk-group has been judged as very practical. It is a good indicator for trunk congestion occurring in the early phase of focused overload. Overload in nodes may be detected in the same way as for stream by analyzing the total Reject Seizure Ratio (RSR) excluding those caused by TM controls

$$RSR(c) = \frac{X_{nod}^j(c) \neq 0}{X_{nod}^j}$$

Node congestion as observed in simulation is a consequence of focused overload, if no control is activated in time. Congestion appears first in the nodes adjacent to the HTR node expands then to the remaining part. Routing congestion indicates an ineffective usage of routing expressed by the completion ratio $U_{ij}$

$$U_{ij} = \frac{Palt_{ij}^{succ}}{Palt_{ij}}$$

characterizing the success quota of accesses on alternate routes.

Fig 3.2 Surveillance items for detection of congestion

3.2 DECISION RULES FOR NM CONTROLS

The classification of the effects related to focused overload into separate phenomenon done in section 3.1 allows a practicable assignment of suitable control actions. In Tab 3.1 welltried control measures are presented. Theoretical Approaches through e.g. definition of constraints have been done previously by many authors as in /5/, /7/ and /10/ etc.

TRUNK RESERVATION: Trunk Reservation is applied selective for HTR Streams i.e. about $0.5\sqrt{n}$ trunks, see /6/ with $n$ as the size of the corresponding trunk group is reserved for others non HTR streams.
**ADJUSTING** the parameters of the local overload control consists of two tasks: a) increase the threshold of the incoming call rate limiter at an appropriate limit (e.g., 0.6 CAPS) and b) activate the trunk group control to limit the number of setup processes in the overloaded trunk groups.

**CANCELATION** of alternate routing means that streams which have been identified as HTR, have no more access on alternate route.

**RESTRICTION** of number of paths is applied on HTR streams and means that only a maximal number of paths is allowed. The definition of the limitation depends on the origin and destination of call streams.

**CALL RATE THROTTLING** is the most effective protective action for control severe focussed overload. There are several well known call rate control mechanisms such as e.g. the window mechanism used for flow control in protocol or others general call rate control as Automatical Call Gapping (ACG) and Automatical Leaky Bucket (ALB). The notation automatic refers to the dynami adjustability of its parameters. The last both mechanisms are known call rate control mechanisms such as e.g. the window mechanism if it is permanently invoked blocks a considerable number of calls even in normal load level. It allows e.g. only 50% of incoming rate at the transition point 1/T_{CG}. The parameter KLB of

### 3.3 APPLICATION OF THROTTLING MECHANISMS

### 3.3.1 PERFORMANCE COMPARISON

Call Gapping works so, that a gap interval timer T_{CG} is started after the first call request is accepted. Calls incoming while this duration will be rejected. The next call after expiration of T_{CG} will be accepted again. The call rate estimated after filtering through Call Gapping is expressed by the term

$$\lambda_{CG}^* = \lambda'_{in}/(1 + \lambda'_{in} \cdot T_{CG})$$

with $\lambda'_{in}$ input rate $T_{CG}$ Gap time parameter of Call Gapping

The call accept condition of Leaky Bucket is working so, that the number of Call requests passed into the Leaky Bucket is registrated by a counter. If the counter is greater than a value $S_{LB}$ incoming call requests will be blocked. After each time interval $T_{LB}$ the counter will be reduced on $S_{LB} - K_{LB}$ (so far the counter remains >= 0). The corresponding throughput rate for Leaky Bucket is presented by

$$\lambda_{LB}^* = \lambda'_{in} \cdot (1 - Pb(S_{LB}, K_{LB}, T_{LB}))$$

with $S_{LB}$ Limit Counter $K_{LB}$ Update factor after each time interval $T_{LB}$ $T_{LB}$ Update time interval

where Pb($S_{LB}, K_{LB}, T_{LB}$) is the loss probability formula for Leaky Bucket by consideration of Poisson input process. The terms $S_{LB}$, $K_{LB}$, $T_{LB}$ are the parameters of the leaky bucket mechanism. A detailed Analysis of this expression may be found in [8], where an imbedded Markow approach has been applied. Referring to the fig. 3.2 it may be noted, that Call Gapping has an advantage in comparison with Leaky Bucket in respect of the number of parameters to be set. In high load level a maximum rate of only one call per T_{CG} is allowed. Leaky Bucket principle however needs 3 parameters. $T_{LB}$ corresponds approximately to T_{CG}, $S_{LB}$ represents somewhat the volume of the bucket and is responsible for the accepting of the natural variation of input call rate therefore is an improvement of throughput in comparison to Call Gapping. The Call Gapping mechanism if it is permanently invoked blocks a considerable number of calls even in normal load level. It allows e.g. only 50% of incoming rate at the transition point 1/T_{CG}. The parameter KLB of

### CALL GAPPPING

**Saturated accepted call rate = 1/T_{CG}**

### LEAKY BUCKET  

(example with $S_{LB}*4$, $K_{LB}*2$)

saturated accepted call rate = $K_{LB}/T_{LB}$

$\overline{\delta}$ : accepted call  \( \overline{\delta'} \) : rejected call

a) principle
leaky bucket is responsible for the leakeness and may lead to bursty traffic if it is set with a great value. Experiences shows of course that with $S_{LB} = 5$ and $K_{LB} = 1$ the Leaky bucket shows good performance characteristics. Leaky Bucket allows in high load level a maximum rate equal $K_{LB}/T_{LB}$. The quasi self-regulation effect of Leaky Bucket e.g. at the load transition point is an important advantage allowing high flexibility and more security for application in dynamic control loop. In this sense Leaky Bucket may even be activated more early and deactivated more late as necessary.

b) Performance comparison

Fig 3.2 Comparison of call rate throttling mechanisms

3.3.2 ACTIVATION PROCEDURE

If the goal of network planning is to specify how and how much trunks and exchange capacity are to install to meet a given GOS, thus the situation to deal here seems somewhat reverse. The solution to look for now is, what traffic stream has to be controlled in which way, in order to keep the performance of existing network optimal and stable. An appropriate formulation of the problem is to drop call streams directed towards a focussed overload point, at a level on which the network show its optimal behaviour. The idea is to act and to react closely on the real network state by taking into account an objective function which goal is the maximization of the network tariff rate expressed by $\sum R_{ij}$

$$\text{MAXIMIZE} \left( \sum_{W_{ij}} R_{ij} \right) = \sum R_{ij}^{opt} \quad (7)$$

$R_{ij}^{opt}$ is the target optimal tariff completion rate.

$$R_{ij}^{opt} = X_{ij}^{opt} \cdot C_{i} \cdot TF(X_{ij}) \quad (8)$$

$X_{ij}^{opt}$ the optimal target completion rate of $X_{ij}$

$C_{i}$ adjusting factor for Leaky bucket (1.0 to 1.4)

$TF(X_{ij})$ Tariff function for stream $X_{ij}$

Once a stream is identified as HTR a calculation of the desired completion call rate for this stream is started. A selective activation of Leaky Bucket for the considered HTR stream will be carried with the following parameter

$$T_{LB} \text{ for } X_{ij} = 1/R_{ij}^{opt} \quad (9)$$

Regarding the question at which point in the network the Leaky Bucket has to be activated effectively, following experiments have been done.

- Throttling of HTR call streams only at it originating node
- Throttling of HTR call streams at all nodes (originating, transiting or terminating nodes), where they are detected

Results obtained from simulation confirms the well known statement that the most effective control point is the originating node.

3.4 SIMULATION RESULTS

The effects of the mentioned control loop should be discussed now. The characteristics of network throughput and setup time versus the increasing total input rate is shown in fig 3.4. An overall network GOS improvement is documented. Especially the GOS of ETR traffic shows now a stable state in high overload. Throughput improvement in the overloaded area may be even noted due to the utilization of reserve resource. Also setup holding time for accepted calls have reasonable values.

Fig 3.4 Network GOS under focussed overload with control

In fig 3.5 the effects of control loop in a real time environment is shown. It is assumed that traffic streams to the focussed overload area (node 13 in the model) have two peaks occurred sequentially in a time interval of ca 15 minutes. The first peak has a duration of ca 30 minutes the second peak with a less magnitude has a duration of ca 20 minutes. Two types of control loop has been applied: a) a manual control loop i.e. a leaky bucket with fixed call rate is used, which value has been derived from the planning data and b) a dynamic control loop with flexible handling of controls in respect of parameters calculation and activation time. The dynamical control
loop works with a period time of 5 minutes. A comparison of them shows that dynamic control loop has a better performance. It may be explained by the fact that dynamical control loop takes advantage of all available network resource at the light overload level. In high overload it reacts more closely to the network real situation through throttling of HTR traffic in time.

CONCLUSION

A range of impact factors leading to a declination of network GOS under focussed overload were investigated. Aspects considered by network planning such as network topology, routing schema, call setup procedure, dimensioning method, exchange internal structure and its overload control have direct influences on the speed and magnitude of the deterioration of GOS. A large-scale planning of these cannot prevent congestion due to a focussed overload of up to 250%. An effective way to control that is the introduction of a control loop with the following features:

- Surveillance of network performance state by detecting separately Trunk-, Routing-, Node- and Stream congestion. The last is the most important element for handling of focussed overload.
- Use of automatical Leaky Bucket for dropping and regulating HTR streams closely to the realtime network state and fulfilling a determined objective function.

REFERENCES

/10/ Lewis A., Leonard G.: Measurements of Repeat Call Attempts in the intercontinental Telephone Service ITC 10, Montreal, June 1983.

Abbreviation und Symbols list

ASR Answer Seizure Ratio
ACG Automatical Call Gapping
ALB Automatical Leaky Bucket
CAPS Call Attemps Per Second
CE Control Element
DBP Deutsche Bundes Post
ETR Easy To Reach
GOS Grad Of Service
HTR Hard To Reach
LACE Line Auxiliary CE
LSAC Local Subscriber Identification System
Auxiliary CE
PSAC Prefix analysis System Auxiliary CE
PCB Process Control Block
SBR Seizure Bid Ratio
TACE Trunk Auxiliary CE
TSAC Trunk resource manager System
Auxiliary CE
TG Trunk Group
TMN Telecommunication Management Network
TM Traffic Management