I/O PROCESSOR PERFORMANCE OPTIMIZATION WITH DIFFERENT AND CHANGING ARRIVAL RATES

Dr. Herbert Jans and Jörg Oehlerich

Fachhochschule Landshut *) Siemens AG Munich
Dept. of Electrical Engineering Dept. of Public Switching
8300 Landshut 8000 München 70
Federal Republic of Germany Federal Republic of Germany

This paper describes the scan cycle optimization and dynamization of an I/O processor with respect to further increase of capacity, unbalanced load, and defective scanned devices. Formulae for throughput calculations and simulations for verification and delay time determination are presented.

1. INTRODUCTION

Modern electronic switching systems (SPC-systems) are structured and organized as distributed processing systems. Fig. 1 illustrates the basic control structure of the Siemens SPC-system EWSD. Signals of subscribers and trunks are preprocessed by the decentralized control units DLUs (Digital Line Unit) and LTGs (Line/Trunk Group). Messages are sent via the MCHs (Message Channel) through the SN (Switching Network) and are stored in the centralized MBUs (Message Buffer Unit). For ISDN calls signaling messages are interchanged via signaling links and NUCs (Nailed-up Connection) in the SN, which are multiplexed onto the CCNC (Common Channel Network Control). The MBUs and the CCNC are scanned by the IOP (Input/Output Processor) which transfers the messages either to the CP (Co-

Fig. 1. BASIC CONTROL STRUCTURE of EWSD

*) This work was done while Dr. Jans was with Siemens AG, Munich

3.3B.6.1
ordination Processor) for further processing or to the B-part of the connection
(LTG or CCNC). The IOP as part of the CP is the centralized message distributor
between CP and periphery control units. The most commonly used operating mecha-
nism for I/O processors is the polling scheme where all connected devices are
scanned in a cyclic manner and from each nonempty queue (input and/or output
direction) one message is transferred. The performance analysis of such IOPs
with regard to different and changing arrival rates is of great importance for
system dimensioning and optimization.

Modeling of polling systems leads to complex queueing systems with multiple
queues, cyclic service and overhead caused by protocol signaling and physical
switch-over time. Because of the bi-directional message transfer the IOP works
in a combined operation as scanner (input) and distributor (output).

Analysis of queueing models with cyclic service, finite and infinite buffers,
exhaustive, gated, and limited service disciplines, with and without overhead
has been the subject of many papers. A comprehensive analysis of polling sys-
tems can be found in [1] together with an extensive reference list. In classic
polling systems the scan cycle is fix without adaptation to changing arrival
rates which occur e.g. in unbalanced load conditions, load peaks, system con-
figuration changes or system faults in connection with change-over [2, 3].

2. MODELING

2.1. General Model Description

Fig. 2 shows the traffic model of the IOP. The following devices must be
connected:
1. CCNC 0, CCNC 1 (one active, one stand-by)
2. MBU:LTGs (Depending on system size up to 8 MBU:LTGs are connected operating
in load-sharing by pairs).
3. MBU:SGCs (Depending on system size 2 or 4 MBU:SGCs are connected. They are
split into two equal groups operating independently of each other).
4. CCG 0, CCG 1 (Central Clock Generator)
5. SYP (System Panel Control)

Message flow:
1. In input direction orders/reports are stored in IBs (Input Buffer) in CCNC
   and MBU:LTGs. From there they are transferred by the IOP: Orders to the IL
   (Input List) and reports to the TLs (Transfer List) in the CP.
2. In output direction commands/reports are stored in OLs (Output List)/TLs
   from which they are transferred by the IOP to the connected devices.

2.2. Arrival Processes

The arrival processes of the nonidentical devices connected to the IOP can be
characterized as follows:
1. Each call exists of an initial message and several subcall messages.
2. The initial message process is assumed to be Markovian with mean arrival
   rate c (calls/s).
3. In feedback mechanisms subcall messages are generated according to the real
   message sequence of a complete call.
4. As a result the average number of subcalls per call θ (messages/call) of the
different devices varies widely (e.g. 10:1).
5. The subcall arrival rates for CCNC (c θ_{CCNC}) and MBU:LTGs (c θ_{LTG}) can
   change rapidly at any time caused by
   - change-over from active to stand-by CCNC (e.g. failure situation) where-
     by the arrival rate can change from zero to c θ_{CCNC} and vice versa.
   - a fault in one MBU:LTG which doubles the arrival rate c θ_{LTG}
   of the
      partner MBU:LTG (load sharing operation).

3.3B.6.2
6. The arrival rates for CCG 0, CCG 1 and SYP are negligible (\( \beta = 0 \); no call processing, only O&M messages).

Fig. 2. IOP SIMULATION MODEL

2.3. IOP Service Process

Each device has to be scanned by the IOP at least once in a scan cycle. From the addressed device at most one message in input and output direction is transferred. The service (transfer) time per addressed device consists of two parts:

1. A constant overhead time \( h_{OH} \) for switch-over from the preceding device and
2. A constant transfer time depending on the source queue status:
   - \( h_{I-} \) (Input, no Output)
   - \( h_{I-O} \) (no Input, Output)
   - \( h_{I+O} \) (Input, Output)

The OL and TL for each MBU:LTG are scanned in an alternating manner. If the first addressed list is empty the IOP changes to the other one automatically. This strategy reduces the total overhead time.

3. ANALYSIS

3.1. Scan Cycle Optimization

Starting point of the performance optimization was a system with channel associated signaling. In this case the subcall arrival rate per device is nearly the same and therefore in the IOP firmware a scan table was stored which consisted of just one address per device. With the introduction of ISDN and the system extension with CCNC (2 devices) the number of subcalls to be handled by
the IOP increases rapidly in connection with a very unbalanced subcall arrival rate to the different types of devices. This would have decreased the call throughput of the IOP if the scan table were only extended by the two addresses for the CCNC devices. Therefore an IOP performance optimization was necessary with the following requirements:
- no changing of IOP hardware (e.g. max. 41 addresses/scan cycle available),
- retention of all hardware and software interfaces.

A performance optimization was achieved by a scan cycle optimization:
1. Extension of the scan table whereby the number of addresses for each device was chosen to be proportional to the belonging subcall arrival rate $c_B$.
2. Because of the lowest subcall arrival rate the MBU:SGCs needed only one scan address in the scan cycle.
3. The number of scan addresses for each MBU:LTG had to be chosen according to the change over subcall arrival rate (partner MBU:LTG fails $\rightarrow 2 \cdot c_B$).
4. Because the IOP had no knowledge which CCNC is active or stand-by the number of scan addresses had to be equal for CCNC 0 and CCNC 1.
5. CCG 0, CCG 1 and SYP were scanned with the lowest possible rate, i.e. once per cycle.

3.1.1. Definition of the Optimized Static Scan Cycle

In a first step a scan cycle of fixed length (optimized static scan cycle) was defined. With the traffic parameters of Table 1 and

$$
N_{CCNC} = \text{No. of CCNCs} = 2 \quad A_{CCNC} = \text{No. of CCNC addresses/scan cycle}
$$
$$
N_{LTG} = \text{No. of MBU:LTGs} \quad A_{LTG} = \text{No. of MBU:LTG addresses/scan cycle}
$$
$$
N_{SGC} = \text{No. of MBU:SGCs} \quad A_{SGC} = \text{No. of MBU:SGC addresses/scan cycle} = 1
$$

and according to the arrival rate proportional condition (see 3.1)

$$
A_{CCNC : A_{LTG : A_{SGC} = \frac{G_{CCNC}}{2} \cdot A_{LTG} : A_{SGC}}
$$

the total number of addresses $A_{total}$ per scan cycle is calculated to

$$
A_{total} = N_{CCNC} \cdot A_{CCNC} + N_{LTG} \cdot A_{LTG} + N_{SGC} \cdot A_{SGC} + \text{CCGO} + \text{CCGl} + \text{SYP} = 63.
$$

Because each MBU:SGC, CCG and SYP must be scanned once per cycle and because of the hardware requirement $A_{total} = 41$ only 34 addresses are left for CCNCs and MBU:LTGs. The distribution of these 34 addresses on the CCNCs and MBU:LTGs according to the proportional condition leads to $A_{CCNC} = 5$ and $A_{LTG} = 3$.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>CCNC</th>
<th>MBU:LTG</th>
<th>MBU:SGC</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of devices</td>
<td></td>
<td>2</td>
<td>8 (7)</td>
</tr>
<tr>
<td>Device addresses/scan cycle</td>
<td>5</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Messages/call:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Input $h_{\text{device},I}$</td>
<td>13</td>
<td>31/8</td>
<td>-</td>
</tr>
<tr>
<td>Output $h_{\text{device},O}$</td>
<td>13</td>
<td>28/8</td>
<td>6/4</td>
</tr>
<tr>
<td>Total number of addresses/scan cycle</td>
<td></td>
<td>10 + 24 (21) + 4 + 3 *) = 41 (38)</td>
<td></td>
</tr>
</tbody>
</table>

(…) values if one MBU:LTG fails (worst case)

*) CCGO + CCGl + SYP

3B.6.4
3.1.2. Calculation of the IOP Saturation Throughput

From the balance equation
\[ \bar{F}_c = \bar{F}_{OH} + \sum_i c B_i \bar{F}_c h_i \]
where \( \bar{F}_c \) is the mean number of arrivals of device type \( i \)

where \( i \) stands for CCNC, MBU:LTG or MBU:SGC and
with \( \bar{F}_{OH} = A_{total} h_{OH} \) = total overhead per scan cycle

the mean scan cycle time \( \bar{F}_c \) is given by
\[ \bar{F}_c = \frac{\bar{F}_{OH}}{1 - c h} \]  

with \( h = \sum_i B_i h_i = h_{I0} (\beta_{CCNC} + N_{LTG} B_{LTG}) + h_{L0}, h_{SGC} B_{SGC} \)

and the assumptions (see Fig. 2)
\[ \beta_{CCNC} = \beta_{CCNC},I = \beta_{SGC},0 \] ; \[ \beta_{LTG} = \beta_{LTG},I = \beta_{LTG},O \] ; \[ \beta_{SGC},I = 0 \]

(Arrival rate and service time definitions see chapter 2.2 and 2.3).

The system is stable if the arrival rate at each device (\( c B_i \)) is less than the scan rate (\( \mu_i \)) of the IOP. Therefore the IOP throughput is limited by the device which saturates first.

With \( \mu_i = A_i / \bar{F}_c \) = mean scan rate of device type \( i \) and the stability condition \( \mu_i \geq c B_i \) the mean cycle time is limited by
\[ \bar{F}_c \leq \frac{A_i}{c B_i} \]  

From Eq. (1) and (2) the saturation call rate \( c_{sat},i \) of device type \( i \) is given by
\[ c_{sat},i \leq \frac{1}{h + \bar{F}_{OH} \frac{B_i}{A_i}} \]

The device of type \( i \) with the lowest saturation rate \( c_{sat},i \) defines the IOP saturation throughput. The maximum permissible throughput \( c_{max} \) (calls/s) with acceptable waiting times and queue lengths was found by simulations:
\[ c_{max} \approx 0.9 c_{sat},i \]

3.1.3. Performance Improvement by Scan Cycle Dynamization

Although static scan cycle optimization leads to a high performance increase the following subjects could not be covered:
1. The scan rates of the MBU:LTGs in normal operation (no MBU:LTG failure) and especially of the stand-by CCNC as well as of the CCG 0, CCG 1, and SYP are too high and therefore not adapted to the arrival rates.
2. No dynamic adjustment to short-time arrival rate variations caused by e.g. unbalanced load or load peaks.

On the other hand, to avoid bank-ups with unacceptable waiting times in case of CCNC change-over or MBU:LTG failures, the designed scan rate must be supplied in a time duration in the order of a scan cycle.
Fig. 4. NON-OPTIMIZED SCAN CYCLE (one ADDRESS per DEVICE)

Fig. 5. OPTIMIZED SCAN CYCLE (see CHAPTER 3.1.1)

Fig. 6. SCAN CYCLE DYNAMIZATION (see CHAPTER 3.1.3)

DEVICE TYPE

<table>
<thead>
<tr>
<th>CCNC,1</th>
<th>MBU:LTG,1</th>
<th>MBU:SGC,1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

MBU:LTG SITUATION

<table>
<thead>
<tr>
<th>.1 ALL MBU:LTG ACTIVE</th>
</tr>
</thead>
<tbody>
<tr>
<td>.2 ONE MBU:LTG FAILS</td>
</tr>
</tbody>
</table>

3.3B.6.6
As a solution a dynamization of the static scan cycle has been implemented: A "success bit" (S-Bit) which decides whether the address is scanned (S-Bit = 1) or not (S-Bit = 0) is attached to each scan address. The flow-chart in Fig. 3 shows that the S-Bit remains set to 1 as long as messages are transferred. It is set to 0 when no input and no output message is stored. In this case in the next cycle the address is skipped and the S-Bit is changed to 1.

Because the time to control the S-Bit is very short compared to the switch-over time, the mean cycle time decreases and therefore the IOP throughput increases. A performance improvement of 10% to 20% was found by simulations.

3.1.4. Simulation Studies and Results

Detailed simulation studies of the IOP were done using the simulation language GPSS. The arrival and service processes were modeled as in reality. For the example system configuration and the traffic parameters of Tab. 1 results are presented in:
- Fig. 4 for the non-optimized static scan cycle with one address per device
- Fig. 5 for the optimized static scan cycle described in chapter 3.1.1
- Fig. 6 in case of static scan cycle dynamization (see chapter 3.1.3).

The left column of Fig. 4 to Fig. 6 shows the saturation throughput of all device types \( c_{sat} \) calculated by Equation (3). The MBU:LTG situation is indicated by the index 1 or 2 (see legend). The results show that in all three cases the CCNC saturates first and limits the maximum IOP throughput. This is a sensible strategy because the initial message of an ISDN call arrives at the CCNC and can be rejected, e.g. by an overload strategy, before unwanted sub-calls are generated.

The maximum IOP throughput \( c_{max} = 0.9 c_{sat} \) is indicated in the middle column. The results show that the optimized static scan cycle leads to a throughput improvement of nearly 60% compared to the non-optimized case. A further increase of about 10% can be achieved through dynamization but the decisive advantage of the adaptation to short-time arrival rate variations has to be kept in mind.

By simulations with \( c_{max} \) the mean waiting times of all related queues or buffers are depicted in the right column. The waiting times in all three cases are small and in the same time range, especially in case of static scan cycle dynamization.

REFERENCES