ICCC 88, 30.10. – 4.11.88, Tel Aviv; Israel

Simulation of Application Layer Protocols for Factory Automation

Martin Bosch, Ottmar Gihr

University of Stuttgart
Institute of Communications Switching and Data Technics
Seidenstraße 36, D–7000 Stuttgart 1
Federal Republic of Germany

Two protocol architectures for factory automation are analysed by means of an event by event simulation technique. One is the standardized Manufacturing Automation Protocol (MAP) architecture and the other is a vendor specific protocol architecture. We show the modelling of these protocol architectures and a performance evaluation by simulation for various topologies, message types and scheduling strategies of the processor handling the upper three layers of these protocol architectures.

1. Introduction

Communication is an essential prerequisite for Computer Integrated Manufacturing (CIM) and especially for factory automation. Computer systems of different vendors can often only communicate with each other, if protocol conversions between the various systems are employed. In general, these protocol conversions must be done by gateways at the application layer with the disadvantage of high communication costs, reduced performance and reduced functionality of the interconnected networks. Therefore, a multi vendor project has been initiated by General Motors in 1980, to develop the standardized *Manufacturing Automation Protocol*. This worldwide project is accompanied by many other groups or multi vendor projects like the European MAP User Group (EMUG) since 1985 or the ESPRIT (European Strategic Program for Research and Development in Information Technology) project CNMA (Communications Network for Manufacturing Applications) since 1986.

MAP is based on the ISO (International Organization for Standardization) Reference Model for Open Systems Interconnection (OSI) [2]. From existing standardized protocols adequate options are chosen where they are practicable. Because of the special application in factory automation, there was a new protocol for the application layer to be specified and proposed for standardization.

Until MAP reaches the status of a stable standard, communicating devices have to use vendor specific protocol architectures. In this contribution we compare the performance of a vendor specific Automation Protocol (AP) architecture [13, 14] with the performance of the MAP architecture. We consider especially layers 5 to 7. Both architectures are depicted in Figure 1. Implementation details for MAP are taken from the CNMA project [1] and are, especially at the Media Access Control (MAC) sublayer, not in conformance with the MAP standard. However, this does not influence our simulation for the higher layers too much.

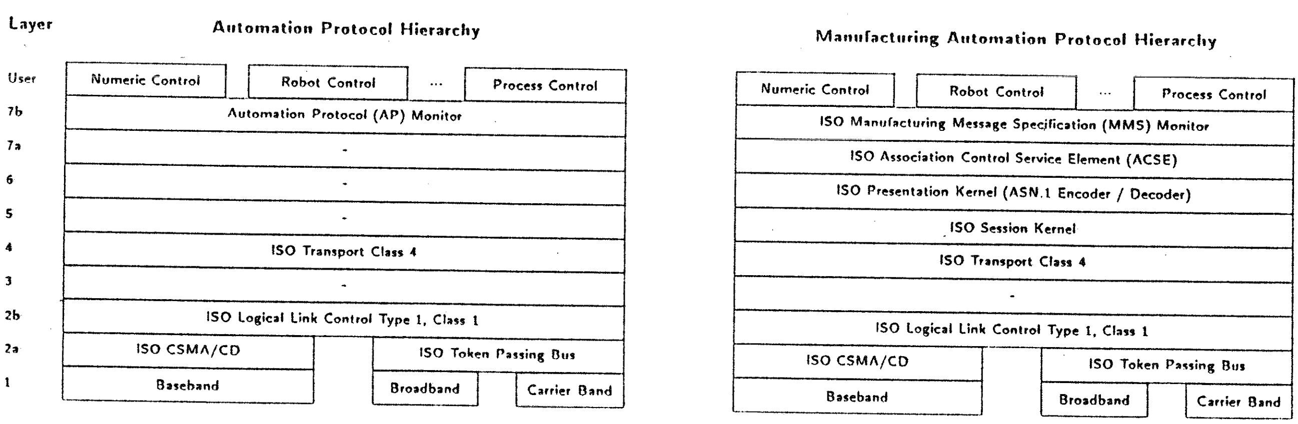

Figure 1 : Protocol Architectures

At layer 1 (physical layer) and sublayer 2a (MAC) the *CSMA/CD Access Method* (Carrier Sense Multiple Access with Collision Detection) [8] is used. In the MAP specification [12] the use of the *Token Passing Bus Access Method* [8] is recommended. The Logical Link Control (LLC) sublayer 2b [7] is using the connectionless, unconfirmed datagram service. Up to now, the implementation of layer 3 (network layer) is not necessary due to the limitation of communication to one *Local Area Network* (LAN). The connection oriented transport protocol class 4 [3] is used at layer 4 (transport layer).

At the moment, layers 5 (session layer) [4] and 6 (presentation layer) [9] of the MAP architecture contain only kernel functions. Especially the encoder and decoder of the used syntax at layer 7, described in *Abstract Syntax Notation One* (ASN.1) [10], are located at layer 6. The *Association Control Service Element* (ACSE) [6] is the basic part of many standardized *Application Service Elements* (ASEs) at layer 7 (application layer) as for example *File Transfer, Access and Management* (FTAM) [5]. Additionally, the ASE *Manufacturing Message Specification* (MMS) [11] has been prepared for factory automation by the Electronic Industries Association (EIA) and has now reached the status of a ISO Draft International Standard (DIS) 9506.

In the AP architecture the AP monitor [13, 14] realizes all necessary functions of layers 5, 6 and especially 7.

2. Modelling

For the AP and MAP architectures of Figure 1 the simulation models with a detailed modelling of the layers 5 to 7, are depicted in Figures 2 and 3, respectively. The layers below the *Transport Service Access Point* (TSAP) are modelled by infinite servers, representing the transport subsystem by delay phases. The corresponding parameters for these phases are obtained by separate simulation runs, by analytical studies or by performance measures of real systems.

In the models, the communication relationship of one transport connection is depicted. The models show on the left side the active station (client or requestor) and on the right side the passive station (server or responder). For each user only those parts are depicted which are necessary for this configuration. Combining the two parts of a figure in mind gives the model of a complete station. The upper parts of the processor in the models refer to the AP monitor in Figure 2 and to the MMS monitor in Figure 3. The user is placed on top of the application layer. In Figure 2 each generator G_i refers to one logical channel and represents one or many users. In Figure 3 only one context is shown for which unconfirmed and confirmed services are generated and flow controlled separately.

Flow controls are depicted as a rhombus. The horizontal queues contain the credits. For passing the rhombus each message picks up a credit.

We consider the following flow control types:

1) channel individual flow control,
2) flow control for segmented messages and
3) flow control due to limited common resources.

In both models all messages enter the monitor at a distribution phase V. The message type is recognized and the message is routed to a queue according to its type.

In Figure 2 there are two different actions to be done after reception of the transport confirmation. In the case of an unconfirmed service a local confirmation has to be returned to the user in order to release occupied resources. In the case of a confirmed service a timer has to be started controlling the arrival of the confirmation (Phase T_1). In Figure 2 also segmented messages are allowed. A special case is a reaction to a message after a waiting time W, which again is modelled as an infinite server.

In Figure 3 no timer is necessary because of the reliably assumed transport connection which creates at least a negative confirmation, even in the case of a fault. Flow control E blocks the messages in Figure 3 in front of the transport system, if there is no credit available, in contrast to Figure 2.

As an *Enhanced Performance Architecture* (EPA), additionally to the MAP architecture *mini-MAP* has been defined to increase performance. In mini-MAP layers 3 to 7a are empty and therefore replaced by an interface. The MMS monitor together with its interface is then similar to the AP monitor due to no layering of functionality being used at the higher layers. Figure 3 could also be used as a model for mini-MAP. The phases of layers 5 to 7a must constantly be set to zero. The phases of the transport system and of the MMS monitor, which then includes the interface, must be modified according to estimated or measured values at a real implementation.

3. Simulation Technique

For our purposes the event by event simulation method is used. The system state is represented by a set of variables and the changes of the system state, called events, are seen to happen immediately, consuming no system time. Therefore, the simulation program processes an event entirely updating all affected system variables and planning possible later events. After processing an event, the simulation program looks for the next event in the order of time and after updating the system time it processes that event. The sequence of events is organized by a calendar in which all next events are stored.

Simulating the system in this way, measures for server utilizations, queue lengths as well as for means and coefficients of variation for transfer times, can be taken. The replication method is used to achieve confidence intervals for the measures, according to the Student-t-theory, subdividing the simulation run into 5 to 10 part tests.

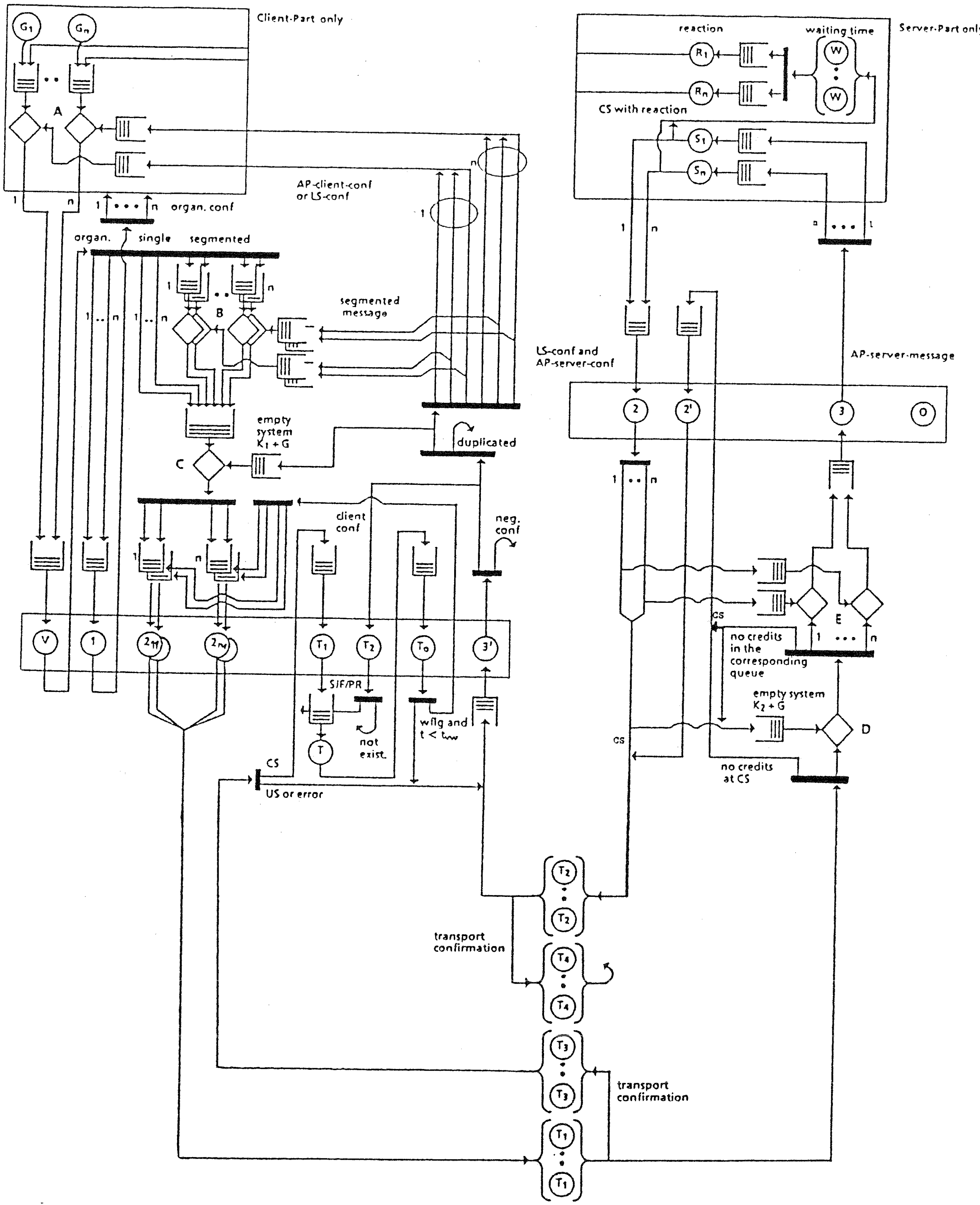

Figure 2 : Model for the Higher Layers of the AP Architecture

	Symbol	Legend	AP Param.	MMS Param.
Explanations	1 .. n	n Logical Multiplex Channels	1 – 3	1 – 3
		Decision		
		Duplication		
		Merging		
Flow Controls	A	Individual for the Channel (Client)		
		or for the Context (Requestor)		
		Amount of Credits (for CS)	10	10
	B	Segmented Message (Window)		
		Window Size	3	
	C	Limitation of Common Resources		
		(Client or Reqestor)		
		K_1 Credits only for C	10	10
		G Credits for C and D	30	30
	D	Limitation of Resources		
		(Server or Responder)		
		K_2 Credits only for D	10	10
		G Credits for C and D (see C)	30	30
	E	Individuell for the Channel (Server)		
		or for the Context (Responder)		
		Amount of Credits	10	10
Abbreviations	ACSE	Association Control Service Element		
	AP	Automation Protocol		
	conf	Confirmation		
	CS	Confirmed Service		
	LS	Local Interface		
	SJF/PR	Special Service Discipline		
	T	Hardware Timer		
	t	Time since the Message Generation		
	tvw	Administration Time		
	US	Unconfirmed Service		
	wflg	Retransmission Flag		
Monitor Phases	O	Overhead Phase	1 ms	1 ms
	1	For Organisatory Messages or M-Open	3 ms	1 ms
	2	For Positive Confirmations from the User	2.5 ms	3 ms
	$2'$	For Negative Confirmations from the User	2.5 ms	3 ms
	2_i	For Messages from the User	5 ms	3.5 ms
	3	For Messages to the User	7.5 ms	3 ms
	$3'$	For Confirmations to the User	4 ms	3 ms
	$3''$	For Transport Confirmations to the User		0.5 ms
	I	For M-Await Initiate		5.5 ms
	T_1	To Start a Timer	1.5 ms	
	T_2	To Stop a Timer	3.5 ms	
	T_0	For Timeouts of a Timer	1.5 ms	
	V	Distribution Phase	1.5 ms	1.5 ms
Other Phases	O	Overhead Phases		1 ms
	A_i	ACSE Phases		1.5 ms
	P_i	Presentation Phases		5 ms
	R	Responder Phase		5 ms
	R_i	Reaction Phases	5 ms	
	S_i	Server Phases or Session Phases	5 ms	3.5 ms
	T_1	Transport Delay (Message)	10.5 ms	10.5 ms
	T_2	Transport Delay (Confirmation)	7.5 ms	7.5 ms
	T_3	Transport Delay (T-Confirmation)	7.5 ms	7.5 ms
	W	Waiting Time until Reaction	500 ms	

Table 1 : Legend to Figures 2 and 3 and Parameters for the Simulation

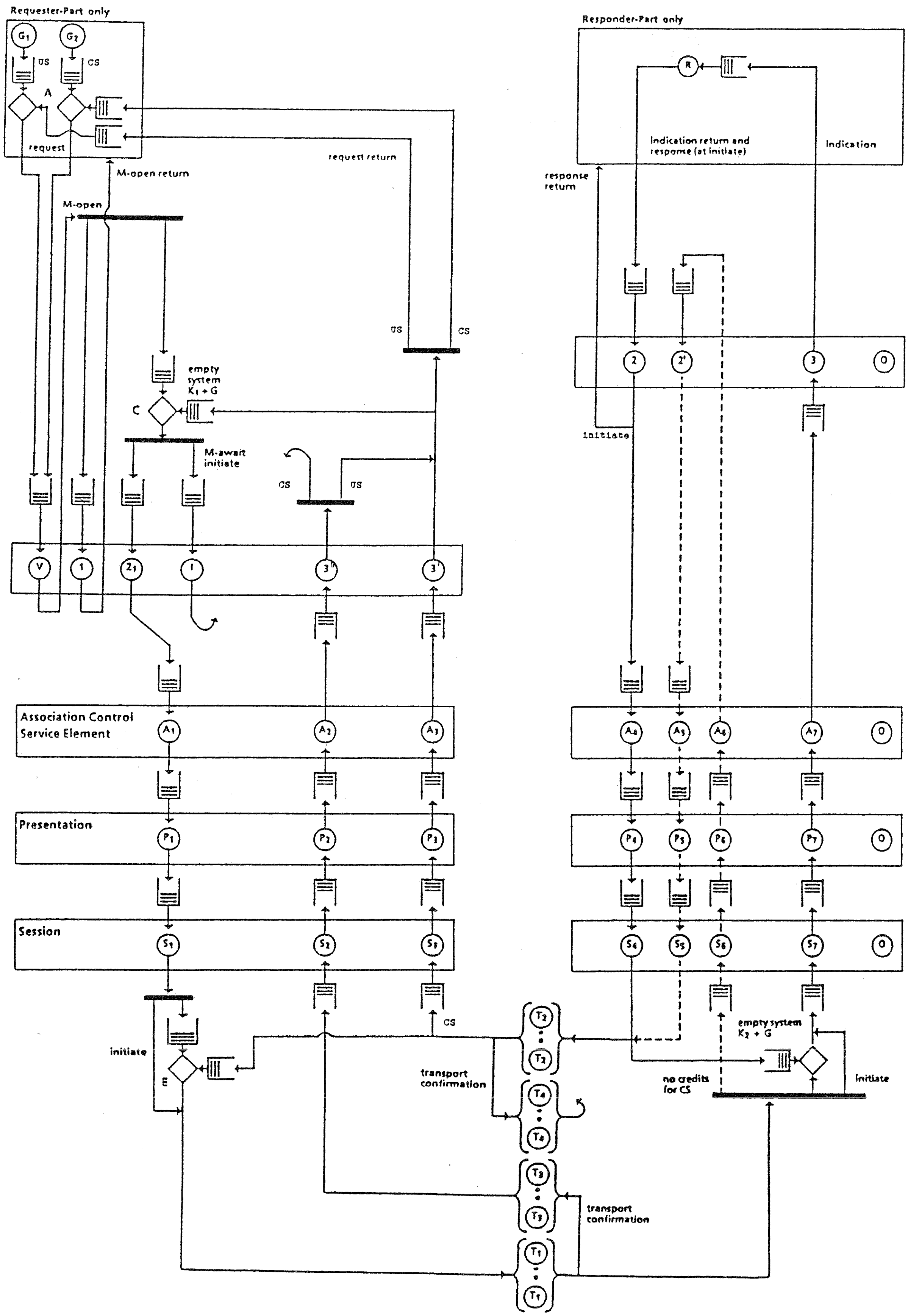

Figure 3 : Model for the Higher Layers of the MAP Architecture

Due to the complexity of models containing several layers, such models cannot be simulated as a whole. Only one to three layers should be simulated simultaneously. Therefore, the lower layers are aggregated to a delay equivalent service center (infinite server). The lower moments of the distribution function of the infinite server are obtained by another simulation of the lower layers, by analytical studies or by measurement of real systems. Usually, the first or the lower two moments are used for an approximation. Therefore the influence of the lower layers is represented, but need not be simulated explicitly at the simulation run for the higher layers.

To do the lower layer performance evaluation, the offered traffic to the considered layer must be known. Therefore, a data flow analysis or data flow estimation is needed before. The performance evaluation of the considered layer results in an aggregation of the submodel to a delay. During simulation of the next higher model, the offered traffic to the submodel is measured and compared with the traffic assumed for the submodel aggregation. If the assumed traffic is out of tolerance of the meassured traffic the submodel performance evaluation is repeated as depicted in Figure 4.

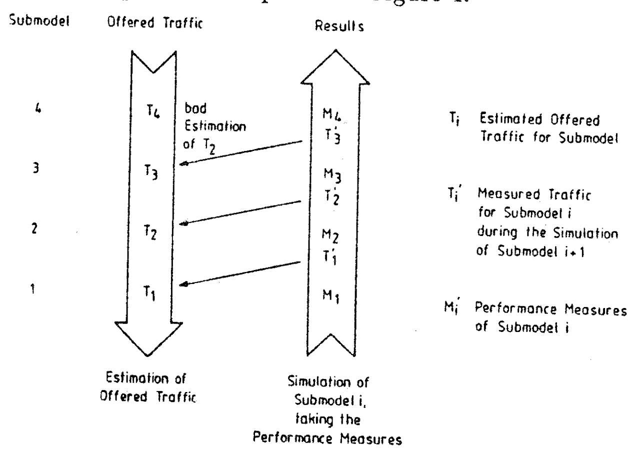

Figure 4 : Iterative Simulation of Submodels

For our purposes, the usage of general simulation tools with userfriendly interfaces is very critical, due to the restricted functionality of the contained model components there and the resulting relatively long simulation time. Therefore, the implementation of simulation programs is done by using a modular simulation library developped at the *Institute of Communications Switching and Data Technics* at the *University of Stuttgart*. The library is a compromise between simulation tools and individually written simulation programs. A pattern of the dynamic data structure and procedures or functions for frequently used tasks in a simulation program as queue handling, server handling, random number generation or statistic evaluation are included in that library. Carefully implementing the simulation programs and looking for the locality of data guarantees an optimal program for the considered problem with relatively short simulation times and acceptable program development times.

4. Results

The parameters of the simulations can be taken from Table 1. The generators are producing traffic according to a Markovian distribution of the interarrival time. The curves are depicted in the stable range of the arrival rate and the simulation points are marked together with their 95% confidence intervals where possible. In Figures 5 – 11 curves can be found for the mean of the transfer time from the generation of a message to the arrival at the receiving user and of the buffer occupation time at the sending station from the generation of a message to the arrival of the confirmation at the sending user.

4.1 Comparison of Various Topologies

During our studies of topologies with one sending and one receiving station, it could always be observed, that the bottleneck is within the sending part of the sending station. Results for confirmed services for one queue per phase can be taken from the solid lines of Figures 5 and 6. The AP architecture has always lower transfer and buffer occupation times than the MAP architecture and higher arrival rates are possible.

In order to shift the bottleneck to the receiving station, the influence of two other stations (sending to the same receiving station with the same offered load as the first station) on the first channel of the receiving station is depicted in Figures 5 and 6 with dotted lines. The border of stability is reached at half the load of the before mentioned configuration and especially the buffer occupation time increases rapidly.

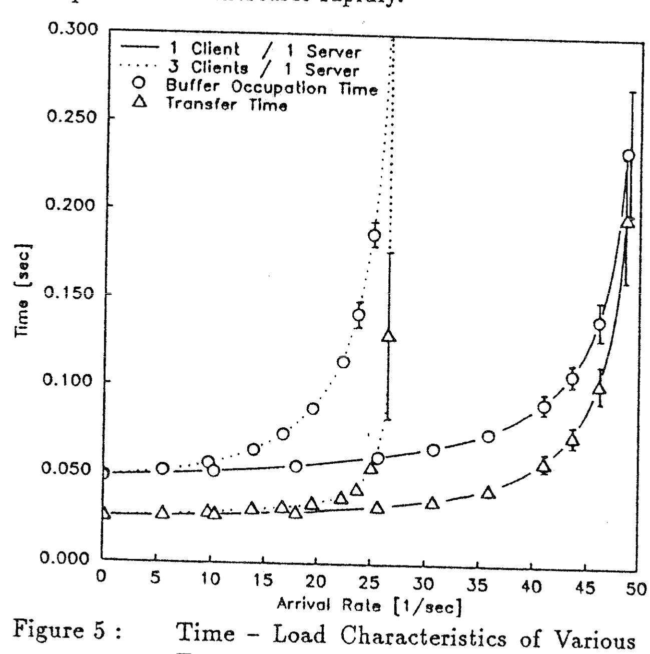

Figure 5 : Time – Load Characteristics of Various Topologies for the AP Architecture for Confirmed Services

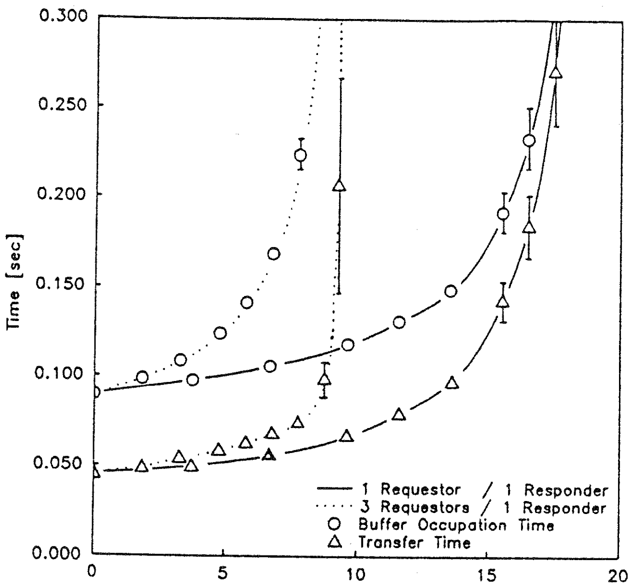

Figure 6 : Time – Load Characteristics of Various Topologies for the MAP Architecture for Confirmed Services

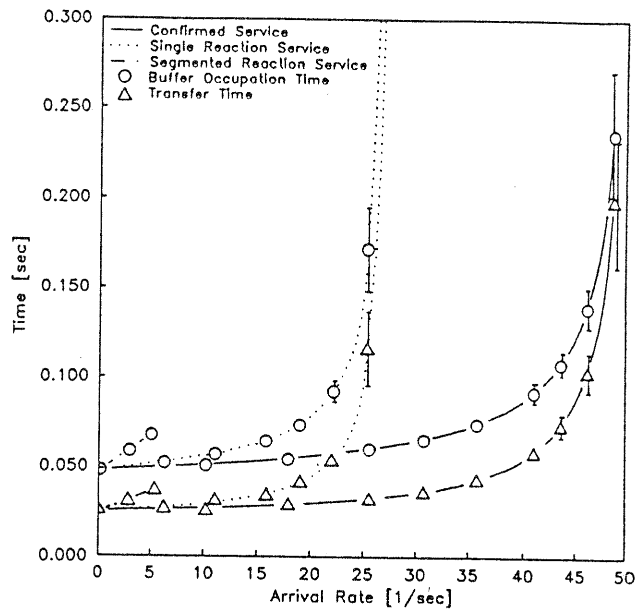

Figure 7 : Time – Load Characteristics of Various Message Types for the AP Architecture

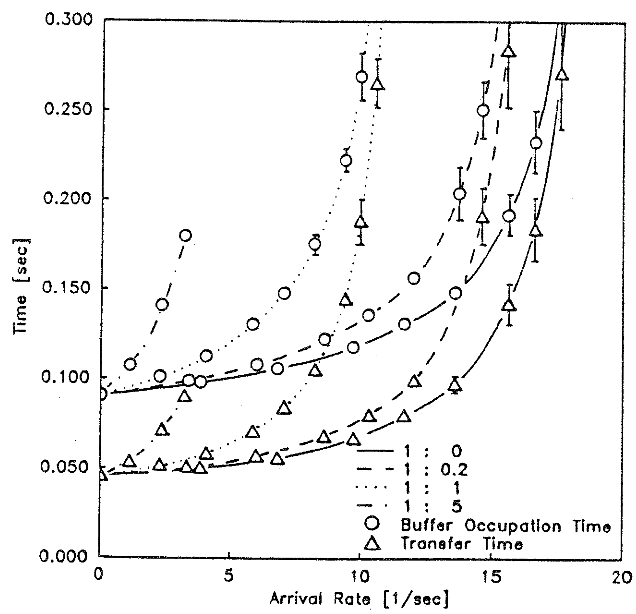

Figure 8 : Corresponding MAP Results (Confirmed Services, Full Duplex) to the Time – Load Characteristics of Various Message Types for the AP Architecture

4.2 Comparison of Various Message Types

For AP various message types can be distinguished: organisatory, unconfirmed, confirmed and segmented messages as well as messages with single reaction or with segmented reaction. From these types confirmed services and messages with single reaction as well as messages with segmented reaction with five segments per message have been choosen to be compared in Figure 7 for a topology with one sending and one receiving station and for one queue per phase. Considered is the original message over its own arrival rate. The curves can be distinguished according to the traffic in the reverse direction, which is zero for confirmed service, equal to the original traffic for single reaction service and five times of the original traffic for segmented reaction service. The load from the reverse direction of the single reaction service has nearly the same effect as the load from the two additional sending stations in Figure 5. The boarder of stability is reached at half of the arrival rate for confirmed services. If the reaction is segmented into five segments, the system becomes unstable for the reverse direction even at an arrival rate of five original messages per second.

Figure 8 depicts the corresponding MAP results to Figure 7, where the traffic in reverse direction is produced by separate generators, due to no existing equivalent message type to the one above. The observed effects are similar to Figure 7. Additionally, the dashed lines show the result for having the higher traffic in the observed direction.

4.3 Comparison of Various Scheduling Strategies

Various scheduling strategies for the phases of the processor for layers 5 to 7 are taken into consideration in this section for a topology with one sending and one receiving station. It is possible to simulate the models as depicted in Figures 2 and 3 with one queue per phase but we are also able to simulate a configuration in which only one queue (mailbox) exists to enter a layer, corresponding to real implementations. We compare the strategies one queue per phase, one mailbox with low internal priority and one mailbox with high internal priority to each other.

In the case of one queue per phase, the priorities of the phases are chosen to handle at first timer events and to prefer the receiving direction against the sending direction. The phase V has lowest priority to keep the traffic out of the system, if it is heavily loaded. A separate study has shown, that prefering the sending direction against the receiving direction in the receiving station (as it is useful for tandem systems) yields nearly the same results and is therefore no longer considered separately. An overhead phase O follows each other phase considering the influence of the operating system (for example task switch). Phase $3'$ is the only exception which is followed immediately by phase T_2. We consider timeout values for the retransmission of a message and for beeing ready to receive a confirmation to be very large, in order to avoid retransmissions. After an administration timeout the message would have to be removed and a negative confirmation would have to be sent to the user. In Figure 3 we take into account, that layers 5 to 7 are running on a single real processor. The MMS monitor has the lowest priority and priorities increase in direction to the transport system.

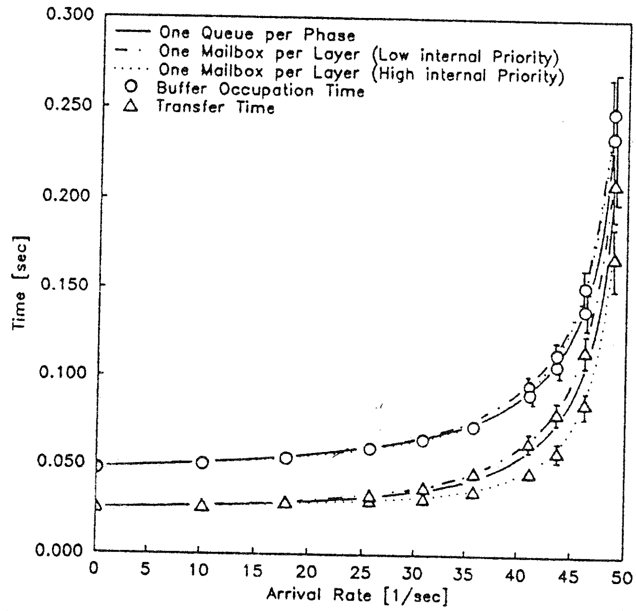

Figure 9 : Time – Load Characteristics of Various Scheduling Strategies of Processor Phases for the AP Architecture for Confirmed Services

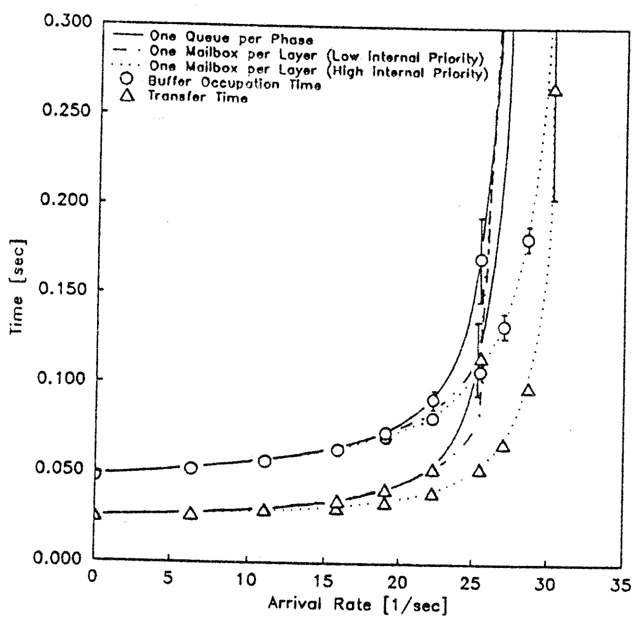

Figure 10 : Time – Load Characteristics of Various Scheduling Strategies of Processor Phases for the AP Architecture for Messages with Single Reaction

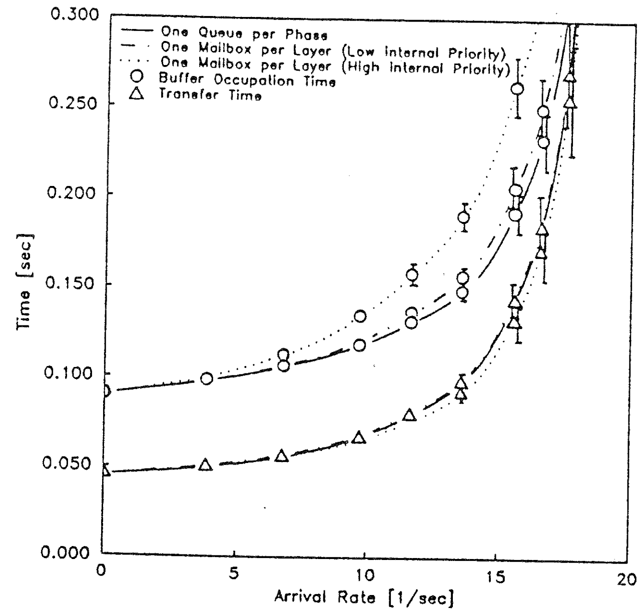

Figure 11 : Time – Load Characteristics of Various Scheduling Strategies of Processor Phases for the MAP Architecture for Confirmed Services

Figure 9 shows this comparison for confirmed services and Figure 10 for messages with single reaction for the AP architecture. In Figure 11 the corresponding curves for confirmed services for the MAP architecture are depicted. The transfer time is always an optimum, if the scheduling strategy with one mailbox is chosen, prefering internal phases, so that traffic being in the system is handled with priority and internal blocking in front of phase 2_{ij} can be avoided. This results however in a long buffer occupation time for the MAP architecture, since high internal priorities of the sending direction are prefered against the returning confirmation in each mailbox at each layer. Figure 10 stresses, that for the AP architecture this strategy is an optimum, due to the full duplex connection resulting in internal phases for both directions. One queue per phase results in optimal buffer occupation times for confirmed services and behaves very similar to one mailbox per layer with low internal priority having relatively long waiting times in front of phase 2_{ij}.

5. Conclusion

We have compared two application layer protocols for factory automation. The more complex standardized MAP architecture is less efficient. Buffer occupation time and transfer time are higher than for the alternative AP architecture and therefore the specification of an Enhanced Performance Architecture is once again recommended. Bounds for the traffic and mean transfer and buffer occupation times for various topologies and message types have been obtained. Comparing different scheduling strategies, it could be seen, that one mailbox per layer with high internal priorities yields always optimal transfer times. Optimal buffer occupation times are obtained for one queue per phase.

Acknowledgement

The authors would like to thank Hans-Hubert Hemberger and Waldemar Hersacher for the implementation of the simulation programs and for obtaining the simulation results.

References

[1] CNMA (ESPRIT Proj. 955); "MMS Interface Specification", Version 5.0, Siemens AG, March 18, 1988

[2] ISO 7498; "Information Processing Systems – Open Systems Interconnection – Basic Reference Model", November 1983

[3] ISO 8072; "Transport Service Definition", May 1984
ISO 8073; "Transport Protocol Specification", May 1984

[4] ISO 8326; "Basic Connection Oriented Session Service Definition", September 1984
ISO 8327; "Basic Connection Oriented Session Protocol Specification", September 1984

[5] ISO DP 8571; "File Transfer, Access and Management", April 1985

[6] ISO DIS 8649; "Service Definition for Common Application Service Elements – Part 2 : Association Control", April 1986
ISO DIS 8650; "Protocol Specification for Common Application Service Elements – Part 2 : Association Control", April 1986

[7] ISO DIS 8802/2; "Local Area Networks – Logical Link Control", 1985

[8] ISO DIS 8802/3; "Local Area Networks – CSMA/CD Access Method", 1985
ISO DIS 8802/4; "Local Area Networks – Token Passing Bus Access Method", 1985

[9] ISO DP 8822; "Connection Oriented Presentation Service Definition", May 1986
ISO DP 8823; "Connection Oriented Presentation Protocol Specification", May 1986

[10] ISO DIS 8824; "Specification of Abstract Syntax Notation One (ASN.1)", 2nd DIS, September 1986
ISO DIS 8825; "Specification of Basic Encoding Rules for Abstract Syntax Notation One (ASN.1)", 2nd DIS, September 1986

[11] ISO DIS 9506; "Manufacturing Message Specification, Part 1: Service Definition", December 1987
ISO DIS 9506; "Manufacturing Message Specification, Part 2: Protocol Specification", December 1987

[12] MAP; "Manufacturing Automation Protocol", Version 3.0, General Motors, Warren/Michigan, April 7, 1987

[13] SINEC; "Automation Protocol (AP 1.0), Teil 1: Protokollspezifikation", Siemens AG, 30. Juni 1987
SINEC; "Automation Protocol (AP 1.0), Teil 2: Technologische Funktionen", Siemens AG, Dezember 1987
SINEC; "Automation Protocol (AP 1.0), Teil 3: AP–Protokollabwickler", Siemens AG, Januar 1988

[14] SINEC; "SINEC Communication Processor (SCP) – Software", Siemens AG, 19. Dezember 1986

Martin Bosch

was born in Leonberg, Federal Republic of Germany, in 1961. The Dipl.-Ing. degree in Electrical Engineering he received from the University of Stuttgart, Federal Republic of Germany. Since 1986 he has been working at the Institute of Communications Switching and Data Technics at the University of Stuttgart as a member of the scientific staff. His interests include the interconnection of heterogeneous communication networks and the performance evaluation of protocol mechanisms and architectures, especially for factory automation.

Ottmar Gihr

was born in 1957 in Mengen (Federal Republic of Germany). He studied at the University of Stuttgart where he got the Dipl.-Ing. degree in Electrical Engineering. Since 1984 he is with the Institute of Communications Switching and Data Technics as a member of the scientific staff. Currently he is supervisor of a research group on protocol analysis. His main interest is the performance evaluation of multi layered protocol architectures.