Deterministic Real-Time Communication with Switched Ethernet

Jürgen Jasperneite [*], Peter Neumann [†], Michael Theis and Kym Watson [‡]

Abstract

In this paper we show, under which conditions switched ethernet can be called deterministic.The influence of the service strategies in switched ethernet is investigated by simulation and compared with analytical bounds on the delays.

1 Introduction

Ethernet is an internationally standardized local area network technology [IEE00]. Although characterized as having a behavior unsuitable for control level network applications under moderate to heavy traffic loads, much of ethernet's non-deterministic reputation derives from the stochastic nature of its collision recovery mechanism. In today's factory communication systems ethernet is used for non-time-sensitive applications. A major step toward deterministic behavior in Ethernet networks is to eliminate the random CSMA/CD bus arbitration. This can be achieved by using the latest Ethernet switch technology instead of a hub-based infrastructure and by directly connecting systems to full duplex ethernet switches. Switch technology divides collision domains into simple point-to-point connections between network components and stations. Collisions no longer occur and the random backoff algorithm is no longer required. The concept of switching or Media Access Control (MAC) bridging, which was introduced in standard IEEE 802.1 in 1993, was expanded upon in 1998 by the definition of additional capabilities in bridged LANs. The aim was to provide additional traffic capabilities so as to support the transmission of time-critical information in a LAN environment [IEE98a], [IEE98b].

By using simulation, [JN01] showed the quantitative influence of system parameters like scheduling strategy and thinking time within the devices on the typical time behavior in a switched Ethernet. Mean values are not sufficient to assess the characteristics of distributed real-time systems. This paper therefore evaluates the real-time characteristics by looking at distributions and upper bounds on transaction times. Simulation techniques and the analytical method Network Calculus (originating from the engineering of Internet networks) are both applied. The results yielded by these methods are discussed and compared.

[*]Jürgen Jasperneite is with Phoenix Contact GmbH, System Development Department, Flachsmarktstr. 8-28, 32823 Blomberg, Germany, Fax:+49-5235-3-39999, e-mail: jjasperneite@phoenixcontact.com

[†]Peter Neumann is with IFAK Institute for Automation and Communication Magdeburg, Steinfeldstr. 3, 39179 Barleben, Germany, Fax:+49-39203-81100, e-mail: neu@ifak.fhg.de

[‡]Michael Theis and Kym Watson are with Fraunhofer Institute IITB, Fraunhoferstr. 1, 76131 Karlsruhe, Germany, Fax:+49-721-6091-413, e-mail: {the, wat}@iitb.fhg.de

2 Case Study

In order to meet the requirements for small, medium and large configurations the station number N is taken from the set $\{10, 50, 100\}$ for all scenarios. The simulation model was realized with OPNET Modeler 7.0B from OPNET Technologies, Inc., by extending the standard models for switches and MAC to incorporate priority queuing. In the simulation experiments only full duplex transmission channels (FDX) with a transmission capacity of $C = 100Mbps$ are used to guarantee a collision-free system. In the switches and devices the scheduling strategies FCFS (First Come, First Served) and PQ (Priority Queuing) are considered. Packet processing is asumed to be store and forward (i.e. a packet must be completely received before the transmission to subsequent switch or station may begin.) In addition, we assume that packet processing times in a switch or station are negligible.

2.1 Topology

Switches enable free network topology configuration. Of the numerous types of topologies only the star and line topology are considered in the scenarios here. The simplest topology is the star topology because it only has one switch. However, the star topology requires maximum cabling which is of concern to automation technology. Originally, fieldbus systems were developed in order to replace parallel cabling of sensors and actuators. For this reason we look at a line topology in which every switch is assigned to one device. This creates a "fieldbus-compliant" cabling concept. In future devices this switch function may be part of the device itself. This will reduce costs for cabling and infrastructure. Due to the large number of switches the line topology is the most unfavorable topology for real-time behavior.

2.2 Load Parameters

In order to create a data traffic mix the following four message types (MT_i), which can be found in every indus-

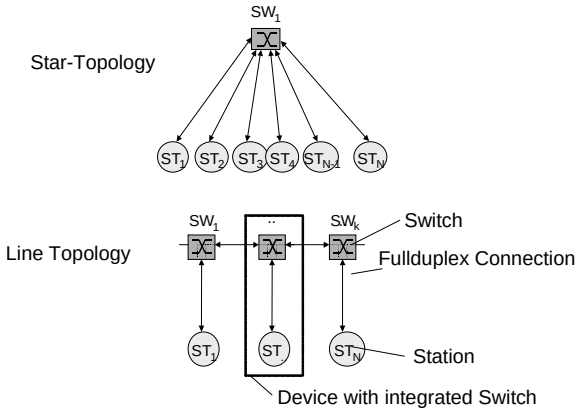

Figure 1. Considered topologies

trial application, are defined from the application viewpoint. There are cyclic as well as acyclic event types.

When evaluating the performance of a network topology we also have to consider the information flow distribution defined by the application. The information flow distribution describes every communication relationship between the devices within a message type. As we can see in actual configurations the central information flow distribution (C) is very important in the field of automation technology. This is the case if there is one special device in the system which is involved in all communication relationships, either as source or sink. This form applies to the application (part) near the process in which a central control system is used. In the simulation experiments we assume a uniform distribution of the destination addresses.

2.3 Scenarios

In these synthetic scenarios a typical automation application is used which consists of a central control system and distributed intelligent field devices. Within the MT_1 message type the cyclic process data exchange forms the major part of the data with its short user data lengths typical of field communication. In addition, events, such as for example alarms from the field devices to the central control system, are transmitted within the MT_2 message type. The scenarios also include acyclic parameterization and network management (MT_3) as well as transfer of large data amounts (MT_4) from the distributed field devices to the central control system.

Type	Description
MT_1	Cyclic send and request of process data objects with response
MT_2	Transmission of events, e.g., alarms
MT_3	Acyclic transmission for, e.g. network monitoring, diagnostics or device configuration
MT_4	Acyclic transmission of datagrams for file transfer or IP-based applications

Table 1. Definition of message types

	Message Type			
Feature	**MT1**	**MT2**	**MT3**	**MT4**
Transaction type	Confirmed Client: $\mathcal{ST}_1$ Server: $\mathcal{ST}_2..\mathcal{ST}_N$	Unconfirmed Publisher: $\mathcal{ST}_2..\mathcal{ST}_N$ Subscriber: $\mathcal{ST}_1$	Confirmed Client: $\mathcal{ST}_1$ Server: $\mathcal{ST}_2..\mathcal{ST}_N$	Confirmed Client: $\mathcal{ST}_1$ Server: $\mathcal{ST}_2..\mathcal{ST}_N$
Distribution information flow	Central $(\mathcal{ST}_1)$	Central $(\mathcal{ST}_1)$	Central $(\mathcal{ST}_1)$	Central $(\mathcal{ST}_1)$
Percentage of packets u_{Mt_i}	80%	10%	9.5%	0.5%
Distribution packet generation	D	R	R	R
Distribution destination addresses	D (cyclic)	D	R	R
Length MAC payload [Bit]	$L^{req}_{MT_1}$: 368 $L^{res}_{MT_1}$: 368	$L^{req}_{MT_2}$: 368	$L^{req}_{MT_3}$: 1024 $L^{res}_{MT_3}$: 1024	$L^{req}_{MT_4}$: 1024 $L^{res}_{MT_4}$: 12000
Distribution Length MAC payload	D	D	D	D
user priority	Standard (5)	Medium (6)	High (7)	Low (4)

Table 2. Defining the workload

The distributions used for the packet interarrival times and the packet destination addresses are either deterministic (D) or random (R) (cf. Table 2). The metrics are shown as a function of the system load. The system load ρ is defined as the ratio of arrival rate to service rate of a service system, whereby the service rate is the link capacity C. Since a switched Ethernet comprises numerous segments with individual dedicated service systems we want to focus on the segment with the highest load in the sense of a bottleneck analysis.

Due to the central information flow distribution of the selected workload, the physical link between station $\mathcal{ST}_1$ and the port of the corresponding switch has the highest load. The service system at the port of the switch to which the central station is connected has the highest load due to the asymmetrical workload. Therefore, it is considered as reference for the system load from now on. The relative system load is within the range $0 < \rho < 1$.

As stated in [Jas02] the generation of transactions of a given type on a given connection is constrained by a leaky bucket regulator (refer to the definition below). In the scenarios we assume that with increasing arrival rate of the sub-application λ_{MT_i} the burstiness also increases. Besides the flow rate r_{MT_i}, the bucket size b_{MT_i} of the leaky bucket arrival curves is increased with increasing load ρ. Based on the analysis of communication requirements [JN00] the bucket size is chosen between $b_{min} = 2$ packets for $\rho = 0.05$ and $b_{max} = 12$ packets for $\rho = 1.2$.

We first describe the analytical approach using Network Calculus and subsequently consider the results using simulation.

3 Network Calculus

3.1 The principles of Network Calculus

Network Calculus is a widely applicable technique to assess the real-time performance of communication networks and is based on the fundamental work of [Cru99a], [Cru99b] and [GP93], [GP94]. For a full introduction the reader is recommended to consult [BT01]. Network Calculus basically considers networks of service nodes and packet flows between the nodes. Whereas traditional queuing theory deals with stochastic processes and probability distributions, Network Calculus involves bounding constraints on packet arrival and service. These constraints allow bounds on the packet delays and work backlogs to be derived, which can be immediately used to quantify the real-time network behavior. Traditional queuing theory, on the other hand, normally yields mean values and perhaps quantiles of distributions. The derivation is often difficult and upper bounds on end-to-end delays may not exist or be computable. The packet arrival process in Network Calculus is described with the aid of so-called *arrival curves* which quantify constraints on the number of packets or the number of bits of a packet flow in a time interval at a service node. Typical packet flows of interest are, for example, the packets of a given message type entering or departing from a service node. Let F be a packet flow and let $R(t)$ be the number of packets of F arriving in the time interval $[0, t]$. We say the flow is constrained by or has arrival curve $\alpha(t)$ if, for all $0 \leq s \leq t$

$$R(t) - R(s) \leq \alpha(t - s)$$

whereby $\alpha(t)$ is a non-negative, non-decreasing function. Although one can consider packet arrival constraints, it is mostly more convenient in the context of this paper to consider the corresponding work arrival curve. We say that a packet P requires $W[sec]$ work in the service node N, if N needs this long to process P. Typically, W can be calculated from the packet length (including overhead), the transmission rate and interframe gap on the medium. Let $R_w(t)$ be the amount of work due to flow F which arrives $[0, t]$. Then F is constrained by a work arrival curve $\alpha_w(t)$ if, for all $0 \leq s \leq t$

$$R_w(t) - R_w(s) \leq \alpha_w(t - s)$$

A service curve describes the service of flow F in node N. Let $R'_w(t)$ be the amount of work completed in the time interval $[0, t]$. Then the non-negative, non-decreasing function $\beta(t)$ is called a (minimal) work service curve for F in N if, for all $t \geq 0$,

$$\begin{aligned} R'_w(t) &\geq R_w \otimes \beta(t) \\ &= \inf_s \{R_w(s) + \beta(t - s) : 0 \leq s \leq t\} \end{aligned}$$

where $\otimes$ denotes the convolution operator. The work output flow is itself constrained by the arrival curve

$$\alpha_w \oslash \beta(t) = \sup_s \{\alpha_w(t + s) - \beta(s) : s \geq 0\}$$

where $\oslash$ denotes the deconvolution operator. In a packet network it is important to consider the packetization of the output. If packets flow from $\mathcal{N}_1$ to $\mathcal{N}_2$ then work arrives at $\mathcal{N}_2$ for further service only when the complete packet has been received. If $\beta(t)$ is a work service curve of a packet flow and δ_{max} is the maximum service time of packets of the flow, then $\beta_1(t) = \beta([t - \delta_{max}]^+)$ is a work service curve in $\mathcal{N}_1$ for the flow including packetization of the output.

The constraints given by an arrival and service curve for a flow suffice to calculate an upper bound on the delay of a packet in the service node and on the work backlog. The packet delay is bounded by the horizontal distance between α_w and β, whereas the work backlog is bounded by the vertical distance:

$$delay \leq \sup_{t \geq 0} \{\inf\{s \geq 0 \; such \; that \; \alpha(t) \leq \beta(t + s)\}\} \quad (1)$$

$$work \; backlog \leq \sup_{t \geq 0} \{\alpha(t) \leq \beta(t)\} \quad (2)$$

Computations with arrival and service curves can be greatly facilitated if concave arrival and convex service curves can be applied. The most important representatives are the leaky bucket arrival curve $\alpha(t) = rt + b$ with rate r and burst or bucket size b and the rate latency service curve $\beta(t) = \max(0, (t - T)R)$ with latency T and rate R. Both are determined by just two parameters. The above delay bound in this case is simply $T + b/R$ (if $r \leq R$), which manifests the dependence on the service latency and the burst size. The output flow is constrained by the leaky bucket arrival curve $\alpha^*(t) = \alpha(t + T)$.

3.2 Results for FCFS and PQ service nodes

We give two specific results on the Network Calculus for service nodes serving two or more flows according to a FCFS or PQ service strategy [refer to [BT01], Corollary 6.2.1 and Corollary 6.4.1, and [Wat02] for these and more general results].

Proposition 3.1 *[FCFS Split] Consider a service node $\mathcal{N}$ serving two packet flows F_1 and F_2 in FCFS order. Assume that $\mathcal{N}$ gives the aggregate flow a rate latency work service curve $\beta(t) = \max(0, (t - T)R)$ and suppose that F_2 is constrained by the leaky bucket work arrival curve $\alpha_2(t) = r_2 t + b_2$. Let $\theta = T + b_2/R$. Define β_1 as follows:*

$$\begin{aligned} \beta_1(t) &= [\beta(t) - \alpha_2(t - \theta)]^+ \; for \; t \geq \theta \\ &= 0 \; for \; 0 \leq t < \theta \end{aligned}$$

Then β_1 is a rate latency work service curve for F_1 with rate $R - r_2$ and latency θ.

Proposition 3.2 *[PQ Split] A service node $\mathcal{N}$ gives prioritized packet flows F_i ($i = 1, \cdots, P$; F_i has higher priority than F_j whenever $i > j$) non-preemptive priority*

service. Let δ_i be the service time of packets of flow F_i for $i = 1, \cdots, P$. Suppose that each F_i is constrained by the leaky bucket work arrival curve $\alpha_i(t) = r_i t + b_i$. For $j = 1, \cdots, P$ define

$$\begin{aligned} L_j &= \max\{\delta_i : i < j\} \\ r_j^H &= \sum_{\{i:i>j\}} r_i \\ b_j^H &= \sum_{\{i:i>j\}} b_i \\ \beta_j(t) &= \max\{t - r_j^H t - L_j - b_j^H, 0\} \end{aligned}$$

If $r_j^H < 1$, then β_j is a rate latency work service curve for F_j.

One could in principle use the above results directly to do a network calculus of the star and line topologies considered in this paper. However, the iterative application of the FCFS and PQ split calculations along a chain of service nodes yields unacceptably high delay bounds due to the rapid increase in burst and latency values. The way around this problem is to consider the service curve offered to a flow in a complete subsystem. If flow F with work arrival curve α is given a work service curve β_1 in node subsystem $\mathcal{S}_1$ before entering node subsystem $\mathcal{S}_2$ where it is given work service curve β_2, then F is given the work service curve $\beta = \beta_1 \otimes \beta_2$ in the combined system $\mathcal{S} = \mathcal{S}_1 \cup \mathcal{S}_2$. The Network Calculus can then be done using the input arrival curve α and the service curve β in $\mathcal{S}$. Compare the "pay bursts only once principle" in [BT01], 1.5.3.

3.3 Network Calculus for the line topology

In the case of the line topology we sketch the procedure used in [Wat02] to find suitable flows and subsystems. For the line topology the packet flow $F_{k,i}$ in a message type (class) k with confirmed service is for $i = 2, ..., N$

$$\mathcal{ST}_1(request) \to \mathcal{SW}_1 \to \cdots \to \mathcal{SW}_i \to \mathcal{ST}_i$$

$$\mathcal{ST}_i(response) \to \mathcal{SW}_i \to \cdots \to \mathcal{SW}_1 \to \mathcal{ST}_1$$

and for unconfirmed service

$$\mathcal{ST}_i(data) \to \mathcal{SW}_i \to \cdots \to \mathcal{SW}_1 \to \mathcal{ST}_1$$

Let $F_{ki}^* = \bigcup_{j>i} F_{k,j}$ be the aggregate flow of packets in class k which are transmitted from $\mathcal{SW}_i$ to $\mathcal{SW}_{i+1}$ for $1 < i < N$.

Let $\mathcal{S}_i$ denote the subsystem $\mathcal{ST}_1(request) \to \mathcal{SW}_1 \to \cdots \to \mathcal{SW}_i$ for $i = 1, \cdots, N$. Suppose that we have derived a service curve for the confirmed service flows $F_{k,i}$ and $F_{k,i}^*$ in $\mathcal{S}_{i-1}$. Note that $F_{k,i-1}$ has a different output port in $\mathcal{SW}_{i-1}$ and is served in parallel. We first compute a work arrival curve for $F_{k,i}^*$ at $\mathcal{SW}_i$ by considering the packetized output of this flow from $\mathcal{S}_{i-1}$. We then apply the PQ and FCFS splits at $\mathcal{SW}_i$ to derive service curves for $F_{k,i+1}$ and $F_{k,i+1}^*$, firstly in $\mathcal{SW}_i$, and then in $\mathcal{S}_i$ by concatenation. By iteration we thus find service curves for the F_{ki} in the subsystem $\mathcal{S}_i$. Using the arrival curve for this flow at $\mathcal{ST}_1$, we can then compute a delay bound for the flow F_{ki} on the path $\mathcal{ST}_1 \to \mathcal{SW}_i$ as well as an arrival curve for F_{ki} at $\mathcal{ST}_i$ where the response packets are generated. The unconfirmed packet flows have to be added at $\mathcal{ST}_i$. We can compute work arrival curves $A_w(F_{ki}, \mathcal{ST}_i)$ at $\mathcal{ST}_i$ by considering the packet output of $\mathcal{S}_i$.

For the highest priority packets (or for all packets in case of FCFS), we can calculate the delay bound on the outbound path from $\mathcal{ST}_1$ to $\mathcal{ST}_N$ as follows. Let $r[packet/s]$ and $b[packet]$ be the packet rate and burst parameters of a leaky bucket arrival curve constraining the total highest priority flow of request packets at $\mathcal{ST}_1$. Let δ_{req} and δ_{rsp} be the service times of its request and response packets respectively. The delay of highest priority packets from $\mathcal{ST}_1$ to $\mathcal{ST}_N$ is bounded by

$$D_H^{out} = \delta_L + b\delta_{req} + N\delta_{max}^{out}$$

where δ_L is the maximum service time in $\mathcal{ST}_1$ of a lower priority packet (0 if there are none) and δ_{max}^{out} is the maximum service time of an outbound request packet (cf. [Wat02]).

On the inbound path from the $\mathcal{ST}_i$ back to $\mathcal{ST}_1$ it is advantageous to aggregate the higher priorities. Let $p(k)$ denote the priority of class k and let

$$\hat{F}_{ki} = \cup\{F_{jm} : p(j) \geq p(k), m \geq i\}$$

be the aggregate of all flows of priority $\geq p(k)$ through $\mathcal{SW}_i$. For $i = 1, \cdots, N$ define the subsystem

$$\mathcal{S}_i^{in} = \{\mathcal{ST}_j : j \geq i\} \cup \{\mathcal{SW}_j : j > i\}$$

The output of this subsystem is the input for $\mathcal{SW}_i$. The input of flow $\hat{F}_{ki}$ to $\mathcal{S}_i^{in}$ is constrained by the work arrival curve

$$A_w(\hat{F}_{ki}, \mathcal{S}_i^{in}) = \sum_{\{j,m:p(j)\geq p(k), m\geq i\}} A_w(F_{jm}, \mathcal{ST}_m)$$

For a class k let μ_k be the maximum service time of an inbound packet of priority at least $p(k)$, and let ν_k be the maximum service time of an inbound packet of priority less than $p(k)$. We set $\nu_k = 0$ if there are no lower priority packets. Define the service curve

$$\tilde{\beta}_w(k, t) = [t - \mu_k - \nu_k]^+$$

Then $\tilde{\beta}_w(k)$ is a packetized work service curve for $\hat{F}_{ki}$ in $\mathcal{SW}_i$. It is shown in [Wat02] that the aggregate flow $\hat{F}_{ki}$ in $\mathcal{S}_i^{in}$ is given the packetized work service curve

$$\tilde{B}_w(\hat{F}_{ki}, \mathcal{S}_i^{in}) = \underbrace{\tilde{\beta}_w(k) \otimes \cdots \otimes \tilde{\beta}_w(k)}_{(N-i+1)-fold}$$

The input of flow $\hat{F}_{ki}$ to $\mathcal{SW}_i$ is constrained by the work arrival curve

$$A_w(\hat{F}_{ki}, \mathcal{SW}_i) = A_w(\hat{F}_{ki}, \mathcal{S}_i^{in}) \oslash \tilde{B}_w(\hat{F}_{ki}, \mathcal{S}_i^{in})$$

Let m be the class with the largest priority less than $p(k)$ and let $G_{mi} = \cup\{F_{mj} : j \geq i\}$ be the aggregate flow of packets of priority m which traverse $\mathcal{SW}_i$ for $1 \leq i \leq N$. We first apply the PQ split at $\mathcal{ST}_N$ to calculate a service curve for G_{mN}:

$$B_w(G_{mN}, \mathcal{SW}_N, t) = [t - \nu_m - A_w(\hat{F}_{kN}, \mathcal{S}_N^{in})(t)]^+$$

We then apply the PQ split at the service node $\mathcal{SW}_i$ to derive a service curve for G_{mi} for $1 \leq i \leq N$:

$$B_w(G_{mi}, \mathcal{SW}_i, t) = [t - \nu_m - A_w(\hat{F}_{ki}, \mathcal{SW}_i)(t)]^+$$

The concatenation of these service curves yields a service curve for G_{m1} in $\mathcal{S}_1^{in} \cup \mathcal{SW}_1$, from which we can compute a delay bound for the inbound path. The sum of the outbound and inbound delay bounds for each confirmed priority class gives the desired end-to-end delay bound.

It is shown in [Wat02] that the end-to-end delay of the highest priority packets (or for all packets in case of FCFS) is bounded by

$$D_H = D_H^{out} + (b + rD_H^{out})\delta_{rsp} + N\delta_{rsp} + (N+1)\nu$$

where ν is the maximum service time of an inbound packet of lower priority (0 if there are none).

4 Real-Time Evaluation

4.1 Simulation Results

For evaluating the real-time behavior of switched Ethernet, mean values are not sufficient because they only allow a statement of the typical time response. Therefore, the distributions of the absolute transaction times recorded by simulation are considered first. Fig. 2 shows the absolute transaction times (globally[1]) determined by simulation as a function of selected relative system loads in a so-called box plot for the star topology with a medium device number ($N = 50$) and FCFS scheduling strategy. Load and system parameters shown in Table 2 apply. The processing time within the stations and the network delay in the switches are set to $T_{TH} = T_S = 0$.

Box plots present a compact graphical visualization of the frequency distribution of data records on the basis of a 5-number summary (pentagram) [Sch98]. The rectangle (box) includes the first quartile $\tilde{X}_{0.25}$ up to the third quartile $\tilde{X}_{0.75}$ and thus comprises 50% of all data in the center of a distribution. The horizontal line in the rectangle locates the median $\tilde{X}$ (second quartile). The upper and lower horizontal line represent the maximum and minimum values of the distribution. The additional circle is the mean value $\bar{X}$. Furthermore, the values for the 95% percentile (dot-and-dash line) and the maximum (dashed

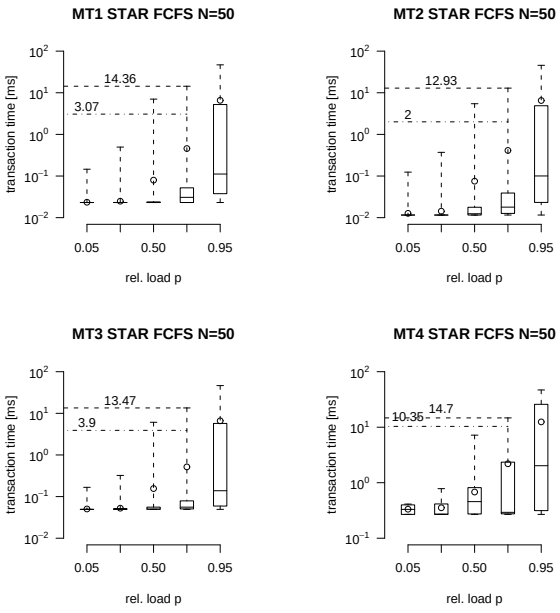

Figure 2. Box plots of the time distribution for star topology and FCFS

line) are shown for a load of $\rho = 0.8$. Generally, all distributions in Fig. 2 have a positive skew. The "collapsed rectangle" for the message types $MT_1 - MT_3$ clearly shows that for low system loads the transaction times come close to the mean value.

For real-time evaluation the upper extreme values are of vital interest. As an example a deterministic upper limit (i.e. absolute upper bound) and a statistical upper limit with $Z_{Max} = 0.95$ (i.e. distribution percentile) are shown for a system load of $\rho = 0.8$. For the chosen load of $\rho = 0.8$ the result for all message types and the star topology is that a deterministic upper limit less than $15\ ms$ can be guaranteed for a *single* transaction. For the statistical time limit the value for $MT_1 - MT_3$ is less than $4\ ms$ in 95% of all cases and equal to $10.36\ ms$ for MT_4.

For the line topology (Fig. 3) the central area of the transaction times has a much greater distance to the lower extreme value than in the star topology (Fig. 2). This "wider" distribution leads to accordingly higher mean values of the transaction times. Due to the high number of switches the jitter is already at a high level even if system loads are small. In the line topology an upper deterministic limit less than $19\ ms$ can be kept for all message types at a load of $\rho = 0.8$. For the statistical time limit the 95% percentile for $MT_1 - MT_3$ is less than $6\ ms$ and equal to $13.03\ ms$ for MT_4. Attention should be drawn to the fact that there is a big difference between the corresponding mean values for both topologies and a comparatively small difference for the extreme values determined for a medium load. In comparison, Figs. 4 and 5 show the box

[1] In this context globally means that the diagrams comprise the transaction times of all links within one message type.

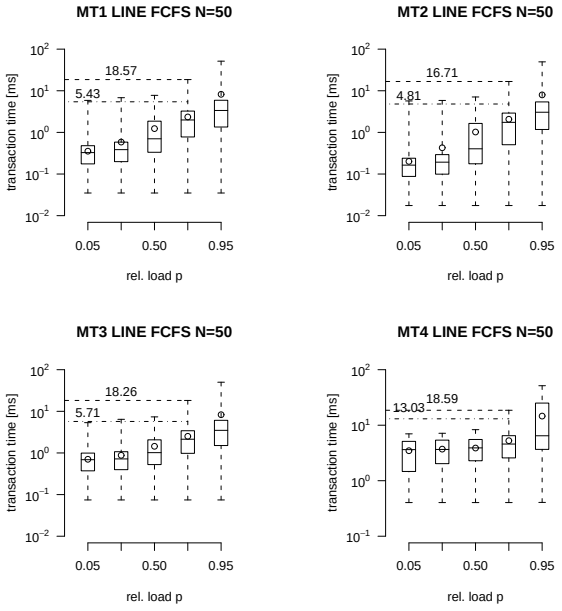

Figure 3. Box plots of the time distribution for line topology and FCFS

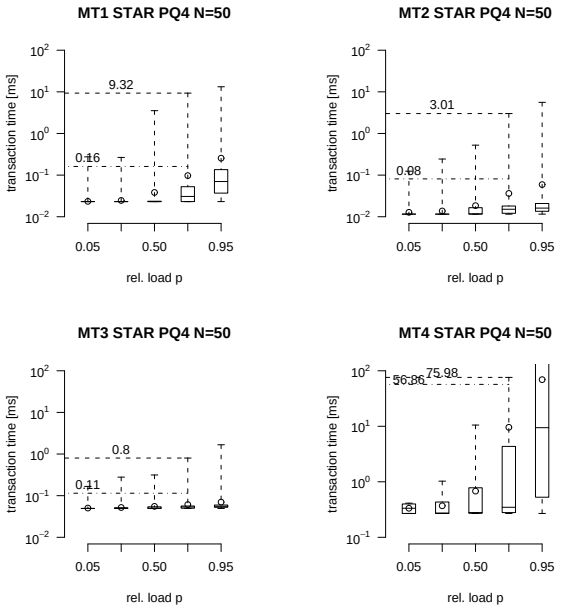

Figure 4. Box plots of the time distribution for star topology and PQ4

plots when the PQ scheduling strategy with four traffic classes ($PQ4$) is applied. The preferential service given to the high priority message type MT_3 in the star topology (Fig. 4) is a typical feature. The rectangles of the box plots have collapsed in the entire load range selected. Even the maximum values are much smaller than in the FCFS scenario. In the selected load profile the message types MT_2 and MT_1 also have a better time treatment with regard to the absolute delay time and the jitter. This is clearly shown with the smaller rectangles and the maximum values. The distance between the 95% percentile and the absolute upper limit for $\rho = 0.8$ and above indicates that there is only a small number of large values in the distribution (long tail). However, the message type MT_4 is given a much less favorable service than in the FCFS scenario beginning from a system load of 50%. This behavior is not critical because this message type has a throughput-oriented definition and does not include time-critical applications (best effort). The priority settings effectively isolate the influence of the long data packets in MT_4 on the time-sensitive transactions of the message types MT_1 to MT_3 with their short data packets.

When applying PQ in line topologies the maximum value for the high-priority message type MT_3 is only weakly dependent on the load. This effect becomes less as the priority drops. As expected the jitter increases due to the numerous switches to be traversed. This leads to the ”wider” distribution of transaction times, which we already know from FCFS and the line topology.

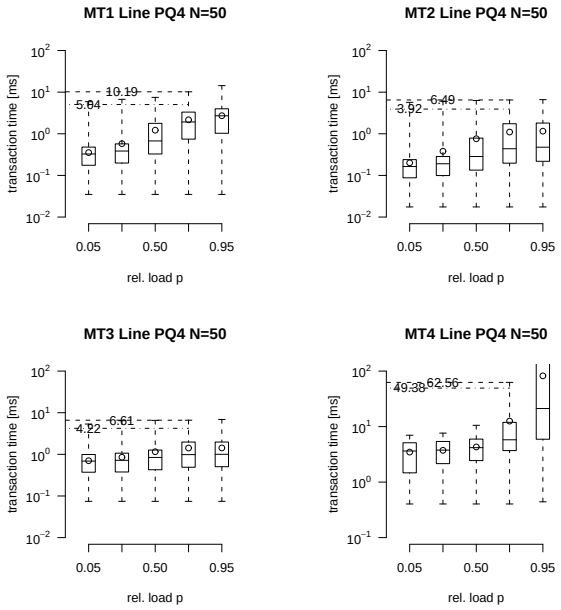

Figure 5. Box plots of the time distribution for line topology and PQ4

4.2 Analytical Delay Bounds

Especially when considering the rare upper extreme values relevant for real-time processing, the question arises as to whether the absolute upper bound can be determined with simulation experiments. Without doubt the maximum value determined by simulation can occur. However, we have to take into account that even higher values may occur when the simulation runs are longer or when the start conditions are changed. In any case

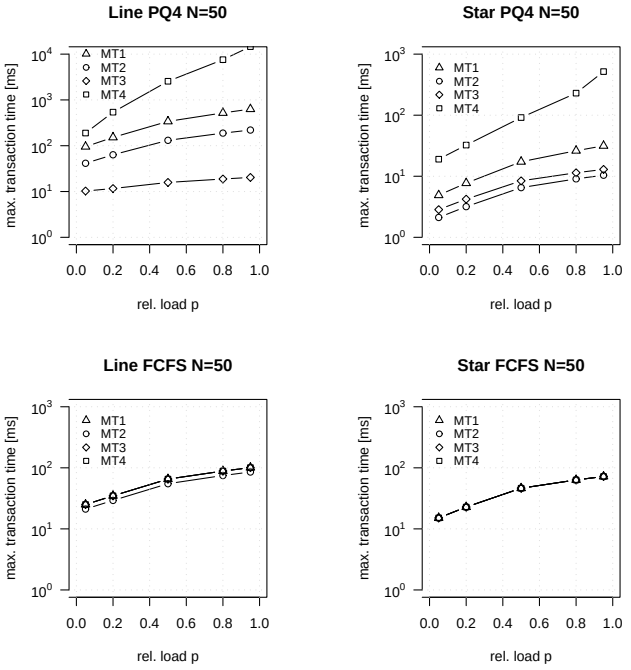

Figure 6. Upper bounds on the transaction times determined by analysis

the simulation results allow the statement that an upper time limit set by the application is exceeded if its value is smaller than the maximum value determined by simulation.

The analytical method presented above was applied to derive sure upper bounds on the transaction times. Fig. 6 shows the analytical results for the transaction time bounds as a function of the relative system load (and implicitly the burst value) for the star and line topologies as well as for the $PQ4$ and FCFS scheduling strategies. For a load of $\rho = 0.8$ the analytically determined values are 2 to 40 times greater than the maximum transaction times determined by simulation. Let us discuss the reasons for these deviations by means of a so-called "bad case" (but possible) scenario for the FCFS scheduling strategy:

Consider a set $\mathcal{P}$ of packets generated in the time interval $[t_0, t_\omega]$ which are to be served in the FCFS switch $\mathcal{SW}$, ($\mathcal{SW}_1$ for the line topology). The packet paths before reaching $\mathcal{SW}$ may be arbitrary. Let W_S be the work in the Service Node $\mathcal{SW}$. We will consider the delay of the packets in the system until leaving $\mathcal{SW}$. Now let t_1 be the first arrival of a packet of $\mathcal{P}$ at $\mathcal{SW}$. Note that this packet need not be the first generated. Let t_S be the last arrival of a packet of $\mathcal{P}$ at $\mathcal{SW}$. Note that this packet need not be the last generated. Denote it by p_S. Cf. Fig. 7.

In the interval $[t_1, t_S]$, work amounting to at least W_S arrives at $\mathcal{SW}$. In the same interval, the amount of work executed is at most $t_S - t_1$. Hence the work remaining at time t_S is at least $W_S - (t_S - t_1)$. Since $\mathcal{SW}$ is FCFS, the delay of p_S in $\mathcal{SW}$ is at least $D(p_S, S) = W_S - (t_S - t_1)$.

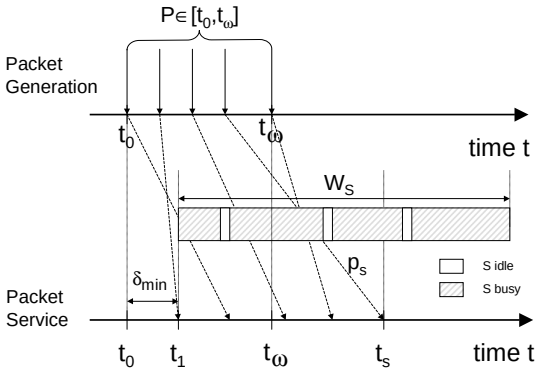

Figure 7. "Bad case" scenario

The delay of p_S from its generation until reaching $\mathcal{SW}$ is at least $D_0(p_S) = t_S - t_\omega$.

Therefore the delay of p_S until leaving $\mathcal{SW}$ is at least

$$D(p_S) \geq D_0(p_S) + D(p_S, S) \geq W_S + t_1 - t_\omega$$

Now $t_1 \geq t_0 + \delta_{min}$, where δ_{min} is the minimum time required by a packet to reach $\mathcal{SW}$. Depending on the network topology and service strategies, it is in general possible to construct packet generation scenarios with larger values for t_1. We conclude that

$$D(p_S) \geq W_S + t_0 - t_\omega + \delta_{min}$$

In order to obtain the highest possible delay of a data packet ("bad case"), scenarios must be found which maximize the right hand side part of the above inequality. The analytical method assumes that on all connections (even within a message type) a burst can be generated at the same time. This results in a high work W_S in the interval. However, in the simulation model the connections within a message type are not configured independently. There is always a finite time $t > 0$ between the bursts on the different connections within a message type MT_i. In the simulation method the work W_S is therefore generated within a larger time interval $t_\omega - t_0$ than in the analytical method and thus leads to smaller time bounds $D(p_S)$.

5 Summary

The analytical method presented in this paper used Network Calculus techniques to derive deterministic guarantees for bounds on transaction times, if the "worst case" source behavior can be characterized. The simulation results confirmed the applicability of these bounds for scenarios with equivalent source behavior.

The presented results show that, with regard to maximum transaction times the results of the analytical method applied here show noticeable deviations from the results of the simulation method. We therefore have to be careful if simulation with the necessary use of random generators is the only method applied for real-time evaluation. We

can only be certain that a limit value set by a technical process is exceeded if the maximum value determined by simulation is higher than this limit value.

An open issue in general is the derivation of a quality metric for the upper bounds determined by the Network Calculus approach. However, in some situations it is possible to calculate ”bad cases” which show that the Network Calculus upper bound is only moderately higher than the worst case. Simulation is a useful approach to generate ”bad cases”.

It was also shown that using more than one traffic class with user priorities leads to a good isolation of time-sensitive and time-uncritical data. In this case the upper time bound of high-priority applications can be reduced by a value comparable with FCFS scheduling, but only at a the cost of the low-priority application.

The question is still open as to whether switched Ethernet has real-time capabilities or not. This primarily depends on the requirements of the technical process and cannot be answered in general. But it was shown that in a typical application with $N = 50$ devices maximum transaction times on the MAC level within milliseconds could be guaranteed for real-time services.

References

[BT01] Le Boudec and P. Thiran. Network calculus: a theory of deterministic queuing systems for the Internet. In *Lecture Notes in computer science*, volume 2050. Springer Verlag, 2001.

[Cru99a] Rene L. Cruz. A Calculus for Network Delay, Part I: Network Elements in Isolation. *IEEE Transactions on Information Theory*, 37:114–131, Jan. 1999.

[Cru99b] Rene L. Cruz. A Calculus for Network Delay, Part II: Network Analysis. *IEEE Transactions on Information Theory*, 37:132–141, Jan. 1999.

[GP93] R. G. Gallager and A. K. J. Parekh. A generalized processor sharing Approach to flow control in integrated services networks: the single-node case. In *IEEE/ACM Transaction on Networking*, volume 1, pages 344–357, June 1993.

[GP94] R. G. Gallager and A. K. J Parekh. A generalized processor sharing Approach to flow control in integrated services networks: the multiple-node case. In *IEEE/ACM Transaction on Networking*, volume 2, pages 137–150, April 1994.

[IEE98a] IEEE, New York. *Media Access Control (MAC) Bridges*, 1998. ANSI/IEEE Std 802.1D-1998.

[IEE98b] IEEE, New York. *Virtual Bridged Local Area Networks*, 1998. ANSI/IEEE Std 802.1Q-1998.

[IEE00] IEEE, New York. *Carrier sense multiple access with collision detection (CSMA/CD) access method and physical layer specifications*, 2000. ANSI/IEEE Std 802.3-2000 Edition (ISO/IEC 8802-3:2000(E)).

[Jas02] J. Jasperneite. Performance Evaluation of a Class-of-service based Local Area Network for using at the Field device level. Internal Research Report, Phoenix Contact, 2002.

[JN00] J. Jasperneite and P. Neumann. Measurement, Analysis and Modeling of Real-Time Source Data Traffic in Factory Communication Systems. In *2000 IEEE International Workshop on Factory Communication Systems*, pages 327–334, Porto, Portugal, Sept. 2000.

[JN01] J. Jasperneite and P. Neumann. Switched Ethernet for Factory Automation. In *8th IEEE International Conference on Emerging Technologies and Factory Automation*, pages 205–212, Antibes - Juan les Pins, France, Oct. 2001.

[Sch98] R. Schlittgen. *Einführung in die Statistik – Analyse und Modellierung von Daten*. Oldenbourg Verlag, München/Wien, 1998.

[Wat02] K. Watson. Network Calculus in Star and Line Networks with centralized communication. Technical report, Fraunhofer-Institut für Informations- und Datenverarbeitung (IITB), Karlsruhe, April 2002. Bericht-Nr. 10573.