Available at
www.ElsevierComputerScience.com
POWERED BY SCIENCE DIRECT
Computer Standards & Interfaces 27 (2005) 105–115
COMPUTER STANDARDS & INTERFACES
www.elsevier.com/locate/csi

# Real-time performance evaluation of urgent aperiodic messages in FF communication and its improvement☆

Zhi Wang[a,*], Xingfa Shen[a], Jiming Chen[a], Yeqiong Song[b], Tianran Wang[c], Youxian Sun[a]

[a] *National Laboratory of Industrial Control Technology, Zhejiang University, ZheDa Road 38, Hangzhou 310027, China*
[b] *LORIA-ENSEM, 2, av. de la Forêt de Haye, 54516, Vandoeuvre-lès-Nancy, France*
[c] *Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110016, China*

Received 12 September 2003; received in revised form 17 April 2004; accepted 7 May 2004
Available online 2 June 2004

**Abstract**

In distributed control system (DCS), computation tasks, which usually reside in different nodes and communicate with one another to accomplish a common goal, have to be executed timely. Therefore, there is a need for special-purposed real-time industrial networks. Foundation Fieldbus (FF) is characterized by explicitly distinguishing between periodic and aperiodic messages. Centralized Media Access Control (MAC) is utilized by FF to support periodic messages, and distributed MAC to support aperiodic messages. It is indicated that FF's current mechanism using Pass Token (PT) priority and PT circulation period cannot effectively guarantee the real-time requirement of important aperiodic messages. A detailed analysis shows the method of computing Actual PT Circulation Period (APTCP) should be improved. An improved method of computing APTCP is proposed in this paper, and then a method of determining Setting PT Circulation Period (SPTCP) is presented accordingly. In the end, simulation validates the effect of the improved method on the real-time performance of urgent aperiodic messages in FF.


*Keywords:* Foundation Fieldbus; Real-time; Aperiodic traffic; Distributed control system; Token circulation period

## 1. Introduction

The increasing complexity of control systems, as well as the large dimension of applications, such as process control, factory automation, space vehicle system, etc., has lead to the development of distributed control system (DCS), where control commands and state information are exchanged through networks. Within DCS, it is very necessary to execute computation tasks timely, which usually reside in different nodes and communicate with one another to accomplish a common goal. It is difficult to ensure timely results of tasks in a DCS without a network that supports the timely inter-task messages [1–3]. Therefore the temporal property of the underlying network is important, and special-purposed real-time industrial networks are in need.

☆ This work is supported by National Natural Science Foundation of China (NSFC 60203030, 60084001), National Basic Research Program of China (2002CB312200) and Advance Research Program of France-China(PRA SI03-02).
* Corresponding author. Tel./fax: +86-571-87951879.
*E-mail address:* wangzhi@iipc.zju.edu.cn (Z. Wang).

0920-5489/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.csi.2004.05.001

Foundation Fieldbus (FF), one of eight IEC international Fieldbus standards proposed by Fieldbus Foundation, is just a kind of special-purposed real-time industrial networks. FF-liked networks are generally characterized by the obligation to respect stringent temporal constraint, which must be met to guarantee the correctness and safety of field devices [4–6]. In order to achieve above object, centralized Media Access Control (MAC) is utilized by FF to support periodic messages, and distributed MAC to support aperiodic messages.

Concerning periodic messages, the main problem is constructing schedule table and schedule algorithm. Similar researches can be found in Refs. [7–14]. As for aperiodic messages, FF utilizes the mechanism of Pass Token (PT) priority and PT circulation period to meet its temporal constraint. At the aspect of guaranteeing important aperiodic messages, there are only few mechanisms, such as in Refs. [15,16]. Moreover, these researches are not detailed enough. This paper not only analyzes the effect of PT circulation period on Worst-Case Response Time (WCRT) of aperiodic messages in detail, but also points out main cause of current mechanism's deficiencies. Furthermore, this paper proposes an improved one and validates it with simulation.

The remainders of this paper are organized as follows: Section 2 describes the primary transfer procedures of periodic and aperiodic messages in FF and corresponding models. Then in Section 3, the effect of Setting PT Circulation Period (SPTCP) and Actual PT Circulation Period (APTCP) on real-time communication of aperiodic messages is evaluated from different aspects of view, such as guaranteeing *urgent* aperiodic messages, high level utilization of FF, effectiveness of FF on best-effort transfer of *normal* aperiodic messages, etc. In Section 4, an improved APTCP computing mechanism and then a method of setting SPTCP are proposed. The simulation results for the proposed mechanism are presented and analyzed in Section 5. Finally, in Section 6 some conclusions are given.

## 2. Communication model of aperiodic messages

### 2.1. Network and message models

Consider a DCS where there are $n_n$ nodes interconnected by a FF network. Assume there are $n_P$ periodic messages $M_P^i$=($C_P^i$, $T_P^i$, $D_P^i$)($i$∈[1, $n_P$]) and $n_A$ aperiodic messages $M_A^i$=($C_A^i$, $T_A^i$, $D_A^i$)(i∈[l, $n_A$]) within the DCS. Wherein, $T_P^i$, $D_P^i$ and $C_P^i$ correspond to periodicity, deadline and transaction duration of periodic message $M_P^i$, respectively. $T_A^i$, $D_A^i$ and $C_A^i$ correspond to periodicity, deadline and transaction duration of aperiodic messages $M_A^i$, respectively. For aperiodic messages, their arrival durations usually are irregular, however it is assumed that there is a minimum inter-arrival time in order to guarantee its temporal constraint.

Within FF, aperiodic messages are classified into *urgent*, *normal* and *available*. Therefore, the following description for concrete aperiodic messages with its priority and node index is presented.

$$M_U^{i,j} = (C_U^{i,j}, T_U^{i,j}, D_U^{i,j}) \quad (j \in [1, n_U^i]) \tag{1}$$

$$M_N^{i,j} = (C_N^{i,j}, T_N^{i,j}, D_N^{i,j}) \quad (j \in [1, n_N^i]) \tag{2}$$

$$M_V^{i,j} = (C_V^{i,j}, T_V^{i,j}, D_V^{i,j}) \quad (j \in [1, n_V^i]) \tag{3}$$

$$M_A^{i,j} = (C_A^{i,j}, T_A^{i,j}, D_A^{i,j}) \quad (j \in [1, n_U^i + n_N^i + n_V^i]) \tag{4}$$

$$\sum_{i=1}^{n_n} (n_U^i + n_N^i + n_V^i) = n_A \tag{5}$$

where, $M_U^{i,j}$, $M_N^{i,j}$, and $M_V^{i,j}$ correspond to $j$th *urgent*, *normal* and *available* aperiodic messages in $i$th node, respectively.

### 2.2. Transfer procedure of aperiodic messages

In order to meet the requirement of aperiodic messages with different levels of criticality in temporal aspect, FF provides PT priority and PT Circulation Period (PTCP).

Corresponding to different aperiodic messages' priorities, PT priorities are differentiated into three classes: *urgent*, *normal* and *available*. For an aperiodic message, it is transmitted only when its priority is not less than current PT priority and its transfer time is less than Maximum Token Hold Time (MTHT), which is set in PT frame. PT returns to Link Active Scheduler (LAS) when no aperiodic message with proper priority exists or MTHT expires [16].

PTCP is the time duration between PT reaches a same node twice consecutively. Within FF, there are two kinds of PTCP, Setting PT Circulation Period and Actual PT Circulation Period, where the former is set online or offline by an operator and the latter is measured online. Within FF, APTCP equals to the measured value of the time duration between PT reaches a particular node with the minimum address (without losing generality, the minimum address is set as 1 in this paper) twice consecutively.

To enhance real-time response of critical aperiodic messages in FF, LAS changes PT priority online according to the difference between SPTCP and APTCP. The detail is shown in Fig. 1.

## 2.3. PT visiting model

Intervals between PT twice consecutive visiting a same node are different because of PT experiencing different loads of periodic and aperiodic messages during these intervals. However, PT priority is only changed at node 1.

To describe the behavior of a sequence by which PT visits a node, a pair of subscripts are introduced, *visit*(*c,i*), where *c* indicates the order of PT visiting and *i* indicates the address of node being visited. That means *visit*(*c,i*) indicates the beginning instant that PT visits node *i* at its *c*th order. According to definition, *visit*(*c,i*) has following property: if $(i \neq n_n)$, *visit*(*c,i*) is followed by *visit*(*c,i*+1), otherwise by *visit*(*c*+1,1), on the other hand, if $(i \neq 1)$, the previous one before *visit*(*c,i*) is *visit*(*c,i* − 1), otherwise is *visit*(*c* − 1, $n_n$).

Further, to describe the elapsed time between PT visiting node *i* and node *m* that starts from *visit*(*c,i*) and ends at *visit*(*c,m*), $\mathrm{Del}_c^{i,j}$ is introduced. Obviously, *visit*(*c,i*) and $\mathrm{Del}_c^{i,j}$ have the following property,

$$\mathrm{Del}_c^{i,j} = \begin{cases} visit(c, i+j) - visit(c,i) & (i+j) \leq n_n \\ visit(c + (i+j)/n, (i+j)\%n_n) - visit(c,i) & \\ & (i+j) > n_n \end{cases} \quad (6)$$

Let *APTCP(c)* be the elapsed time of APTCP at its *c*th order, which means *APTCP(c)* is from the time PT visits node 1 at its *c*th order to the time PT visits node 1 at its (*c*+1)th order. According to the definition of APTCP, *APTCP(c)* equals to *visit*(*c*+1,1) − *visit*(*c*,1) or $\mathrm{Del}_c^{l,n_n}$.

The relationships among *visit*(*c,i*), $\mathrm{Del}_c^{i,j}$, APTCP and SPTCP are illustrated in Fig. 2.

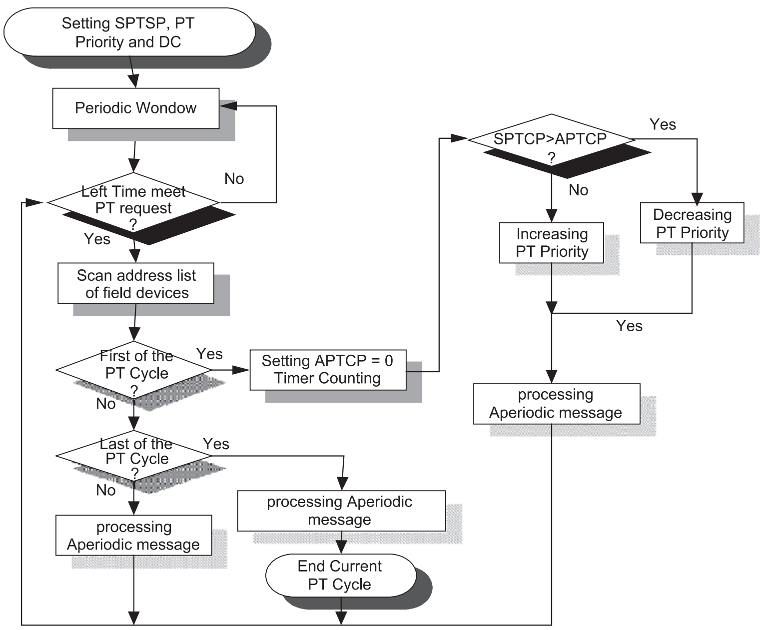


Fig. 1. LAS schedule for aperiodic messages.

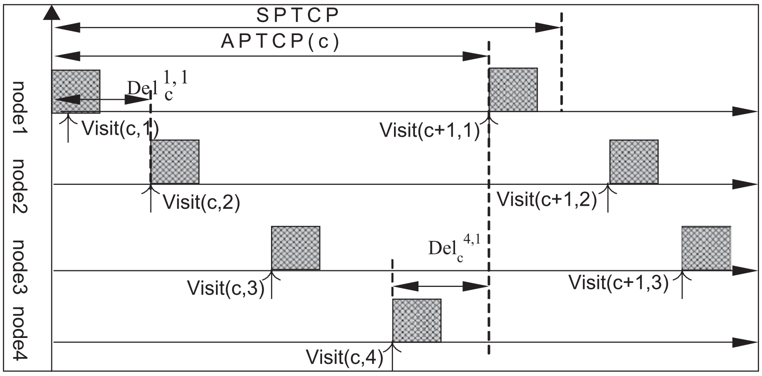


Fig. 2. Relationship among related parameters of PT visiting model.

### 2.4. Changing procedure of PT priority

From Sections 2.2 and 2.3, PT priority changes online at the instant of $visit(c+1,1)$ according to difference between SPTCP and $APTCP(c)$, and to the PT priority in last cycle. Therefore, the following state transfer graph for PT priority illustrated is presented in Fig. 3.

### 2.5. Queue delay for urgent aperiodic messages

For convenient description of waiting procedure of aperiodic messages, their priority are neglected temporally, and just regard all aperiodic messages as the same priority.

As for $M_{\mathrm{A}}^{i,j}$, it cannot be transferred immediately it arrives at node $i$ and must wait until the arrival of PT at this node. Further, $M_{\mathrm{A}}^{i,j}$ is transferred according to First Come First Service (FCFS) rule within node $i$. Therefore, queue delay of $M_{\mathrm{A}}^{i,j}$ consists of the time for waiting arrival of PT at node $i$, for transmitting other aperiodic messages arriving earlier in this node, and for transmitting itself. It is obvious that the worst-case condition for $M_{\mathrm{A}}^{i,j}$ is that the following two situations occur simultaneously: $M_{\mathrm{A}}^{i,j}$ arrives at node $i$ just after PT leaves this node, and all other aperiodic messages in node $i$ arrive simultaneously except for $M_{\mathrm{A}}^{i,j}$ arriving later. The corresponding response procedure is illustrated in Fig. 4.

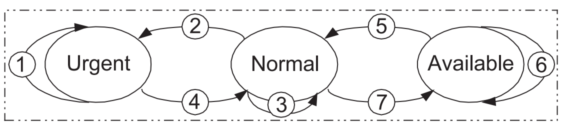


Fig. 3. State transfer graph for PT priority.

Let $R_{\mathrm{A}}^{i}$ denote response time of aperiodic message $M_{\mathrm{A}}^{i,j}$ under the worst-case situation, then

$$R_A^{i,j} = (\mathrm{Del}_{\mathrm{c}}^{i,l} - PT_f) + \sum_{x=i+1}^{N} \mathrm{Del}_c^{x,1} + \sum_{l=c+1}^{c+\lambda} \sum_{x=1}^{N} \mathrm{Del}_l^{x,1} + \sum_{x=1}^{i-1} \mathrm{Del}_{c+\lambda+1}^{x,1} + \sum_{x=1}^{n_A^i} C_A^{i,x} \qquad (7)$$

where $\lambda$ denotes the number of PT cycles that $M_{\mathrm{A}}^{i,j}$ must wait for proper PT priority, while $PT_f$ denote the time of sending and returning PT without transmitting aperiodic messages.

Note that Eq. (7) is only a general expression of WCRT of any aperiodic message. Further, the last item in Eq. (7) maybe includes only $C_{\mathrm{A}}^{i,j}$ instead of all aperiodic messages in node $i$. In fact, the contents contained in the last term are related to all the previous several items in Eq. (7). The items from the second to the fifth are related to priority of considered aperiodic message. Therefore, Eq. (7) is only an upper bound of WRCT.

### 2.6. Transfer capability of aperiodic window (AW)

LAS sends PT for aperiodic messages only within Aperiodic Window (AW), whose transfer capability varies in different instants and is related to the content in the corresponding Periodic Window (PW) [11]. Therefore, to calculate the transfer capability, the content of periodic window, the schedule of periodic messages, which is determined by the arrival pattern

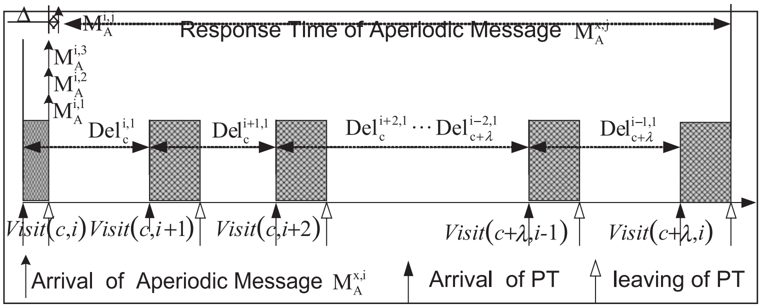


Fig. 4. Response procedure of an aperiodic message.

of periodic messages, must be considered. That means the transfer capability of aperiodic window integrated with arrival pattern of periodic messages has to be calculated.

Let $PW(i)$ and $AW(i)(i\in[1, N_{\mathrm{Mic}}])$ denote $i$th periodic window and aperiodic window, respectively, where $N_{\mathrm{Mic}}$ denote the number of microcycles within Scheduling Table (ST) [16]. If ST is constructed by the rule of Least Common Multiple (LCM) and Highest Common Factor (HCF), $N_{\mathrm{Mic}}$ equals to LCM/HCF of the periods of all periodic messages.

Let $W_{\mathrm{A}}^{i}$ and $W_{\mathrm{P}}^{i}$ be the time lengths of $AW(i)$ and $PW(i)$, respectively. According to above definition, the following equation is given

$$W_{\mathrm{P}}^{i} = \sum_{j=1}^{n_{\mathrm{P}}} (ST[j, i^{*}]) C_{\mathrm{P}}^{j} \tag{8}$$

where $ST[j, i^{*}]$ denotes whether periodic message $j$ is scheduled in the $i$th microcycle. $i^{*}$ denotes the order of the $i$th microcycle in ST, $i^{*}=[i\% N_{\mathrm{Mic}}]$.

$$W_{\mathrm{A}}^{i} = \mathrm{MicP} - W_{\mathrm{P}}^{i} \tag{9}$$

where MicP denotes time length of a microcycle. Let $\Omega_{\mathrm{A}}^{i,j}$ denote the number of microcycles for completing the transfer of $M_{\mathrm{A}}^{i}$ under worst-case situation.

$$\Omega_{\mathrm{A}}^{i,j} = \min(\psi) \cap \sum_{i=\beta}^{\beta+\psi-1} W_{\mathrm{A}}^{i} \geq \sum_{k=1}^{n_{\mathrm{A}}} f(C_{\mathrm{A}}^{K}, \psi, \beta) \tag{10}$$

Note that Eq. (10) does not contain $C$ the order of PT cycle or PT visiting, as Eq. (7) does. This is because actual traffic of aperiodic messages and the aperiodic window are uncertain. Within Eq. (7), $C$ is just for the convenience of description. Furthermore, Eq. (10) only explains how to calculate $\Omega_{\mathrm{A}}^{i,j}$, particularly the load of aperiodic messages is in the right part of Eq. (10), which will be adjusted according to actual permitted load of aperiodic messages.

Another point worthwhile to note in Eq. (10) is $\beta$, the order of aperiodic window. Since the variety of aperiodic window will lead aperiodic messages to suffer different waiting time, which is determined by the transfer capability of aperiodic window from the arrival of an aperiodic message to completing its transfer. Therefore, the worst-case order of aperiodic window need be considered, which means the aperiodic window starting from $\beta$ owns the less transfer capability. That is,

$$\beta = \left( maw \middle| \max\left( \min(\psi) \cap \sum_{i=maw}^{maw+\psi-1} W_{\mathrm{A}}^{i} \geq \sum_{k=1}^{n_{\mathrm{A}}} f(C_{\mathrm{A}}^{j}, \psi, \beta) \right) \right) \tag{11}$$

In the following section, $\beta$ denotes the worst-case order of aperiodic window unless explicitly stated.

Furthermore, considering the priority of aperiodic messages, the situation is complicated. For *urgent* aperiodic messages in node $i$, only the time length of aperiodic messages being transferred before PT arriving node $i$ need be considered, where the time length is related to current PT priority. For a *normal* aperiodic message in node $i$, it is allowed transferring only when the PT priority changes into *normal*. Therefore, the maximum waiting number of PT cycles for PT priority changing from *urgent* into *normal* has to be considered. Besides, the time length of other aperiodic messages being transferred before completing the regarded aperiodic message needs to be considered too. The detailed analysis of the WRCT of aperiodic messages will be done in Section 3.

## 3. The effect of SPTCP and APTCP on real-time communication of aperiodic messages

In the above section, the real-time capability of aperiodic messages is evaluated under the assumption that the instant of PT changing priority is known. However, the effect of SPTCP and APTCP is neglected actually. In this section, how to set SPTCP is investigated in detail. The following analysis is done by guaranteeing the real-time requirement of *urgent* aperiodic messages and by meeting the real-time requirement of *normal* aperiodic messages with best efforts (Just for convenience, *normal* and *available* aperiodic messages are regarded as the same).

### 3.1. Deadline of requirement of urgent aperiodic messages

To guarantee the real-time requirement of *urgent* aperiodic messages, every *urgent* aperiodic message must be transmitted before its deadline. Under this requirement, SPTCP should be less than the deadline of every *urgent* aperiodic message. Therefore SPTCP must at least arrive once during the period of two consecutive arrivals of an *urgent* aperiodic message. Consequently,

$$\mathrm{SPTCP_U} \le \min(D_{\mathrm{U}}^{i,j})(i\in[1,n],\ j\in[1,n_{\mathrm{U}}^{i}]) \tag{12}$$

Besides, the transfer capability of aperiodic window had better transfer all ongoing *urgent* aperiodic messages during one SPTCP in order to decrease the overhead during a PT cycle.

$$W_{\mathrm{A}}(\mathrm{SPTCP_U}) \ge L_{\mathrm{U}} \tag{13}$$

where $L_{\mathrm{U}}$ denotes the left capability after transferring all possible aperiodic messages, $W_{\mathrm{A}}$(SPTCP) denotes transfer capability of the aperiodic window during the period of SPTCP.

$$L_{\mathrm{U}} = \min(D_{\mathrm{U}}^{i,j}) - \sum_{i=1}^{n}\sum_{j=1}^{n_{\mathrm{U}}^{i}} C_{\mathrm{U}}^{i,j} - \delta\ (i\in[1,n],\ j\in[1,n_{\mathrm{U}}^{i}]) \tag{14}$$

$$W_{\mathrm{A}}(\mathrm{SPTCP}) = \left(\min\left(\sum_{i=maw}^{maw+\lfloor\frac{\mathrm{SPTCP}}{\mathrm{MicP}}\rfloor-1} W_{\mathrm{U}}^{i}\right)\right) \tag{15}$$

where $\delta$ denotes the overhead for transferring PT token during a PT cycle. Accordingly, PT priority is in *urgent* in this situation.

Therefore, $\mathrm{SPTCP_U}$ should meet the following condition,

$$\mathrm{SPTCP_U} \ge (\psi \mid \min(W_{\mathrm{A}}(\psi)) \ge L_{\mathrm{U}}) \tag{16}$$

### 3.2. Unfairness of APTCP and the dependence relation of node address on urgent aperiodic messages

It should be noted that the above expression is gotten under the worst case among all nodes. In fact, PT priority changes in node 1 when the previous APTCP is less than SPTCP. That means part of nodes maybe suffer from less delay and dependence relation of *urgent* aperiodic messages on node address.

For node *k*, the dependence relation is stated from the condition of SPTCP.

$$\mathrm{SPTCP_U^{k}} \ge (\psi \mid (\psi \mid \min(W_{\mathrm{A}}(\psi)) \ge L_{\mathrm{U}}^{k})) \tag{17}$$

where $W_{\mathrm{A}}$ and $L_{\mathrm{U}}^{k}$ meet the following condition,

$$W_{\mathrm{A}}(\mathrm{SPTCP_U^{k}}) \ge L_{\mathrm{U}}^{k}$$

$$L_{\mathrm{U}}^{k} = \min(D_{\mathrm{U}}^{k,j}) - \sum_{i=k+1}^{n}\left(\sum_{j=1}^{n_{\mathrm{U}}^{i}} C_{\mathrm{U}}^{i,j} + \sum_{j=1}^{n_{\mathrm{N}}^{i}} C_{\mathrm{N}}^{i,j}\right) - \sum_{i=1}^{k}\sum_{j=1}^{n_{\mathrm{U}}^{i}} C_{\mathrm{U}}^{i,j} - \delta(j\in[1,n_{\mathrm{U}}^{k}])$$

If *normal* aperiodic messages are taken as the block factor just as the real schedule theory does, it is obvious that node 1 suffers from maximum block factor. That is due to the computing mechanism of APTCP that causes PT priority changes only in node 1.

### *3.3. Lost control of SPTCP on guaranteeing urgent aperiodic messages*

PT changes into *normal* as soon as APTCP being less than SPTCP occurs, which will lead *urgent* aperiodic messages to suffering from larger delay since both *normal* and *urgent* aperiodic messages can be transferred during the next PT cycle. In this case, SPTCP cannot guarantee the real-time requirement of *urgent* aperiodic messages at all. This phenomenon is referred to as lost control of SPTCP. In order to avoid this phenomenon, the only way is to prevent APTCP being less than SPTCP from occurring. That means

$$\mathrm{SPTCP_L} \le \min(W_\mathrm{A}^i) + \delta \tag{18}$$

### *3.4. Transfer opportunity for normal aperiodic messages*

As for *normal* aperiodic messages, they get transfer opportunity only when PT priority is *normal* or *available*. That means APTCP being less than SPTCP must occur. It is difficult to find exact APTCP since aperiodic messages arrive irregularly. Therefore, just consider two extreme cases, one is that all aperiodic messages arrive at the maximum rate, the other is none of aperiodic messages exists during a PT cycle.

For case 1, the following condition must be met, otherwise *normal* aperiodic messages never get chance to be transferred.

$$\delta + W_\mathrm{P}(\mathrm{SPTCP_N^1}) \le \mathrm{SPTCP_N^1} \quad \text{(case 1)} \tag{19}$$

However, it is difficult to find a proper number of *urgent* aperiodic messages that can both guarantee the real-time requirement of *urgent* aperiodic messages and meets that of *normal* aperiodic messages with best efforts. If let SPTCP equal to $\mathrm{SPTCP_N^2}$ according to the case 2, the result is obviously improper since that it cannot guarantee the real-time requirement of *urgent* aperiodic messages at all.

$$\sum_{i=1}^{n}\left(\sum_{j=1}^{n_\mathrm{U}^i} C_\mathrm{U}^{i,j} + \sum_{j=1}^{n_\mathrm{N}^i} C_\mathrm{N}^{i,j}\right) + \delta + W_\mathrm{P}(\mathrm{SPTCP_N^2}) \le \mathrm{SPTCP_N^2} \quad \text{(case 2)} \tag{20}$$

Furthermore, the transfer capability of FF must at least meet all possible messages during a long time[17]. Accordingly,

$$\sum_{i=1}^{n}\left(\sum_{j=1}^{n_\mathrm{U}^j} C_\mathrm{U}^{i,j}\cdot\lambda_\mathrm{U}^{i,j} + \sum_{j=1}^{n_\mathrm{N}^j} C_\mathrm{N}^{i,j}\cdot\lambda_\mathrm{N}^{i,j}\right)\cdot t + \delta\cdot\frac{t}{\mathrm{SPTCP_{Lon}}} + W_\mathrm{P}(t) \le t \tag{21}$$

Besides, $\mathrm{SPTCP_{Lon}}$ should meet

$$\mathrm{SPTCP_{Lon}} \ge \left(1 - \sum_{i=1}^{n}\left(\sum_{j=1}^{n_\mathrm{U}^i} C_\mathrm{U}^{i,j}\cdot\lambda_\mathrm{U}^{i,j} + \sum_{j=1}^{n_\mathrm{N}^i} C_\mathrm{N}^{i,j}\cdot\lambda_\mathrm{N}^{i,j}\right) - \frac{1}{\mathrm{MacP}}\cdot\sum_{i=1}^{n} W_\mathrm{P}^i\right)^{-1}\cdot\delta \tag{22}$$

where $\mathrm{SPTCP_T}$ denotes the necessary SPTCP for all possible messages.

### *3.5. Lower efficiency of APTCP on normal aperiodic messages*

Furthermore, a less SPTCP can meet the real-time requirement of *urgent* aperiodic messages, but this SPTCP may introduce a bad waste of aperiodic window to transferring *normal* aperiodic messages. If SPTCP is set according to the extreme result in case 1, x and y may become very big and the utilization of

Table 1
Transfer capability of aperiodic window and overhead

| Microcycle (MicP) | 20 (ms) | | Overhead per PT cycle ($\delta$) | | | | 2 (ms) | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Aperiodic window | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| $W_A^i$ (ms) | 6 | 8 | 8 | 12 | 14 | 10 | 8 | 12 | 14 | 8 | 8 | 12 |

Table 2
Parameters of *urgent* and *normal* aperiodic messages

| Order of nodes | | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| *Urgent* aperiodic messages | Service time (ms) | 2 | 2 | 2 | 2 | 2 |
| | Arrival period (ms) | 40 | 45 | 50 | 48 | 50 |
| | Arrival rate (1/ms) | 50 | 60 | 70 | 60 | 70 |
| *Normal* aperiodic messages | Service time (ms) | 5 | 5 | 5 | 5 | 5 |
| | Arrival rate (ms) | 100 | 90 | 80 | 90 | 90 |

network resource may decrease. This phenomenon can be obviously seen from the following.

$$x \cdot \delta + \sum_{i=1}^{n} \sum_{j=1}^{n_U^i} C_U^{i,j} \le W_A(x \cdot \mathrm{SPTCP}_N^1) \quad (23)$$

$$y \cdot \delta + \sum_{i=1}^{n} \left( \sum_{j=1}^{n_U^i} C_U^{i,j} + \sum_{j=1}^{n_N^i} C_N^{i,j} \right) \le W_A(y \cdot \mathrm{SPTCP}_N^1) \quad (24)$$

where $x$ and $y$ are the number of PT cycles for PT changing into *normal* first time and completing transfer of *normal* aperiodic messages.

Example: Considering a given schedule in Table 1 and a set of periodic and aperiodic messages in Table 2, the following value for SPTCP can be obtained.

According to the given load of periodic and aperiodic messages and the conditions SPTCP must meet, the following results can be obtained:

$$\mathrm{SPTCP_U} \le 40 \text{ ms}; \mathrm{SPTCP_U^O} \ge 20 \text{ ms}; \mathrm{SPTCP_L} \le 8 \text{ ms};$$
$$\mathrm{SPTCP_N^1} \ge 8 \text{ ms}; \mathrm{SPTCP_N^2} \ge 80 \text{ ms}; \mathrm{SPTCP_{Lon}} \ge 20 \text{ ms}$$

From the obtained results, it can be seen obviously that there are lots of conditions that SPTCP has to meet, and most of these conditions are conflicting. Even the two extreme conditions ($\mathrm{SPTCP_N^1}$, $\mathrm{SPTCP_N^2}$) are neglected, it is still difficult to find a proper SPTCP from the left conditions.

## 4. Improvement of APTCP computing mechanism

### 4.1. Impact of various PT priority states on the response time of urgent and normal aperiodic messages

The analysis in previous sections indicates that the WCRT of *urgent* and *normal* aperiodic messages varies significantly under different conditions, and temporal constraints of these messages cannot be effectively guaranteed by current mechanism utilizing PT priority and PT circulation period. In current mechanism, PT priority changes only at node 1 within a new PT cycle if there is enough difference between previous APTCP and SPTCP. Obviously, this mechanism is too simple to effectively adapt to the dynamic traffic of aperiodic messages and meet to the respond time requirement of *urgent* and *normal* aperiodic messages.

Although minimum arrival period for aperiodic messages is assumed, it is just for convenience of analysis. In fact, the following case may occur: none aperiodic message arrives in a PT cycle, which causes PT priority to maintain or change to *normal.* Then in the next PT cycle, all aperiodic messages arrive at their maximum rates. Consequently *urgent* aperiodic messages suffer from larger response time. As for an *urgent* aperiodic message, a fairly small SPTCP cannot guarantee the time requirement of its each instance, unless setting SPTCP as small as $\mathrm{SPTCP_L}$. However, the latter selection will cause *normal* aperiodic messages to lose opportunity to transmit forever. As for a *normal* aperiodic message, a small SPTCP will make PT priority maintain at *urgent* because of the overhead of sending and returning PT even no *urgent* aperiodic messages is transmitted (as for this argument, you can refer to Section 3).

Therefore, it can be argued that the goal of PT priority and aperiodic messages priority cannot be realized only through a properly set SPTCP under current APTCP computing mechanism. Consequently, it is necessary to propose a new one to enhance timely responsibility of FF for actual load to guarantee the temporal constraint of *urgent* and *normal* aperiodic messages.

### 4.2. Improvement of APTCP computing mechanism

It is obvious that the new mechanism should timely respond the real-time requirement of each *urgent* aperiodic message. To achieve this goal, the improved APTCP (IAPTCP) should timely indicate the real-time requirement of each *urgent* aperiodic message in each node. Naturally, improved mechanism proposed in this paper is that IAPTCP equals to the elapsed time between PT arriving consecutively twice at any same node instead of node 1 only. Additionally, this paper introduces how to set ISPTCP (SPTCP under the improved mechanism).

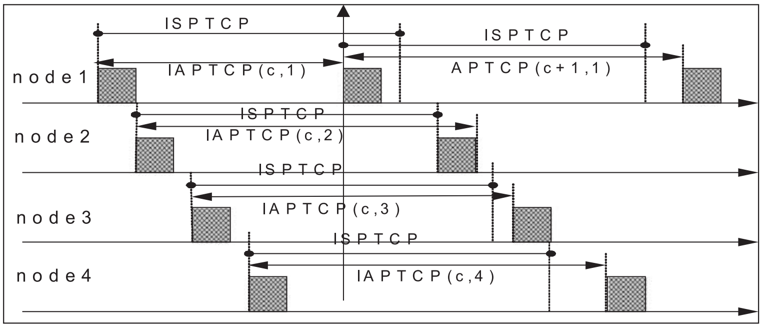


Fig. 5. Relationship between *IAPTCP*(*c*,*i*) and ISPTCP.

Let *IAPTCP*(*c,i*) denote the elapsed time between *visit*(*c,i*) and *visit*($c+1$,*i*), which is illustrated in Fig. 5.

It is obvious that new mechanism can maintain the upper bound for WRCT of *urgent* aperiodic messages through an easy method by setting a proper SPTCP.

Here, $M_{\mathrm{U}}^{i,j}$ is taken as example for explaining how to set ISPTCP. Assume that $M_{\mathrm{U}}^{i,j}$ arrives at node *i* just after PT leaves this node and current PT priority is *normal*. Assume that *IAPTCP*(*c,x*) being larger than SPTCP occurs at instant of *visit*($c+1$,*x*), then PT priority changes from *normal* to *urgent*. Furthermore, assume that PT priority at least keeps *urgent* until *visit*($c+1$,*i*). In this scenario, the WCRT of $M_{\mathrm{U}}^{i,j}$ can be determined.

Let $R_{\mathrm{U}}^{i,j}$ denote response time of $M_{\mathrm{U}}^{i,j}$ under the new mechanism.

$$R_{\mathrm{U}}^{i,j} = \mathrm{SPTCP} + (\zeta_{\mathrm{U}}^{i,j} - 1)\cdot \mathrm{MicP} + \sum_{l=1}^{n_{\mathrm{P}}} ST[l,(\beta + \zeta_{\mathrm{U}}^{x,i})*] \cdot C_{\mathrm{P}}^{l} + \sum_{l=1}^{n_{\mathrm{U}}^{i}} C_{\mathrm{U}}^{k,l} \qquad (25)$$

where $\dddot{\zeta}_{\mathrm{U}}^{x,y}$ denote the number of microcycles for completing the transfer of $M_{\mathrm{U}}^{i,j}$.

$$\dddot{\zeta}_{\mathrm{U}}^{x,y} = \min(\psi) \cap \sum_{i=\beta}^{\beta+\psi-1} W_{\mathrm{A}}^{i} \geq \dddot{\Psi}_{\mathrm{U}}^{i,j}$$

$$\dddot{\Psi}_{\mathrm{U}}^{x,y} = \begin{cases} \sum_{k=x}^{y} \sum_{j=1}^{nd_k^u} C_{\mathrm{U}}^{k,l} & (y > x) \\ \sum_{k=x}^{n} \sum_{l=1}^{n_{\mathrm{U}}^{k}} C_{\mathrm{U}}^{k,l} + \sum_{k=1}^{y} \sum_{j=1}^{nd_k^n} C_{\mathrm{U}}^{k,l} & (y < x) \end{cases}$$

For convenience of describing its WCRT of $M_{\mathrm{U}}^{i,j}$, let $\dddot{\xi}_{\mathrm{U}}^{i,x}$ denote the left delay after the promotion of PT priority that occurs at the instant of *visit*(*c,x*) ($x \geq i$) or *visit*($c,+1,i$) ($x < i$) under the new mechanism.

$$\dddot{\xi}_{\mathrm{U}}^{i,x} = (\dddot{\zeta}_{\mathrm{U}}^{x,i} - 1)\cdot \mathrm{MicP} + \sum_{l=1}^{n_{\mathrm{P}}} ST[l,(\beta + \dddot{\zeta}_{\mathrm{U}}^{x,i})*] \cdot C_{\mathrm{P}}^{l} + \sum_{l=1}^{n_{\mathrm{U}}^{i}} C_{\mathrm{U}}^{k,l} \qquad (26)$$

In order to guarantee the time constraints of *urgent* aperiodic messages, ISPTCP must meet the following requirement:

$$D_{\mathrm{U}}^{i,j} \geq \dddot{R}_{\mathrm{U}}^{i,j} = \mathrm{ISPTCP}^{i} + \dddot{\xi}_{\mathrm{U}}^{i,x} \qquad (27)$$

Furthermore, determine different SPTCPs can be determined according to the traffic of *normal* and Available aperiodic messages permitted before the promotion of PT priority.

$$\mathrm{ISPTCP}^{i} = \min(D_{\mathrm{U}}^{i,j} - \dddot{\xi}_{\mathrm{U}}^{i,x} \mid i,x \in n,\ j \in n_{\mathrm{U}}^{i}) \qquad (28)$$

For $M_{\mathrm{U}}^{i,j}$, there are different positions for the occurrence of *IAPTCP*(*c,x*) or *IAPTCP*($c+1, x$) being larger than SPTCP, among which *IAPTCP*($c,\ i+1$)($c \neq n$) or *IAPTCP*($c+1, 1$) ($c = n$) is the most effective position for preventing the transfer of *normal* aperiodic messages. Under this case, expression (29) is derived

$$\mathrm{ISPTCP}^{i} = \min(D_{\mathrm{U}}^{i,j} - \dddot{\zeta}_{\mathrm{U}}^{i,(i+x)\%i} \mid i,x,\in,\ j \in n_{\mathrm{U}}^{i}) \qquad (29)$$

Since

$$D_{\mathrm{U}}^{i,j} \geq \dddot{R}_{\mathrm{U}}^{i,j} = \mathrm{ISPTCP}^{i} + \dddot{\xi}_{\mathrm{U}}^{i,(i+x)\%n}.$$

## 5. Performance evaluation of improved mechanism

A simulation model has been constructed using OPNET Modeler 8.0. The main objective of the simulation study is to gain insight into the performance of the proposed APTCP computing mechanism. It is assumed that eight nodes are connected using FF. Each node produces two aperiodic message streams: one is a Poisson stream with *urgent* priority, packet length of 60 bits and arrival rate of 10 packet/s, the other is a Poisson stream with *normal* priority, packet length of 120 bits and arrival rate of 20 packet/s (for the simplicity of simulation, periodic messages are not to be considered). The service capability of processing module is 31,250 bps. PT is transferring between LAS and nodes, which takes one fifth of the service time of an *urgent* aperiodic message. In our simulation, both SPTCP and ISPTCP are set to 0.05 s.

Firstly, previous APTCP computing mechanism is considered, where the APTCP is computed and accordingly PT priority is changed only at node 1 in each cycle. Then, the improved APTCP computing mechanism is considered, where the APTCP is computed and accordingly PT priority is changed at every node in each cycle. Time duration of both simulations is set to 1000 s. From each simulation, the WCRT data of *urgent* aperiodic messages in each node can be obtained. These data are processed and plotted as shown in following two figures.

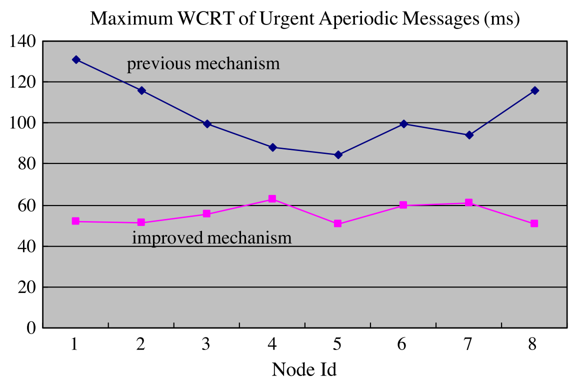


Fig. 6. Maximum WCRT of *urgent* aperiodic messages.

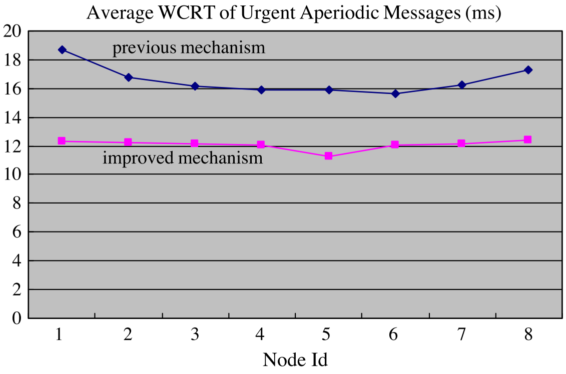


Fig. 7. Average WCRT of *urgent* aperiodic messages.

As shown in Fig. 6, our improved APTCP computing mechanism decreases the maximum WCRT of *urgent* aperiodic messages dramatically. The similar result can be found in Fig. 7, where the average WCRT of *urgent* periodic messages goes down slightly in the improved mechanism than in the previous mechanism. By adjusting the PT priority in every node dynamically, this new mechanism becomes more sensitive to the change of real traffic load in FF and improves the capability of FF to meet the real-time requirement of *urgent* aperiodic messages greatly.

Moreover, Figs. 6 and 7 show that the fairness between different nodes increases significantly. In the previous mechanism, the average WCRT of *urgent* aperiodic messages in former nodes is much high than that in latter nodes. However, in our improved mechanism, the average WCRT of *urgent* aperiodic messages is nearly equal in different nodes.

## 6. Conclusions

Supporting real-time traffic of critical aperiodic messages using FF is a complicated issue since it is concerned with ST, APTCP, SPTCP and PT priority, etc. This paper first proposes an integrated message transmission model, which integrates periodic messages with aperiodic messages together. Then formulas for the response time of *urgent* and *normal* aperiodic messages are given. Through the formula, deficiency of current APTCP computing mechanism in meeting real-time traffic of *urgent* and *normal* aperiodic messages is found. Furthermore, this paper proposes a new APTCP computing mechanism and gives a methodology for setting SPTCP. Finally, the simulation results show that the proposed APTCP computing mechanism improves real-time performance of urgent aperiodic messages significantly. The ongoing work is to optimize aperiodic message priorities under some criterion, such as response time, loss rate, utilization, etc.

## References

[1] H. Kopetz, Real Time System Design Principles for Distributed Embedded Application, Kluwer Academic Publishing, Boston, USA, 1997.

[2] G.C. Buttazzo, Hard Real-Time Computing Systems: Predictable Scheduling Algorithms and Application, Kluwer Academic Publishing, Boston, USA, 1997.

[3] W. Zhao, J. Stankovic, K. Ramamritham, A window protocol for transmission of time constrained messages, IEEE Transactions on Computers 39 (9) (1990) 1186–1203.

[4] S. Zhang, E. Lee, A. Burns, H. Yang, Finding the minimum available transmission time for the timed token medium access control protocol, Proc. of ECRTS'99 (11th Euromicro Conference on Real-Time Systems York, England, UK), (June 9-11, 1999) 185–194.

[5] E. Tovar, F. Vasques, Distributed computing for the factory-floor: a real-time approach using WorldFIP networks, Computers in Industry 44 (1) (2000) 11–30.

[6] J. Thomesse, The Fieldbus, Proc. of SICICA'97 (3rd IFAC Symposium on intelligent Components and Instruments for Control Application), (June 1997) 473–478, Annecy, France.

[7] P. Raja, G. Noubir, Static and dynamic polling mechanism for Fieldbus networks, Operating System Overview 27 (3) (1993) 34–43.

[8] L. Almeida, P. Pedreiras, J. Fonseca, FTT-CAN: why and how, IEEE Transactions on Industrial Electronics 49 (6) (2002) 1189–1201.

[9] L. Almeida, E. Tovar, J.A. Fonseca, F. Vasques, Schedulability analysis of real-time traffic in WorldFIP networks: an integrated approach, IEEE Transactions on Industrial Electronics 49 (5) (2002) 1115–1174.

[10] Z. Wang, Y.Q. Song, H.B. Yu, Y.X. Sun, Worst-case Response Time of Aperiodc Message in WorldFIP and Its Improvement in Real-time Capability, ISA TRANSACTION, accepted and to appear.

[11] F. Simonot, Y.Q. Song, Real-time communication using TDMA-based multi-access protocol, Computer Communications 20 (6) (1997) 435–448.

[12] S. Cavalieri, A. Stefano, O. Mirabella, Pre-run-time scheduling to reduce schedule length in the Fieldbus environment, IEEE Transactions on Software Engineering 21 (11) (1995) 865–880.

[13] S.H. Hong, Bandwidth allocation scheme for cyclic-service Fieldbus networks, IEEE/ASME Transactions on Mechatronics 6 (2) (2001 (June)) 197–204.

[14] Z. Wang, Y. Zhou, J.M. Chen, Y.Q. Song, Y.X. Sun, Real-time characteristic of FF like centralized control Fieldbus and its state-of-art, IEEE International Symposium on Industrial Electronics (ISIE 2002) 1 (2002 (July)) 140–145(Italy).

[15] S.H. Hong, Y.C. Kim, Implementation of a bandwidth allocation scheme in a token-passing fieldbus network, IEEE Transactions on Instrumentation and Measurement 51 (2) (2002 April) 246–251.

[16] Z. Wang, H.P. Dai, Y.Q. Song, Y.X. Sun, The effect of Pass Token period on Real-time Communication of Urgent Aperiodic Message and Its improvement, IEEE International Symposium on Industrial Electronics (ISIE 2003) Vol. 1 (2003 (June)) 472–474.

[17] Z. Wang, Y.Q. Song, H.B. Yu, Y.X. Sun, Stability analysis for multi-class multi-queue single server system under polling table, American Control Conference (ACC2002) 6 (2002 (May)) 4732–4737.

**Zhi Wang**, born in 1969, received his MS degree from Southeast University in 1997 and PhD degree from Shenyang Institute of Automation of Chinese Academy of Science in 2000. He has been doing Post-Doctoral Research in Zhejiang University and National Polytechnique Institute de Lorraine from 2000 to Sep. 2002. He was a senior Visitor and Researcher in LORIA TRIO from July 2002 to Aug. 2002 and from Sep. 2003 to Nov. 2003. In July 2002, he was promoted to Associate Professor. At present, he is the leader of RoDaC (Open Realtime Distributed Control Architecture) group and mainly researchers networked control and embedded system, including discrete event system, realtime schedule, IP networks, sensor network and open distributed control system based on IEC Function block.

**Xingfa Shen** received his B.S. degree in Electrical Engineering from Zhejiang University, China, in 2000. He is currently a PhD student in College of Information Science and Engineering of University of Zhejiang University, China. His current research interests include realtime network scheduling, network modeling and wireless sensor networks protocols.

**Jiming Chen** received his B.S. degree in Electrical Engineering from Zhejiang University, China, in 2000. He is currently a PhD student in College of Information Science and Engineering of University of Zhejiang University, China. His current research interests include real-time schedule, weakly real-time analysis, Real-time QoS of industrial Ethernet, fieldbus-based distributed control system.